Scalable Computing: Practice and Experience

Scientific International Journal
for Parallel and Distributed Computing

ISSN: 1895-1767

Universitatea de Vest
din Timişoara

Volume 21(3) September 2020

SUBSCRIPTION INFORMATION: please visit http://www.scpe.org

Scalable Computing: Practice and Experience

Volume 21, Number 3, September 2020

TABLE OF CONTENTS

Special Issue on Evolving IoT and Cyber-Physical Systems: Advancements, Applications, and Solutions:

Regular Papers:

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 345–348, DOI 10.12694:/scpe.v21i3.1568

INTRODUCTION TO THE SPECIAL ISSUE ON EVOLVING IOT AND CYBER-PHYSICAL SYSTEMS: ADVANCEMENTS, APPLICATIONS, AND SOLUTIONS

ANAND NAYYAR[*], PIJUSH KANTI DUTTA PRAMANIK[†] AND RAJNI MOHANA[‡]

Internet of Things (IoT) is regarded as a next-generation wave of Information Technology (IT) after the widespread emergence of the Internet and mobile communication technologies. IoT supports information exchange and networked interaction of appliances, vehicles and other objects, making sensing and actuation possible in a low-cost and smart manner.

On the other hand, cyber-physical systems (CPS) are described as the engineered systems which are built upon the tight integration of the cyber entities (e.g., computation, communication, and control) and the physical things (natural and man-made systems governed by the laws of physics).

The IoT and CPS are not isolated technologies. Rather it can be said that IoT is the base or enabling technology for CPS and CPS is considered as the grownup development of IoT, completing the IoT notion and vision. Both are merged into closed-loop, providing mechanisms for conceptualizing, and realizing all aspects of the networked composed systems that are monitored and controlled by computing algorithms and are tightly coupled among users and the Internet. That is, the hardware and the software entities are intertwined, and they typically function on different time and location-based scales. In fact, the linking between the cyber and the physical world is enabled by IoT (through sensors and actuators). CPS that includes traditional embedded and control systems are supposed to be transformed by the evolving and innovative methodologies and engineering of IoT.

Several applications areas of IoT and CPS are smart building, smart transport, automated vehicles, smart cities, smart grid, smart manufacturing, smart agriculture, smart healthcare, smart supply chain and logistics, etc. Though CPS and IoT have significant overlaps, they differ in terms of engineering aspects. Engineering IoT systems revolves around the uniquely identifiable and internet-connected devices and embedded systems; whereas engineering CPS requires a strong emphasis on the relationship between computation aspects (complex software) and the physical entities (hardware).

Engineering CPS is challenging because there is no defined and fixed boundary and relationship between the cyber and physical worlds. In CPS, diverse constituent parts are composed and collaborated together to create unified systems with global behaviour. These systems need to be ensured in terms of dependability, safety, security, efficiency, and adherence to real-time constraints. Hence, designing CPS requires knowledge of multidisciplinary areas such as sensing technologies, distributed systems, pervasive and ubiquitous computing, real-time computing, computer networking, control theory, signal processing, embedded systems, etc.

CPS, along with the continuous evolving IoT, has posed several challenges. For example, the enormous amount of data collected from the physical things makes it difficult for Big Data management and analytics that includes data normalization, data aggregation, data mining, pattern extraction and information visualization. Similarly, the future IoT and CPS need standardized abstraction and architecture that will allow modular designing and engineering of IoT and CPS in global and synergetic applications. Another challenging concern of IoT and CPS is the security and reliability of the components and systems.

Although IoT and CPS have attracted the attention of the research communities and several ideas and solutions are proposed, there are still huge possibilities for innovative propositions to make IoT and CPS vision

[*] Graduate School, Duy Tan University, Da Nang, Vietnam
[†] National Institute of Technology, Durgapur, India
[‡] Jaypee Institute of Information Technology, Wakanghat, India

successful. The major challenges and research scopes include system design and implementation, computing and communication, system architecture and integration, application-based implementations, fault tolerance, designing efficient algorithms and protocols, availability and reliability, security and privacy, energy-efficiency and sustainability, etc.

It is our great privilege to present Volume 21, Issue 3 of Scalable Computing: Practice and Experience. We had received 30 research papers and out of which 14 papers are selected for publication. The objective of this special issue is to explore and report recent advances and disseminate state-of-the-art research related to IoT, CPS and the enabling and associated technologies. The special issue will present new dimensions of research to researchers and industry professionals with regard to IoT and CPS.

Vivek Kumar Prasad and Madhuri D Bhavsar in the paper titled "Monitoring and Prediction of SLA for IoT based Cloud described the mechanisms for monitoring by using the concept of reinforcement learning and prediction of the cloud resources, which forms the critical parts of cloud expertise in support of controlling and evolution of the IT resources and has been implemented using LSTM. The proper utilization of the resources will generate revenues to the provider and also increases the trust factor of the provider of cloud services. For experimental analysis, four parameters have been used i.e. CPU utilization, disk read/write throughput and memory utilization.

Kasture et al. in the paper titled "Comparative Study of Speaker Recognition Techniques in IoT Devices for Text Independent Negative Recognition" compared the performance of features which are used in state of art speaker recognition models and analyse variants of Mel frequency cepstrum coefficients (MFCC) predominantly used in feature extraction which can be further incorporated and used in various smart devices.

Mahesh Kumar Singh and Om Prakash Rishi in the paper titled "Event Driven Recommendation System for E-Commerce using Knowledge based Collaborative Filtering Technique" proposed a novel system that uses a knowledge base generated from knowledge graph to identify the domain knowledge of users, items, and relationships among these, knowledge graph is a labelled multidimensional directed graph that represents the relationship among the users and the items. The proposed approach uses about 100 percent of users' participation in the form of activities during navigation of the web site. Thus, the system expects under the users' interest that is beneficial for both seller and buyer. The proposed system is compared with baseline methods in area of recommendation system using three parameters: precision, recall and NDGA through online and offline evaluation studies with user data and it is observed that proposed system is better as compared to other baseline systems.

Benbrahim et al. in the paper titled "Deep Convolutional Neural Network with TensorFlow and Keras to Classify Skin Cancer" proposed a novel classification model to classify skin tumours in images using Deep Learning methodology and the proposed system was tested on HAM10000 dataset comprising of 10,015 dermatoscopic images and the results observed that the proposed system is accurate in order of 94.06% in validation set and 93.93% in the test set.

Devi B et al. in the paper titled "Deadlock Free Resource Management Technique for IoT-Based Post Disaster Recovery Systems" proposed a new class of techniques that do not perform stringent testing before allocating the resources but still ensure that the system is deadlock-free and the overhead is also minimal. The proposed technique suggests reserving a portion of the resources to ensure no deadlock would occur. The correctness of the technique is proved in the form of theorems. The average turnaround time is approximately 18% lower for the proposed technique over Banker's algorithm and also an optimal overhead of O(m).

Deep et al. in the paper titled "Access Management of User and Cyber-Physical Device in DBAAS According to Indian IT Laws Using Blockchain" proposed a novel blockchain solution to track the activities of employees managing cloud. Employee authentication and authorization are managed through the blockchain server. User authentication related data is stored in blockchain. The proposed work assists cloud companies to have better control over their employee's activities, thus help in preventing insider attack on User and Cyber-Physical Devices.

Sumit Kumar and Jaspreet Singh in paper titled "Internet of Vehicles (IoV) over VANETS: Smart and Secure Communication using IoT" highlighted a detailed description of Internet of Vehicles (IoV) with current applications, architectures, communication technologies, routing protocols and different issues. The researchers also elaborated research challenges and trade-off between security and privacy in area of IoV.

Deore et al. in the paper titled ”A New Approach for Navigation and Traffic Signs Indication Using Map Integrated Augmented Reality for Self-Driving Cars” proposed a new approach to supplement the technology used in self-driving cards for perception. The proposed approach uses Augmented Reality to create and augment artificial objects of navigational signs and traffic signals based on vehicles location to reality. This approach help navigate the vehicle even if the road infrastructure does not have very good sign indications and marking. The approach was tested locally by creating a local navigational system and a smartphone based augmented reality app. The approach performed better than the conventional method as the objects were clearer in the frame which made it each for the object detection to detect them.

Bhardwaj et al. in the paper titled ”A Framework to Systematically Analyse the Trustworthiness of Nodes for Securing IoV Interactions” performed literature on IoV and Trust and proposed a Hybrid Trust model that seperates the malicious and trusted nodes to secure the interaction of vehicle in IoV. To test the model, simulation was conducted on varied threshold values. And results observed that PDR of trusted node is 0.63 which is higher as compared to PDR of malicious node which is 0.15. And on the basis of PDR, number of available hops and Trust Dynamics the malicious nodes are identified and discarded.

Saniya Zahoor and Roohie Naaz Mir in the paper titled ”A Parallelization Based Data Management Framework for Pervasive IoT Applications” highlighted the recent studies and related information in data management for pervasive IoT applications having limited resources. The paper also proposes a parallelization-based data management framework for resource-constrained pervasive applications of IoT. The comparison of the proposed framework is done with the sequential approach through simulations and empirical data analysis. The results show an improvement in energy, processing, and storage requirements for the processing of data on the IoT device in the proposed framework as compared to the sequential approach.

Patel et al. in the paper titled ”Performance Analysis of Video ON-Demand and Live Video Streaming Using Cloud Based Services” presented a review of video analysis over the LVS & VoDS video application. The researchers compared different messaging brokers which helps to deliver each frame in a distributed pipeline to analyze the impact on two message brokers for video analysis to achieve LVS VoS using AWS elemental services. In addition, the researchers also analysed the Kafka configuration parameter for reliability on full-service-mode.

Saniya Zahoor and Roohie Naaz Mir in the paper titled ”Design and Modeling of Resource-Constrained IoT Based Body Area Networks” presented the design and modeling of a resource-constrained BAN System and also discussed the various scenarios of BAN in context of resource constraints. The Researchers also proposed an Advanced Edge Clustering (AEC) approach to manage the resources such as energy, storage, and processing of BAN devices while performing real-time data capture of critical health parameters and detection of abnormal patterns. The comparison of the AEC approach is done with the Stable Election Protocol (SEP) through simulations and empirical data analysis. The results show an improvement in energy, processing time and storage requirements for the processing of data on BAN devices in AEC as compared to SEP.

Neelam Saleem Khan and Mohammad Ahsan Chishti in the paper titled ”Security Challenges in Fog and IoT, Blockchain Technology and Cell Tree Solutions: A Review” outlined major authentication issues in IoT, map their existing solutions and further tabulate Fog and IoT security loopholes. Furthermore, this paper presents Blockchain, a decentralized distributed technology as one of the solutions for authentication issues in IoT. In addition, the researchers discussed the strength of Blockchain technology, work done in this field, its adoption in COVID-19 fight and tabulate various challenges in Blockchain technology. The researchers also proposed Cell Tree architecture as another solution to address some of the security issues in IoT, outlined its advantages over Blockchain technology and tabulated some future course to stir some attempts in this area.

Bhadwal et al. in the paper titled ”A Machine Translation System from Hindi to Sanskrit Language Using Rule Based Approach” proposed a rule-based machine translation system to bridge the language barrier between Hindi and Sanskrit Language by converting any test in Hindi to Sanskrit. The results are produced in the form of two confusion matrices wherein a total of 50 random sentences and 100 tokens (Hindi words or phrases) were taken for system evaluation. The semantic evaluation of 100 tokens produce an accuracy of 94% while the pragmatic analysis of 50 sentences produce an accuracy of around 86%. Hence, the proposed system can be used to understand the whole translation process and can further be employed as a tool for learning as well as teaching. Further, this application can be embedded in local communication based assisting Internet of Things (IoT) devices like Alexa or Google Assistant.

Anshu Kumar Dwivedi and A.K. Sharma in the paper titled ”NEEF: A Novel Energy Efficient Fuzzy Logic Based Clustering Protocol for Wireless Sensor Network” proposed a a deterministic novel energy efficient fuzzy logic-based clustering protocol (NEEF) which considers primary and secondary factors in fuzzy logic system while selecting cluster heads. After selection of cluster heads, non-cluster head nodes use fuzzy logic for prudent selection of their cluster head for cluster formation. NEEF is simulated and compared with two recent state of the art protocols, namely SCHFTL and DFCR under two scenarios. Simulation results unveil better performance by balancing the load and improvement in terms of stability period, packets forwarded to the base station, improved average energy and extended lifetime.

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 349–357, DOI 10.12694:/scpe.v21i3.1697

MONITORING AND PREDICTION OF SLA FOR IOT BASED CLOUD

VIVEK KUMAR PRASAD AND MADHURI D BHAVSAR[*]

Abstract. Internet of Things (IoT) and cloud computing are the expertise captivating the technology. The most astonishing thing is their interdependence. IoT deals with the production of an additional amount of information that requires transmission of data, storage, and huge infrastructural processing power, posing a solemn delinquent. This is where cloud computing fits into the scenario. Cloud computing can be treated as the utility factor nowadays and can be used by pay as you go manner. As a cloud is a multi-tenant approach, and the resources will be used by multiple users. The cloud resources are required to be monitored, maintained, and configured and set-up as per the need of the end-users. This paper describes the mechanisms for monitoring by using the concept of reinforcement learning and prediction of the cloud resources, which forms the critical parts of cloud expertise in support of controlling and evolution of the IT resources and has been implemented using LSTM. The resource management system coordinates the IT resources among the cloud provider and the end users; accordingly, multiple instances can be created and managed as per the demand and availability of the support in terms of resources. The proper utilization of the resources will generate revenues to the provider and also increases the trust factor of the provider of cloud services. For experimental analysis, four parameters have been used i.e. CPU utilization, disk read/write throughput and memory utilization. The scope of this research paper is to manage the Cloud Computing resources during the peak time and avoid the conditions of the over and under-provisioning proactively.

Key words: Internet of Thing, Cloud Computing; Monitoring; Reinforcement Learning; Prediction; LSTM; Resource Management; Trust

AMS subject classifications. 68M14

1. Introduction.

When Cloud computing combines with the heights of IoT, which continuously processes the information/ data stream, this combination proves to be a boon to the business applications/industries and allowing the services to work from anywhere, anytime, and from any devices. The abilities of the Cloud range from assigning infinite pool of resources through virtualization, handle on Demand services and rapid elasticity. The cloud refers to an IT environs which is designed for remote provisioning of the resources in measured and scalable ways [1]. Much of the internet is devoted to the access of content based IT resources published via WWW (World Wide Web). The cloud can be grounded on any protocols that permit for the remote access of the infrastructure/ resources. This provides mainly three services and is known as SaaS(Software as a service), PaaS (Platform as a service) and IaaS (Infrastructure as a service) [2]. In this paper, the monitoring and prediction approach of the cloud is described. The cloud resources management supports to harmonize the IT resources and its management required by both the end users and the cloud provider. The core part for the same will is performed by Virtual Infrastructure Manager (VIM) which coordinated to the server hardware to create the numbers of Virtual Machines(VMs) or instances as per the user demand [3].

Following are the automated tasks that should be implemented for resource management:

- Creating server images by managing Virtual IT resources.
- Releasing and allocating the virtual IT resources into the available physical infrastructure with responses, such as IT resource instances termination, resuming, starting and pausing.
- Coordinating resources such as replication of the resources fault tolerance and load balancing.
- Applying security policies and usages of cloud service instances
- Monitoring operative circumstances of the IT resources of the cloud.

The resources should be always available when the user needs this, for the same the SLA (service level agreement) should be maintained [4], A SLA is a form of the document that acts as a contract in between the

[*]Nirma University, India (vivek.prasad@nirmauni.ac.in, madhuri.bhavsar@nirmauni.ac.in)

service provider and the end user. The SLA defines the standard service [5] that the cloud service provider is obligated for.

The prediction [6] for the advance violation for the SLAs can be fruitful to maintain the SLAs in proactive ways. This is an encouraging task because this allows the cloud service provider not only to study from the past disappointment but actually avoid them in the first occurrence, this is achieved by monitoring the cloud resource utilization.

SLA Monitor: The SLA monitoring mechanism [7] is castoff to witness the runtime performance of the cloud facilities to ensure that they are satisfying the contractual QoS needs to which are listed in the SLAs. The data collected by the SLA monitoring is aggregated into the SLA reporting metrics to recognize the numerous influences which might create downtime in the cloud. The system can proactively failover or repair the cloud facilities when incomparable conditions occur (identified because of the operational policies by using metrics), like as when the SLA monitoring intelligence found the cloud facility as "down". Or in other words, the SLA monitoring construction key concern is to proactively monitor the SLA in the mandate to predict the probable violations before they happen.

The repeated violation of the SLAs experienced by the end user's will damage the image of the cloud service provider. Hence it's a duty of the provider to reduce the violations happened because of SLA degradation. A good monitoring and prediction technique will solve these types of delinquent.

Motivation: Dynamic provisioning via resource scheduling is a difficult task and requires proper management of cloud resources. As Cloud contains an infinite pool of resources and managing these resources, optimally is an open research challenge. Hence motivated by these, a self-learning based framework has been implemented through cloud resource monitoring using Reinforcement learning and prediction of the cloud resources using Long Short Term Memory has been analyzed on IoT based Cloud. This framework reduces the conditions of the over and underprovisioning. It ensures the timely management of cloud resources and increased revenue for the Cloud Service Provider.

Contributions

- The paper deals with the proactive technique to maintain the SLAs with the viewpoint of the cloud service provider.
- The prediction technique will make use of the historical and present status of the cloud, which will be fulfilled because of the monitoring mechanism.
- The cloud service provider can create trust and reputation for their services by making use of these features.
- These techniques will be useful for admission control and capacity management.
- The conditions like over provisioning and under-provisioning can be solved out.

The structure of this paper is systematized as follows. Section 2 designates related studies. Section 3 deliberates system architecture and the proposed approach. Section 4 deliberates implementation and evaluation. Section 5 concludes the paper with future work.

2. Related studies. Mian et al [8] discussed the cost model that balanced the SLAs violation penalties and the cost of the resources. Singh et al.[9] describes the various categories of resources and their associated cloud application requirements. In another research done by Ayad et al. [10], an action used to trigger when the availability of virtual machines found to be unsatisfactory and the intelligent will automatically move to the adjoining healthy virtual machine.

Kousiouris, G. et al.[11], proposed a solution for prediction was it predicts the user's behavior patterns through time series analysis. Rimal et al.[12] Compared the cloud mechanisms in terms of its architecture, load balancing terminology, security, the framework of the programming, storage and virtualization. In another approach Buyya et al.[13] The resource manager acts as an admission controller and reallocates the resources as and when the resources deviate.

3. Proposed Approach And Its Architecture. Reinforcement learning (RL) is one of the prominent areas of machine learning [14]. It is about captivating the appropriate actions to maximize the reward in a specific condition. It has been deployed by various machines and software to identify the finest likely behavior it should implement in the specific scenario. There are numerous research has been done to implement RL and is well suited to cloud environs as they do not need a priori information of the application performance.

It learns the environment as the task/job runs. To work with RL the policies are required from which the positive and negative rewards will be generated, which tends to change over time and stops when the goal will be received. RL works with the fundamentals of the Bellman equation, as described below

$$Q(s,a) = r(s,a) + \gamma \max_{a} Q(s',a) \tag{3.1}$$

$Q(s,a)$ is the target, $r(s,a)$ is the reward of taking that action at that corresponding state and $\gamma \max_a Q(s',a)$ is the discounted max Q value among all possible actions from next state.

LSTM: Long short term memory (LSTM) is an extension of the recurrent neural network, which essentially extends their memory [15]. These extended memories are used to store the imperative pieces of knowledge/ experiences that have very elongated time pauses in between. The LSTM's can delete, read and write data from its memory. It has three gates, which are named as output, forget and input gate for its memory. These gates regulate whether or not to let fresh input in, let it influence the output at the present time step and delete the data as this might not useful for the time being.

Architecture and Methodologies used:

- **Monitoring Agent and Reinforcement Learning:**Due to the vibrant or changing nature of resource loads and the complexity of the cloud ecosystem, it is difficult to set up the mathematical model for the energy-efficient resource provision policy. The lack of efficient resource provisioning due to real-time dynamic management of resources is handled with the aid of model-free Reinforcement Learning methodology. The RL intellect with the environmental circumstances of the cloud ecosystem and accomplish the best suitable management policies without precise domain knowledge. Hence this grantee the best effort to the resource allocation problem in the cloud. As the users' requests are dynamic, and the providers are very complex to handle these requests in real-time. RL seems to be a suitable candidate to handle such situations.
- **SLA Analyzer and Match Making Algorithm:** Every cloud users want to be sure about the quality of their services, which is issued by the cloud provider. In the cloud ecosystem, this quality guarantee includes assurances on services enactment and performance. The Service Level Agreement is used to manage the performance of the services. The SLA analyzer manages these agreed qualities of services by using metrics and various threshold values. The matchmaking algorithm is used for the resource allocation mechanism by comparing the number of incoming requests from the cloud users and various resources available in the Cloud ecosystem. SLA usages the matchmaking algorithm to define its metrics and calculate various performance parameters to maintain the quality of the service.
- **Cloud Manager:**The cloud manager takes care of the cloud resources and makes sure that they are working optimally and is integrating with the cloud users and their services.
- **Prediction Manager and LSTM:**LSTM analyzes or observes the future scope (prediction) of the cloud resources and identifies the behavior of the resource demand based on the present input sequences.

As shown in Fig. 3.1 the entire architecture and its tasks can be carried out in the following sequences.

Step 1: Monitoring Agent (Reinforcement Learning) will monitor the resources and will identify the time stamp or episodes by which the particular workload will be finished by making use of the policy to reach the final goal. The agent automatically identifies the standard conduct within the precise content and will exploit its performance.

Step 2: SLA Analyzing service will investigate the various parameters that need to be adjusted to maintain the SLA and form the policy by making use of metrics.

Step 3: Matchmaking algorithms will act as a capacity management process and will identify the resources available as per the current demand. Also identifies the tasks details and its associated infrastructure requirement, whether it's available or not, steps to bring the resources as per the current need.

Step 4: The resource manager will contain the database to carry out the resource management tasks and will receive the input from the matchmaking.

Step 5: Based on the data collected by the resource manager, the prediction approach using LSTM will be implemented to identify future resource demand.

The architecture discussed here will improve the availability and trust factor of the cloud and results in the increased revenue to the cloud service provider.

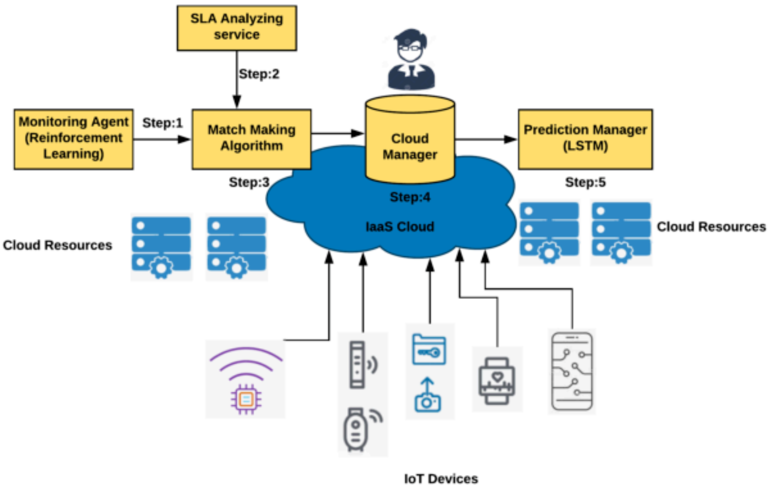


Fig. 3.1. *Architecture of the Cloud for SLA management*

4. Implementation and Evaluation. As discussed in section 3 the reinforcement learning experimentation have been carrying out into the data set of 1750 VMs of distributed data center and is used by various services that usages IoT plate form for the data processing. For the analysis of the experiments, the Openstack based private cloud have been used, with INTEL XEON SILVER CPU with 8 CPU cores and 16 number of threads and 128 GB of RAM. The parameter used is described in Table 4.1. The results are shown in Fig. 4.1 and Fig. 4.2 which make use of the CPU utilization. Here the workload has been classified as less than and greater than the threshold defined. The RL intellect with the environmental circumstances of the cloud ecosystem and accomplish the best suitable management policies without precise domain knowledge. Hence this grantee the best effort to the resource allocation problem in the cloud. The threshold is defined based upon the elastic nature of the cloud resources. As in the case of Figure 4.3 the CPU utilization is up to 35 percent.In an another case like in Figure 4.4 the variable peak rises up to 50 percent of the utilization of the CPU.

Once the results of the monitoring will be identified based on the epochs and rewards, the prediction approach using LSTM will determine the future demand of the resources for a particular type of workload in a proactive way as displayed in Fig.4.3, and Fig.4.4 for different ranges of CPU utilization.As such this can be analyzed that in Fig. 4.3 the range of the CPU utilization is upto 35 % and as per figure 4.4 the utilization has been reached till 50 %. These percentage of the work can be treated as different workload criteria. The parameter used here is depicted in Table 4.2. A total of five LSTM layers has been used and the every layer the neurons has been increased with percentage of 100. The grid search technique has been implemented to carry out the experimental work and to identify the hyper parameters.

Table 4.1
Parameters for reinforcement learning

Parameters	Values
Discount( )	0.8
Tau	0.01
Batch size	32
Layers	(50,50)
Learning rate	0.001
Epsilon decay fraction	0.4
Memory fraction	0.80
Memory type	Deque
Process_observation	Standardizer
Process_target	Normalizer

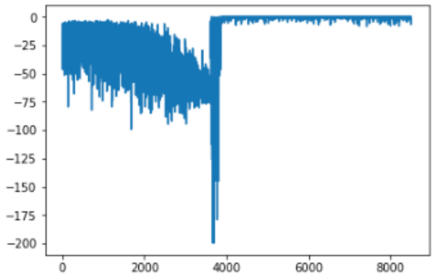


FIG. 4.1. *Graph for rewards generation for the cloud environment policy is to manage the resources in less than 70% threshold. X-Axis: Episodes and Y –Axis: Rewards*

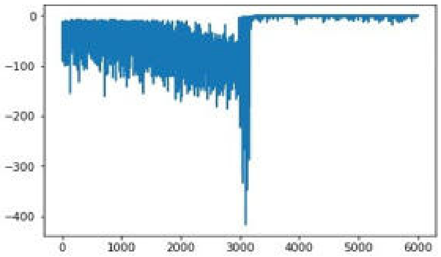


FIG. 4.2. *The CPU utilization is very high and more than the threshold defined i.e 70%. X-Axis: Episodes and Y –Axis: Rewards*

TABLE 4.2
Parameters used for LSTM

Parameters	Values
Batch size	64
Epochs	120
Time steps	10
Input layer	10 nodes
Output layer	10 nodes
Parameters for input layer	4 * LSTM output size * (weights of LSTM output size + 1 Bias variable)
Parameters for output layer	4 * LSTM output size * (weights of LSTM output size + 1 Bias variable)
Optimizer	Adam

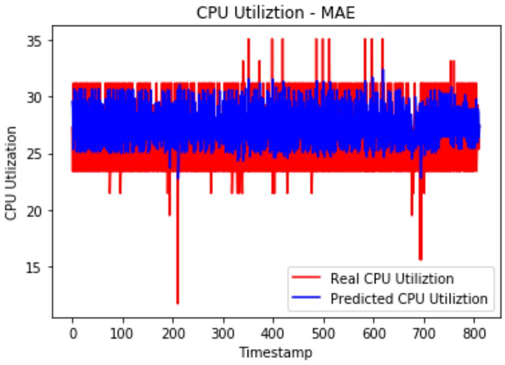


FIG. 4.3. *Predicted values using LSTM for CPU utilization where the usages of the CPU is from 0 to 35 percent*

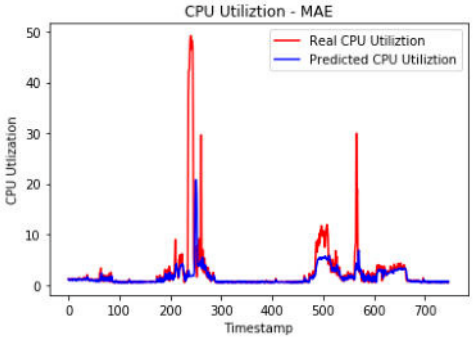


FIG. 4.4. *Predicted values using LSTM for CPU utilization with a range of 0 to 50 percent of CPU usage*

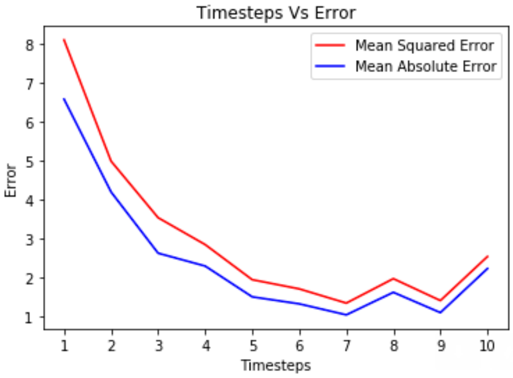


FIG. 4.5. *Variations in error with changing timestamps*

Results and Discussion

The results discussed can be well suited for capacity management of the cloud resources and is based on real-time monitoring and prediction features which usages the reinforcement learning and LSTM to support this framework. The monitoring results will identify the possible scenarios in which the type of tasks (workloads) can be completed in ideal situations with maximized enactment. The prediction results indicate the RMSE (root mean square error) rates as 1.09 for Fig. 4.3 and 1.28 for Fig. 4.4. The MAE (mean absolute error) for Fig. 4.3 and Fig. 4.4 are 0.89 and 0.85.

The changes in the timestamp or setting the window size for the prediction tend to deviate the values for the errors too and have been shown in Fig. 4.5 and the same can be identified by grid search too. Likewise, the parameters such as Disk read throughput[KB/s] and Disk write throughput [KB/s] were analyzed, and the concept of LSTM were implemented.The MAE and RMSE for Disk Read throughput has been calculated as 1.79 and 2.04 and is its graph is shown in Fig.4.6. similarly for Disk write throughout the values for MAE and RMSE are 1.36 and 2.15 and its graph is depicted in Fig.4.7.

The another parameter which has been taken into the consideration is memory utilization in percentage and the experimental value for MAE and RMSE values is calculated as 0.98 and 1.52.The same is depicted in the Fig. 4.8. By understanding the future usage demand from the current and past usage patterns of the resources, the cloud service provider can manage their resources. The prediction of the resources such as CPU Utilization, Disk Read Throughput, Disk write Throughput, and Memory utilization is of great importance for handling dynamic scaling of the resources support. It will achieve enhanced efficiency in terms of energy and cost consumption, and this also maintains the QoS.

Table 5.1 shows the various characteristics of Cloud Computing and its mapping with the proposed approach. It signifies that the proposed approach can handle Scalability [16] , Elasticity[17], Adaptability [18],

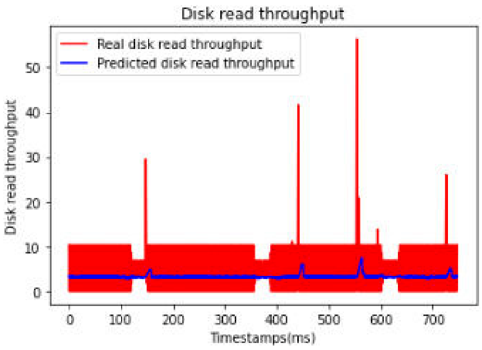


FIG. 4.6. *Disk read throughput[KBs] prediction using LSTM*

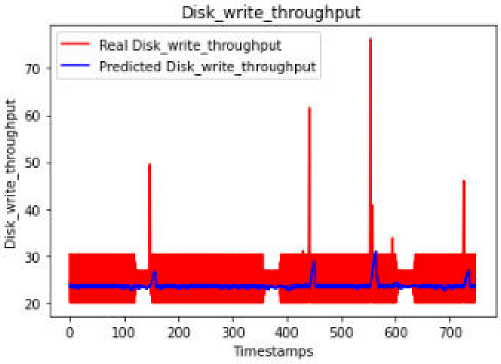


FIG. 4.7. *Disk write throughput[KBs] prediction using LSTM*

Autonomicity [19], Comprehensiveness [20], and Availability [21]. The Extensibility [22], Intrusiveness [23], Resilience [24], and Reliability [25] still needs to be identified and tested for the proposed scheme.

For the fulfillment of the cloud services with proper QoS requirements, a required amount of resources are provisioned by the Cloud Service Provider. Hence based upon the QoS, the SLA will be designed and defined for the smooth conduction of the services [26]. Even the SLA violations are detected regularly to impose the penalty among the parties [27]. So there is a requirement to provide an adequate amount of resources dynamically by the service provider and always avoid the SLA violations proactively. Our proposed approach will proactively avoid the conditions of the SLA violations and thus will be useful for managing the resources as well.

5. Conclusion and future work. To satisfy the end users, SLA violations should be avoided by the cloud service provider. Most of the research proposed the solutions or explanations of violations after they have occurred, the research paper solves this by making use of a proactive approach using the mechanism of monitoring and prediction. Reinforcement learning and LSTM has been used to implement the same. The proposed solution takes the input from the monitoring data and accordingly does the prediction about the resources and manages the capacity of the resources as per the demand.The proposed technique will help us to solve many real-time problems in the cloud environment, such as can be used for matchmaking algorithms, SLA management, capacity planning, and admission control.

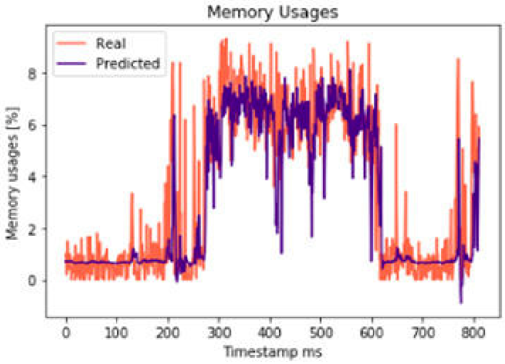


Fig. 4.8. *Predicted values using LSTM for Memory utilization with a range of 0 to 10 percent of Memory usage*

Table 4.3
Cloud Characteristics and its mapping with our proposed approach

Cloud Characteristics	**Proposed Approach**
Scalability	Yes
Elasticity	Yes
Adaptability	Yes
Automaticity	Yes
Comprehensiveness	Yes
Availability	Yes
Extensibility	No
Intrusiveness	No
Resilience	No
Reliability	No

REFERENCES

[1] Josep, A. D., Katz, R., Konwinski, A., Gunho, L. E. E., Patterson, D., and Rabkin, A.. A view of cloud computing. Communications of the ACM, 53 (4) (2010).

[2] Iosup, A., Ostermann, S., Yigitbasi, M. N., Prodan, R., Fahringer, T., and Epema, D. Performance analysis of cloud computing services for many-tasks scientific computing. IEEE Transactions on Parallel and Distributed Systems, 22(6), 931-945 (2011).

[3] Yeung, M., El Ajaltouni, E., Phillips, A., and Andersen, P. U.S. Patent Application No. 15/655,607 (2019).

[4] Gupta, S., Gupta, S. C., Majumdar, R., and Rathore, Y. S.. Measuring Cloud Security from risks perspective. In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence) (pp. 214-220). IEEE (2016).

[5] Casola, V., De Benedictis, A., Eraçu, M., Modic, J., and Rak, M.. Automatically enforcing security slas in the cloud. IEEE Transactions on Services Computing, 10(5), 741-755 (2017).

[6] Leitner, P., Ferner, J., Hummer, W., and Dustdar, S.. Data-driven and automated prediction of service level agreement violations in service compositions. Distributed and Parallel Databases, 31(3), 447-470 (2013).

[7] Emeakaroha, V. C., Netto, M. A., Brandic, I., and De Rose, C. A.. Application-level monitoring and SLA violation detection for multi-tenant cloud services. In Emerging Research in Cloud Distributed Computing Systems (pp. 157-186). IGI Global (2015).

[8] Prasad, V. K., and Bhavsar, M. Efficient Resource Monitoring and Prediction Techniques in an IaaS Level of Cloud Computing: Survey. In International Conference on Future Internet Technologies and Trends (pp. 47-55). Springer, Cham (2017).

[9] Singh, S., and Chana, I. QoS-aware autonomic resource management in cloud computing: a systematic review. ACM Computing Surveys (CSUR), 48(3), 42 (2016).

[10] Ayad, A., Dippel, U.: Agent-based monitoring of virtual machines. In: 2010 International Symposium in Information Technology (ITSim), vol. 1, pp. 1–6. IEEE (2010)

[11] Kousiouris, G., Menychtas, A., Kyriazis, D., Gogouvitis, S., Varvarigou, T., Dynamic, behavioral-based estimation of resource provisioning based on highlevel application terms in cloud platforms. Future Gener. Comput. Syst. 32, 27–40 (2014)

[12] Rimal, B. P., Choi, E., and Lumb, I.. A taxonomy and survey of cloud computing systems. In 2009 Fifth International Joint Conference on INC, IMS and IDC (pp. 44-51). IEEE (2009).

[13] Buyya, R., Broberg, J., and Goscinski, A.. Cloud computing. Principles and Paradigms. Wiley (2011).

[14] Rao, J., Bu, X., Xu, C. Z., Wang, L., and Yin, G.. VCONF: a reinforcement learning approach to virtual machines auto-configuration. In Proceedings of the 6th international conference on Autonomic computing (pp. 137-146). ACM (2009)

[15] Lai, C. F., Chien, W. C., Yang, L. T., and Qiang, W. (2019). LSTM and Edge Computing for Big Data Feature Recognition of Industrial Electrical Equipment. IEEE Transactions on Industrial Informatics.

[16] Yang, J., Qiu, J., and Li, Y. A profile-based approach to just-in-time scalability for cloud applications. In 2009 IEEE International Conference on Cloud Computing (pp. 9-16). IEEE (2009).

[17] Coutinho, Emanuel Ferreira, Flávio Rubens de Carvalho Sousa, Paulo Antonio Leal Rego, Danielo Gonçalves Gomes, and José Neuman de Souza. Elasticity in cloud computing: a survey. annals of telecommunications-annales des télécommunications 70, no. 7-8 (2015): 289-309.

[18] Khan, Suleman, Muhammad Shiraz, Ainuddin Wahid Abdul Wahab, Abdullah Gani, Qi Han, and Zulkanain Bin Abdul Rahman. A comprehensive review on adaptability of network forensics frameworks for mobile cloud computing. The Scientific World Journal (2014).

[19] Caromel, D. ProActive Parallel Suite: Multi-cores to Clouds to autonomicity. In 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing (pp. xi-xii). IEEE (2009).

[20] Durao, Frederico, Jose Fernando S. Carvalho, Anderson Fonseka, and Vinicius Cardoso Garcia. A systematic review on cloud computing. The Journal of Supercomputing 68, no. 3 : 1321-1346 (2014).

[21] Lei, L., Dagang, L., Lianwen, J., and Lihong, M.. Constructing a high availiable private cloud storage platform based on OpenStack Swift. Experimental Technology and Management, (5), 37 (2015).

[22] Copil, Georgiana, Daniel Moldovan, Hong-Linh Truong, and Schahram Dustdar. Sybl: An extensible language for controlling elasticity in cloud applications. In 2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, pp. 112-119. IEEE (2013).

[23] Bolte, Matthias, Michael Sievers, Georg Birkenheuer, Oliver Niehörster, and André Brinkmann. Non-intrusive virtualization management using libvirt. In 2010 Design, Automation and Test in Europe Conference and Exhibition (DATE 2010), pp. 574-579. IEEE (2010).

[24] Suciu, George, Alexandru Vulpe, Simona Halunga, Octavian Fratu, Gyorgy Todoran, and Victor Suciu. Smart cities built on resilient cloud computing and secure internet of things. In 2013 19th international conference on control systems and computer science, pp. 513-518. IEEE (2013).

[25] Garg, Ritu, and Mamta Mittal. Reliability and energy efficient workflow scheduling in cloud environment. Cluster Computing 22, no. 4 (2019): 1283-1297.

[26] Vivek Kumar Prasad and Madhuri Bhavsar, Preserving SLA Parameters for Trusted IaaS Cloud: An Intelligent Monitoring Approach", Recent Patents on Engineering (2019) 13: 1.

[27] Prasad, Vivek Kumar, and Madhuri D. Bhavsar. Monitoring IaaS Cloud for Healthcare Systems: Healthcare Information Management and Cloud Resources Utilization. International Journal of E-Health and Medical Communications (IJEHMC) 11.3 (2020): 54-70.

Edited by: Anand Nayyar
Received: Feb 10, 2020
Accepted: May 31, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 359–368, DOI 10.12694:/scpe.v21i3.1704

COMPARATIVE STUDY OF SPEAKER RECOGNITION TECHNIQUES IN IOT DEVICES FOR TEXT INDEPENDENT NEGATIVE RECOGNITION

NEHA R. KASTURE [*], POOJA JAIN [†], AND TAPAN KUMAR [‡]

Abstract. Speaker recognition (SR) or identification is the subset of broad area of Pattern recognition. Given the features of the voice print, the recognition system identifies the speaker from the knowledge of the speaker models stored in the database. In today's world when many of our works are done through voice, recognition of the speaker is necessary.Recently, SR has also gained importance in Internet of Things (IoT) like setting up of smart environments for home, industries or educational and commercial applications. The race for high accuracy needs making the devices used in these smart environments as close to human hearing capacity as possible. Speaker identification is mostly used to establish negative recognition [1].Negative recognition is when the system decides whether a person is who he disagrees to be thus preventing a person from exploiting multiple identities. Only biometrics will be suitable to establish such identification. The feature extraction of voice sample along with comparative analysis of its methods is of fundamental interest in this paper. We try to compare the performance of features which are used in state of art speaker recognition models and analyse variants of Mel frequency cepstrum coefficients(MFCC) predominantly used in feature extraction which can be further incorporated and used in various smart devices.

Key words: Speaker identification, Pattern recognition, MFCC, Feature extraction, Inverted MFCC, GMM

AMS subject classifications. 68M11

1. Introduction. Speech or voice is a dominant mode of communication in everyday life in many of the Internet of Things (IoT) devices and comprises of unique features relevant to the user. So, although the primary function of a speech is to convey the message, the same speech is also used as bio metric feature to recognize the identity of the person.The last few decades have witnessed speaker recognition technology emerging in various commercial domains like bio-metric, banking applications, indexing or structuring of audio information and Diarization[2]. The emergence of smart devices and home assistants like Google Home or Alexa have brought in ample opportunities for authentication and use of speech samples[3] This biometric characteristic in the person's voice can be used to control the IoT devices. Automatic speaker recognition systems enables to recognize a speaker and hence authenticate it for making any transaction [4]. Speaker recognition systems can be broadly categorized as: speaker identification and speaker verification [5]. Speaker identification attempts to find "who is speaking" from a set of known speakers. This method is also called as Closed set identification because the unknown speaker belongs to the group of speakers in the database whose models are present in advance for matching. In Open set identification problem, the speaker can be an outsider not present in the finite pool of speakers known to the system. Speaker verification differs from recognition in the sense that it confirms if he/she is the authenticated person behind the speech sample.

Furthermore, speaker recognition can also be distinguished as text dependent or text independent. Text dependent speaker identification requires speaker to utter predefined word or phrase from a limited vocabulary, while text independent speaker identification is more flexible and does not restrict user to utter the predefined keyword. The research paper [6] focuses on the scope of text-independent speaker verification using short utterances.

The speaker recognition system operates in two phases. The first is the training phase or enrollment phase where model is created from the speech samples of the different speakers who will need identification. This step is usually completed before the system goes live for voice identification. The second phase is testing phase or

[*] Department of CSE, IIIT Nagpur, India (nehakasture86@gmail.com)
[†] Department of CSE, IIIT, Nagpur, India (pooja.jain@cse.iiitn.ac.in)
[‡] Department of ECE, IIIT Nagpur, India (tapan.jain@ece.iiitn.ac.in)

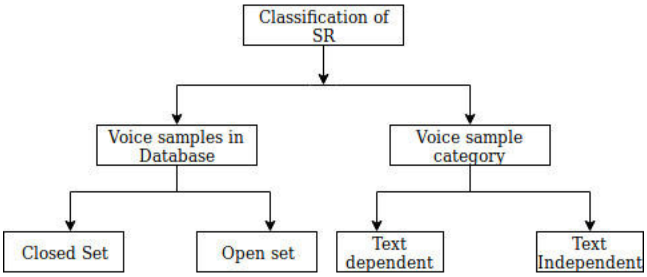


Fig. 1.1. *Broad Classification of Speaker Recognition Systems*

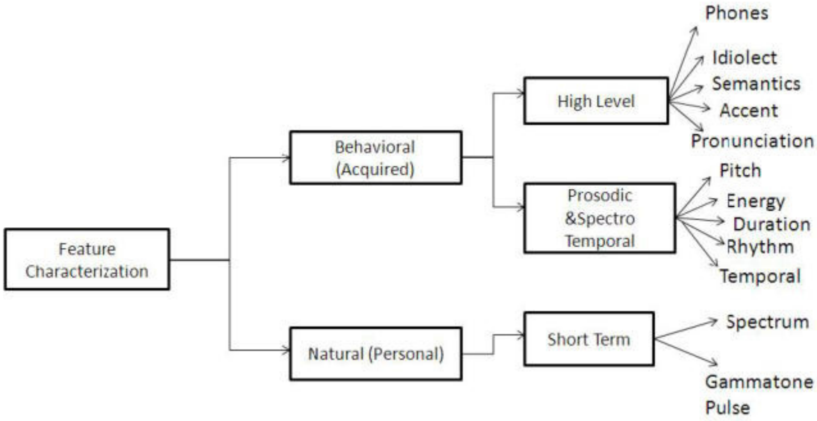


Fig. 1.2. *Broad Classification of Features*

verification phase where signal is matched with the models of speaker available in the database. The implicit assumption here is that each sample belongs to only one speaker. Signals are some quantifiable outputs. Once, the signal is received the fundamental interest lies in extracting a set of feature vectors from speech signals [7]. Feature extraction is one of the preliminary steps of acoustic modelling which quantifies the properties of input speech signal.

In this work we focus on studying and and analysing feature extraction techniques and its effectiveness when working with different Speaker recognition models like Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM). This study aims to choose better combination of feature and model to improve the accuracy of speaker recognition which can be then employed in various scenarios of Human Computer Interaction (HCI).

2. Acoustic Feature Extraction. Feature extraction is crucial to extract characteristics from spoken utterances received from the front end of the model. Features can be broadly classified as (1) short-term spectral (2) voice source (3) spectro-temporal (4) prosodic (5) high-level features based on their physical interpretations [8]. The authors [9] detail the language models for recognition of tamil language. The speech signals were segmented at phonetic levels on the basis of their acoustic characteristics. Spectral analysis determines the frequency content of an arbitrary signal. Spectral features can be obtained by by converting the time based signal into the frequency domain using the Fourier Transform, like: fundamental frequency, frequency components, spectral density, etc. These spectral features can be used to identify characteristics like notes, pitch, rhythm, and melody.

The most popular feature extraction technique used recently in tasks of speaker recognition in different applications is MFCC [10]. But pure MFCC approach modeled on human auditory system is observed to be efficient in non-noisy environments. With the increase in vocabulary or ambient noise the performance of MFCC features seems to decline. Here we analyse the variations of MFCC features which are more robust in nature for the task of SR. See Section 2.

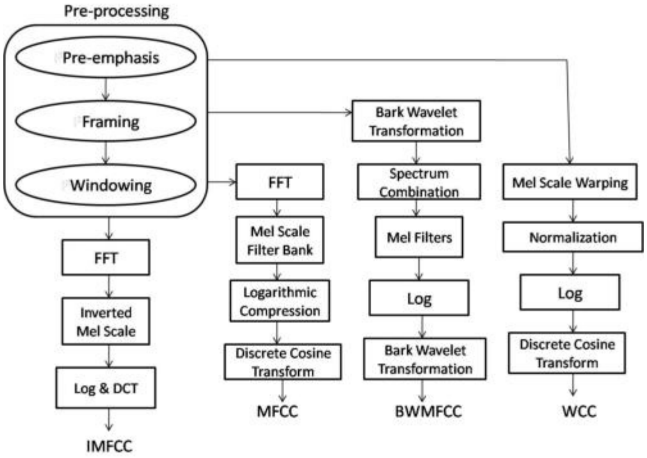


FIG. 2.1. *Pipeline of Feature Extraction*

2.1. Pre-processing. The speech waveform sampled at 8 Khz is used as input for feature extraction. Following steps are common for extracting MFCC, IMFCC and Fused MFCC [11].

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j|y, v, w; \theta) = \frac{\alpha_i(t)a_{ij}^{w_t}\beta_j(t+1)b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i(t)a_{ij}^{w_t}\beta_j(t+1)b_j^{v_{t+1}}(y_{t+1})} \tag{2.1}$$

2.1.1. Pre-emphasis. This step is applied to improve Signal to noise ratio. Higher modulating frequencies are more susceptible to noise than lower ones. Hence, higher frequencies need to be boosted artificially. First order finite impulse response (FIR) filter is applied for spectral flattening as shown in equation 2.2 [12].

$$H(z) = 1 - \alpha z^{-1}, \quad 0.9 \le \alpha \le 1 \tag{2.2}$$

The value of αused for experimentation is usually 0.97.

2.1.2. Framing. Frames should not be too long or too short. Longer frame result in rapid changes in signal properties across the window, thus negatively affecting the time resolution, while too short a frame comes at a cost of affecting the frequency resolution of the signal. So, there always exists a trade-off between time and frequency resolution [13]. So for such non-stationary signals as speech, usually 256 samples in each frame can be chosen with 128 overlapping samples in the adjacent frames with the intention of extracting any vital information occurring at the edges of the frames. In terms of time duration 25ms frame generated every 10 ms with a overlap of 15ms is a popular approach.

2.1.3. Windowing. Distortions in the frame boundaries can give rise to unwanted effects in the frequency response. A window function works with signal in such a way that it smooths the frame at the beginning and end at nearly zero to maintain the continuity [12]. Many window functions like rectangular window, flat top window and hamming window and hanning window are available to implement this step. Out of these possible options Hamming window is the most popular technique used in majority of feature extraction methods as it introduces minimum distortion. The equation 2.3 shows the hamming window function:

$$h[n] = \begin{cases} 0.54 - 0.46\cos\dfrac{2\pi n}{N}, & 0 \le n \le N \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

2.2. MFCC Features. Most of the Speaker Recognition tasks today employ MFCC method for extraction of features [10]. Introduced in early 1980's, these features based on human auditory system [14]are still relevant. MFCC's are representative of vocal tract information. The significant feature of MFCC is its use of perceptually inspired Mel-spaced filter bank processing of the Fourier Transform. Another advantage is flexibility of use achieved through cepstral analysis. Following are the steps for the extraction of MFCC features.

2.2.1. Fourier Transform. Fast Fourier Transformation (FFT) is generally applied on each frame to calculate the components of frequency from the signal in time domain called as spectral values. FFT output is a set of complex numbers containing both real and imaginary part where, real values are dealt with and imaginary part is ignored. In a way, output of FFT and DFT transformation is same the only difference is in terms of computational complexity [15], FFT increases the processing rate of the signal. The following equation shows the DFT for input frame $x(n)$ of 256 samples. 256-point FFT can be used to convert frame of 256 samples into its equivalent DFT.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\prod kn}{N}}, \quad 0 \leq k \leq N-1 \tag{2.4}$$

2.2.2. Mel Scaled Filterbank. The spectrum obtained from the above step contains lot of fluctuations and not the whole spectrum details are useful. Only the envelope of the spectrum is of use here. Hence the spectral envelope is obtained by multiplying the spectrum with Mel scaled filterbank. Each filter in the filterbank is a triangular filter which is uniformly spaced on the Mel frequency axis, having more filters in the low frequency region and less number of filters in the higher frequency region. Mel Frequency analysis is very close to how humans perceive sounds. Also, it is proved by experimentation that sensitivity of human ears is more towards low frequency than high frequency. The voice utterance does not follow linear scale frequency which is used in FFT hence, Mel scale is used which is linear upto 1kHz and logarithmic at higher frequencies. The equation sated below shows the relation between Linear scale and mel scale frequency.

$$Mel(f) = \log_{10}\left(1 + \frac{f}{700}\right) \tag{2.5}$$

2.2.3. Logarithmic Compression. This step aims to take log of spectral envelope obtained from the step above, since human ears cannot hear sounds in linear scale. Each co-efficient of envelope is multiplied by 20 to get the spectral envelope in dB.

2.2.4. Discrete Cosine Transform. This final step ensures conversion of log Mel spectrum into its spatial domain. This is achieved by taking Discrete Cosine Transform (DCT) which divides a finite sequence into discrete vector.Thus, DCT yields cepstral coefficients [16]as follows:

$$c_n = \sum_{k=1}^{K} S_k \cos\left[n(k - \frac{1}{2})\frac{\prod}{k}\right], n = 1, 2....L \tag{2.6}$$

where K is the number of log-spectral coefficients calculated in previous step, S_k are the log-spectral coefficients, and L is required number of cepstral coefficients that we want. The MFCC feature is finally achieved from lowest 12-15 DCT coefficients.

2.3. Bark Wavelet MFCC Feature. The DCT and FFT algorithms used in MFCC feature extraction do not prove to be a good option if the signal to be processed is non-stationary. The bark wavelet feature introduced by [17] proves as an anti-noisy feature that can substitute MFCC and overcome the disadvantages of fixed time-frequency resolution of DCT. Humans perception of speech is non linear if actual frequency is used but linear if Bark frequency is used. The relationship between linear frequency and Bark frequency can be represented as shown below:

$$b = 13 \cdot \arctan(0.76f) + 3.5 \cdot \arctan\left(\frac{f}{7.5}\right)^2 \tag{2.7}$$

where b represents bark frequency and f represents linear frequency. Following steps comprise of the general philosophy behind bark wavelet:

- Gaussian function is chosen as mother function of Bark wavelet to satisfy time and bandwidth product least.
- To maintain consistency with the frequency group,mother wavelet is chosen to have the equal bandwidth in the Bark domain.
- Unit bandwidth of 1 Bark keeps the consistency with the frequency group.

2.3.1. Pre-processing. The Pre-processing stage consisting of Pre-emphasis, Framing and Windowing is same as that of extracting MFCC features.

2.3.2. Bark Wavelet Transformation. Bark Wavelet Transformation can be performed by using the following equation on every frame:

$$s_k(n) = \sum_{l=0}^{N-1} S(l) W_k(l) e^{\frac{j2\Pi nl}{N}} \tag{2.8}$$

where N is the number of zeros in FFT, S(l) is the frequency spectrum of Speech signal, $s_k(n)$ is the speech spectrum of the k^{th} sub-band and $W_k(l)$ is a discrete form of $W_k(f)$ which is expressed as follows:

$$W_k(f) = c_2 2^{-4[13\arctan(0.76f + 3.5\arctan(\frac{f}{7.5})^2 - (b_1 + k\Delta b)]^2} \tag{2.9}$$

where normalization factor c_2 can be calculated as

$$c_2 \sum_{k=0}^{K-1} W_k(b) = 1, \quad 0 < b_l \leq b \leq b_h \tag{2.10}$$

where $[b_l, b_h]$ is the Bark frequency bandwidth.

2.3.3. Spectrum Combination. Frequency Synthesis is obtained using the equation 2.11

$$s(n) = \sum_{k=0}^{K-1} s_k(n) \tag{2.11}$$

where $s(n)$ is the frequency synthesis spectrum.

2.3.4. Mel Filters. Signal $s(n)$ is passed through mel filters to reduce the effect of tone and pitch in the feature co-efficients and emphasize the original formant of speech.

2.3.5. Logarithm. Here, log of spectrum obtained through mel filters is taken as follows:

$$d(m) = \log(\sum_{n=0}^{N-1} |s(n)|^2 H_m(n), \quad 0 \leq m < L \tag{2.12}$$

where $H_m(n)$ is triangular mel frequency band pass filters, L is number of filters and N is sample number $s(n)$.

2.3.6. Bark Wavelet MFCC features. Finally the Bark Wavelet MFCC features (BWMFCC) is obtained by performing Bark Wavelet Transform on $d(m)$ as follows:

$$BWMFCC(n) = \sum_{m=0}^{L-1} |s(n)|^2 W_n(m) \cdot d(m), \quad 0 \leq m \leq L-1, \quad 0 \leq n < M \tag{2.13}$$

2.4. Wavelet Cepstral Coefficient. Another short term feature vector that is effective at keeping the effects of noise at bay is Wavelet Cepstral Coefficient (WCC) that uses Discrete wavelet transform (DWT). The detailed guidelines for DWT implementation is mentioned in [18]. A typical wavelet transform can be given as:

$$W_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\Psi\left(t - \frac{b}{a}\right) dt \tag{2.14}$$

where the function $\Psi(t)$ is a mother wavelet, a is scaling factor and b is translation parameter. DWT obtains the spectrum using multilabel resolution technique. In comparison with FFT used in MFCC, DWT distributes the signal into smaller frequency domains to obtain the local frequency spectrum. The advantage of such decomposition is, if the parts of signal is distorted by noise, the whole frequency spectrum won't be affected much. Thus making DWT more robust towards noise. Wavelet packet transform (WPT) offers a flexible multi-resolution approach which can vary the window length to suit better time or frequency resolution. This results in better time frequency characteristics of WPT but at the cost of computational overhead. The traditional WPT does not warp frequencies as per human auditory perception system. Therefore the work proposed by [19] has combined the advantages of multi-resolution WPT and Mel scale to give Wavelet Packet Based Mel Frequency Cepstral Features.

2.4.1. Pre-processing. The speech is initially sampled and subjected to common set of pre-processing steps as mentioned earlier consisting of Pre-emphasis, Framing and Windowing.

2.4.2. Mel scale warping. Mel scale warping consists of 3 sub-steps:

- Fast Fourier Transform (FFT) is used to transform the pe-processed signal from time domain to frequency domain.
- The frequency spectrum obtained through FFT is Mel-warped using triangular mel filter banks.
- The signal is again converted to time domain by Inverse FFT to carry out the further processing.

2.4.3. Wavelet packet decomposition. The speech signal is decomposed at depth 7 (level 7), with Daubechies type (db4) wavelet. The resultant wavelet consists of maximum frequency of 31.25 Hz producing 128 sub-bands.

2.4.4. Best basis formulation. 35 sub-bands out of 128 total frequency sub-bands are selected for further processing since higher frequency coefficient represents maximum amount of energy. The sub-band signal energies in each frame can be computed as

$$E_j = \frac{\sum_{j=1}^{N_j}[W_j^p f(i)]^2}{N_j}, \quad j = 1...35 \tag{2.15}$$

2.4.5. Log and DCT. Finally, The logarithmic compression is performed and DCT is taken to reduce the dimension of sub-band energies.

2.5. Inverted MFCC. MFCC effectively captures the low frequency region than high frequency region. Hence it is capable of extracting the formants[20] lying in the lower range of frequency. But this extraction of formants in lower frequency range neglects the formants if any lying in higher range of frequency. This essentially happens because of filter bank structure [21] where higher number of closely spaced overlapping triangular filters appear in lower frequency region of Mel filter bank and less number of overlapping triangular filters in higher frequency area. The approach of [22] is based on reversing the the normal MFCC filter bank structure to capture the characteristics in higher frequencies missed out by MFCC. This feature is called as Inverted MFCC (IMFCC). The initial steps of Pre-processing and FFT are common in this appraoch. The variation in implementation lies in the complementary way in which filter bank is used.

2.5.1. Inverted Mel Scale. The complementary filter bank structure is obtained by reversing the original from the general mid point of frequency range i. e., 0-4kHz in speaker recognition applications. Thus, reversing is done at 2kHz point of original filter bank. Mathematical expression for i^{th} filterbank of the same can be given as:

$$\widehat{\Psi}_i(k) = \Psi_{Q+1-i}\left(\frac{M_s}{2} + 1 - k\right) \tag{2.16}$$

where $\widehat{\Psi}(k)$ is the response of inverted Mel scale filter, $\Psi(k)$ is the original Mel scale filter response, Q is the number of filters, $(1 \leq i \leq Q)$ M_s is the number of points in DFT and $(1 \leq k \leq M_s)$ The relationship between inverted mel scale and original can be expessed as:

$$\widehat{f_{mel}}(f) = f_{mel}(f_{high}) + f_{mel}(f_{low}) - f_{mel}\left[\frac{F_s}{2} + \frac{F_s}{M_s} - f\right] \tag{2.17}$$

where $f_{mel}(f)$ is the relative pitch in the inverted scale corresponding to f, the actual frequency in Hz. Also to maintain uniformity in DFT calculation Inverted Mel Scale is made to have common boundary points as with the actual Mel Scale such that $\widehat{f_{mel}}(f_{low}) = f_{mel}(f_{low})$ and $\widehat{f_{mel}}(f_{high}) = f_{mel}(f_{high})$. This flipping in the Mel scale gives fine represenatation of high frequency regions not otherwise not justified by MFCC Mel Scale. Filter outputs $\{\hat{e}(i)\}_{i=1}^{Q}$ are computed from energy spectrum $|Y(k)|^2$ as

$$\hat{e}(i) = \sum_{k=1}^{M_s/2} |Y(k)|^2 \cdot \widehat{\Psi}_i(k) \tag{2.18}$$

2.5.2. Log and DCT. Logarithm of filter bank energies is taken as:

$$\{\log_{10}[\hat{e}(i)]\}_{i=1}^{Q} \tag{2.19}$$

The last step is to obtain the inverted MFCC coefficients by taking the DCT of log energies obtained in 2.19 as shown below:

$$\hat{C_m} = \sqrt{\frac{2}{Q}} \sum_{l=0}^{Q-1} \log[\hat{e}(i+1)] \cdot \cos\left[m \cdot \left(\frac{2l-1}{2}\right) \cdot \frac{\Pi}{Q}\right] \tag{2.20}$$

Usually with MFCC features we choose the first 19 coefficients as features to model the speaker but in case of IMFCC we choose the last 19 coefficients to model the speakers.

3. Speaker Modeling: Gaussian Mixture Model. The basic purpose of this step is building a model for any speaker 's' such that 'x' feature vector extracted from the utterance of speaker 's' can be represented by a unique model. Thus, matching an unknown voice sample with the speaker model can result in recognition of the correct speaker. One of the most universal modeling framework used for SR task is Gaussian Mixture Model (GMM)[23]. GMM's are very popular in text independent SR applications where the speaker is not restricted to use any pre-defined phrase as a voice sample. Gaussian distribution is particularly identified as a mean and a deviation about the mean. For a D-dimensional feature vector x, $\{\overrightarrow{x_t} \in \mathbb{R}^{\mathbb{D}} : 1 \leq t \leq T\}$, the mixture density used for the likelihood function is defined as [24]:

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i p_i(x) \tag{3.1}$$

where GMM is denoted by λ, M is is the number of Gaussian components, w_i is the prior probability or mixing weight of the i^{th} Gaussian component constrained to $\sum_{i=1}^{M} w_i = 1$, and $p_i(x)$ is given by

$$p_i(x) = \frac{1}{(2\Pi)^{\frac{D}{2}} \mid \sum_i \mid^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu_i)'(\sum_i)^{-1}(x-\mu_i)\right\} \tag{3.2}$$

where $p_i(x)$ is is the D-variate Gaussian density function with mean vector μ_i and covariance matrix $\sum_i$. Collectively the GMM model is denoted as $\lambda = \{w_i, \mu_i, \sum_i\}$ where $i = 1 \ldots M$. The average log-likelihood of feature vector X with respect to model λ is defined as,

$$LL_{avg}(X \mid \lambda) = \frac{1}{T} \sum_{t=1}^{T} log \sum_{k=1}^{K} p(x_t \mid \lambda) \tag{3.3}$$

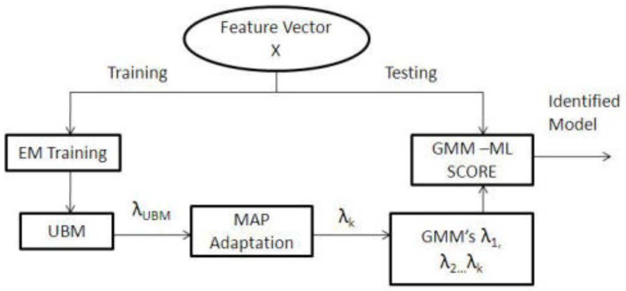


Fig. 3.1. *Training and Testing using GMM-UBM*

where $p(x_t \mid \lambda)$ is calculated as shown in equation 3.1. It has been empirically observed that diagonal matrix GMMs are better in performance and also computationally more efficient than full matrix GMMs. Estimating the parameters of a full-covariance GMM is very expensive [25]. Hence, diagonal covariance matrices are usually used. Maximum Likelihood (ML) estimation is used in training the GMM to estimate the parameter λ $= \{w_i, \mu_i, \sum_i\}$ where $i = 1 \dots M$ for a feature vector X.

Once the training vectors are ready, the iterative expectation–maximization (EM) algorithm [26] is used to maximize the likelihood with respect to the training data [27]. GMM parameters are refined with each iteration of EM algorithm to increase the likelihood of the estimated model for the observed feature data.K-Means can provide for the initialization of EM algorithm [28]. In general, five iterations are considered to be enough for parameter convergence. In applications pertaining to SR a model should adapt well with different types of speakers, their environments, speaking styles etc. Hence in GMM based SR a speaker-independent universal background model (UBM) is created. This UBM is trained with EM algorithm from hundreds of hours of speech data gathered from a large number of speakers. When a new speaker is enrolled into the system, the UBM adapts its parameters to the feature distribution of newly enrolled speaker. This adapted model is used as a model representing that speaker. Thus, prior knowledge is utilized for estimating model parameters. The maximum a posteriori (MAP) method [29] is used to extract speaker-specific GMM from the UBM. In the Testing phase, the MAP-adapted model and the UBM are combined, and the recognizer is called as Gaussian mixture model - universal background model, or “GMM-UBM”. The test features received from the voice sample are compared with the speaker models available in the database and the model with highest log likelihood ratio (LLR) is chosen:

$$LLR_{avg}(X, \lambda_{target}, \lambda_{UBM}) = \frac{1}{T}\sum_{t=1}^{T}\{\log p(x_t \mid \lambda_{target}) - \log p(x_t \mid \lambda_{UBM})\} \tag{3.4}$$

4. Discussion and conclusions. In this work we reviewed two architectures of SR namely GMM and HMM when working with MFCC features and its variants. Its comparative performance is listed in the above table. This study focuses on the performance analysis of MFCC and its variants discussed in the literature. Table 4.1 shows the recognition rate using the discussed feature extraction techniques like MFCC, IMFCC, BWMFCC and WCC as experimented by various researchers. The database used by the authors include TIMIT [31], YOHO [32], VoxForge [33] and also author created manual database. Comparison of MFCC and IMFCC features can be seen on YOHO database where MFCC outperforms IMFCC. IMFCC supports the extraction of information lying in the higher frequency range which is not considered by MFCC. The experimentation done by [34] shows that the results improve when fusion of MFCC and IMFCC is taken. WCC is tested on comparatively smaller corpus of 30 speakers but shows to provide better time and frequency resolution for limited data than MFCC. Bark wavelet based MFCC can be used as a good anti-noise substitute since it overcomes the disadvantage of fixed time-frequency resolution. The robustness of this feature is also demonstrated in detail by introducing the noise component as shown in the work of [30].The advantages and disadvantages of using any of the features mentioned above can be tabulated as shown in Table 4.2. Variations over pure MFCC can improve the system performance if used in applications where robustness is required where humans and smart assistants are involved.

TABLE 4.1
Comparison between MFCC feature variants

Referred System	Database used	Feature Used	Dimension	Modeling Technique	Accuracy achieved (%)
[22]	YOHO	IMFCC	138 speakers GMM Mixing Coefficient=32	GMM	95.23
[22]	YOHO	MFCC	138 speakers GMM Mixing Coefficient=32	GMM	96.82
[19]	VoxForge	WCC	30 speakers GMM Mixing Coefficient=15	GMM	100
[19]	VoxForge	MFCC	30 speakers GMM Mixing Coefficient=15	GMM	93.33
[30]	Author created Word Pronunciation	BWMFCC	16 speakers SNR=Clean Words=30	HMM	95.69
[30]	Author created Word Pronunciation	MFCC	16 speakers SNR=Clean Words=30	HMM	93.74

TABLE 4.2
Advantages and Disadvantages of Features

Feature	Advantages	Disadvantages
MFCC	Good choice for clean speech, Represents human auditory system, Easy and relatively fast to compute	Unsuitable in noisy conditions, performance degrades with larger vocabulary, Only low frequencies are considered and high frequencies are ignored
BWMFCC	Robust to noise, Suitable for larger vocabulary	Works for low signal to noise ratios, Complex due to additional Bark wavelet transformation
IMFCC	Capable of representing information in high frequency region, less computation burden as compared to other variants	Gives better results when fused with MFCC than individual IMFCC
WCC	Frequency spectrum obtained through wavelet is noise-resistant, good time and frequency resolution	To find the optimum mother wavelet, time consuming

REFERENCES

[1] JAMES WAYMAN. Fundamentals of biometric authentication technologies. *Int. J. Image Graphics*, 1:93–113, 01 2001.

[2] HOMAYOON BEIGI. *Fundamentals of Speaker Recognition.* Springer Publishing Company, Incorporated, 2011.

[3] ZHANIBEK KOZHIRBAYEV, BERAT EROL, ALTYNBEK SHARIPBAY, AND MO JAMSHIDI. Speaker recognition for robotic control via an iot device. pages 1–5, 06 2018.

[4] JYOTI SINGHAI AND RAKESH SINGHAI. Automatic speaker recognition: An approach using dwt based featureextraction and vector quantization. *IETE Technical Review*, 24(5):395–402, 2007.

[5] R. TOGNERI AND D. PULLELLA. An overview of speaker identification: Accuracy and robustness issues. *IEEE Circuits and Systems Magazine*, 11(2):23–61, Secondquarter 2011.

[6] ROHAN KUMAR DAS AND SR MAHADEVA PRASANNA. Speaker verification from short utterance perspective: a review. *IETE Technical Review*, 35(6):599–617, 2018.

[7] FRÉDÉRIC BIMBOT, JEAN-FRANÇOIS BONASTRE, CORINNE FREDOUILLE, GUILLAUME GRAVIER, IVAN MAGRIN-CHAGNOLLEAU, SYLVAIN MEIGNIER, TEVA MERLIN, JAVIER ORTEGA-GARCÍA, DIJANA PETROVSKA-DELACRÉTAZ, AND DOUGLAS A. REYNOLDS. A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, 2004(4):101962, Apr 2004.

[8] TOMI KINNUNEN AND HAIZHOU LI. An overview of text-independent speaker recognition: from features to supervectors. *Speech Communication*, 52:12–40, 01 2010.

[9] SS SARASWATHI AND TVG GEETHA. Language models for tamil speech recognition system. *IETE Technical Review*, 24(5):375–383, 2007.

[10] S. MOLAU, M. PITZ, R. SCHLUTER, AND H. NEY. Computing mel-frequency cepstral coefficients on the power spectrum. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, volume 1, pages 73–76 vol.1, May 2001.

[11] SIDDHANT C. JOSHI AND DR. A. N. CHEERAN. Matlab based feature extraction using mel frequency cepstrum coefficients for automatic speech recognition.

[12] YUAN MENG. Speech recognition on dsp: Algorithm optimization and performance analysis. 06 2019.

[13] LINDASALWA MUDA, MUMTAJ BEGAM, AND I. ELAMVAZUTHI. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *CoRR*, abs/1003.4083, 2010.

[14] BEN GOLD AND NELSON MORGAN. *Speech and Audio Signal Processing.* 01 1999.

[15] SHIKHA GUPTA, JAFREEZAL JAAFAR, WAN FATIMAH, AND ARPIT BANSAL. Feature extraction using mfcc. 2013.

[16] BRUCE P. BOGERT. The quefrency analysis of time series for echoes : cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. 1963.

[17] X. Zhang, J. Bai, and W. Liang. The speech recognition system based on bark wavelet mfcc. In *2006 8th international Conference on Signal Processing*, volume 1, Nov 2006.

[18] Y. Zhao, L. Zhang, J. Hu, and T. Liao. Mallat wavelet filter coefficient calculation. In *2013 International Conference on Computational and Information Sciences*, pages 963–965, June 2013.

[19] Smriti Srivastava, Saurabh Bhardwaj, Abhishek Bhandari, Krit Gupta, Hitesh Bahl, and J R. P. Gupta. *Wavelet Packet Based Mel Frequency Cepstral Features for Text Independent Speaker Identification*, pages 237–247. 01 2013.

[20] Ursula G. Goldstein. Speaker-identifying features based on formant tracks. *The Journal of the Acoustical Society of America*, 59:176–82, 02 1976.

[21] Fang Zheng, Guoliang Zhang, and Zhanjiang Song. Comparison of different implementations of mfcc. *J. Comput. Sci. Technol.*, 16:582–589, 11 2001.

[22] S Chakroborty, A Roy, and Goutam Saha. Improved closed set text-independent speaker identification by combining mfcc with evidence from flipped filter banks. *International Journal of Signal Processing*, 4:114–122, 01 2007.

[23] D. A. Reynolds and R. C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, Jan 1995.

[24] Douglas Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10:19–41, 01 2000.

[25] Kuo-Hwei Yuo and Hsiao-Chuan Wang. Joint estimation of feature transformation parameters and gaussian mixture model for speaker identification. *Speech Communication*, 28:227–241, 07 1999.

[26] Arthur Dempster, Natalie Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38, 01 1977.

[27] J.A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4:126, 1998.

[28] Yoseph Linde, Andrés Buzo, and Robert Gray. An algorithm for vector quantizer design. *Communications, IEEE Transactions on*, 28:84–95, 01 1980.

[29] Jean-Luc Gauvain and Chin-Hui Lee. Lee, c.: Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. ieee trans. speech audio process. 2, 291-298. *Speech and Audio Processing, IEEE Transactions on*, 2:291 – 298, 05 1994.

[30] Zhang Jie, Li Guo-liang, Zheng Yu-zheng, and Liu Xiao-ying. A novel noise-robust speech recognition system based on adaptively enhanced bark wavelet mfcc. volume 4, pages 443 – 447, 09 2009.

[31] J S Garofolo, Lori Lamel, W M Fisher, Jonathan Fiscus, D S. Pallett, N L. Dahlgren, and V Zue. Timit acoustic-phonetic continuous speech corpus. *Linguistic Data Consortium*, 11 1992.

[32] Johan Koolwaaij. Speaker identification and assessment on the yoho database. 01 1997.

[33] Voxforge.org. Free speech... recognition (linux, windows and mac) - voxforge.org. `http://www.voxforge.org/`. accessed 06/25/2014.

[34] Sandipan Chakroborty and Goutam Saha. Improved text-independent speaker identification using fused mfcc & imfcc feature sets based on gaussian filter. *Signal Processing*, 35, 11 2009.

Edited by: Anand Nayyar
Received: Mar 3, 2020
Accepted: May 25, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 369–378, DOI 10.12694:/scpe.v21i3.1709

EVENT DRIVEN RECOMMENDATION SYSTEM FOR E-COMMERCE USING KNOWLEDGE BASED COLLABORATIVE FILTERING TECHNIQUE

MAHESH KUMAR SINGH*AND OM PRAKASH RISHI†

Abstract. The Internet is changing the method of selling and purchasing items. Nowadays online trading replaces offline trading. The items offered by the online system can influence the nature of buying customers. The recommendation system is one of the basic tools to provide such an environment. Several techniques are used to design and implement the recommendation system. Every recommendation system passes from two phases similarity computation among the users or items and correlation between target user and items. Collaborative filtering is a common technique used for designing such a system. The proposed system uses a knowledge base generated from knowledge graph to identify the domain knowledge of users, items, and relationships among these, knowledge graph is a labelled multidimensional directed graph that represents the relationship among the users and the items. Almost every existing recommendation system is based on one of feature, review, rating, and popularity of the items in which users' involvement is very less or none. The proposed approach uses about 100 percent of users' participation in the form of activities during navigation of the web site. Thus, the system expects under the users' interest that is beneficial for both seller and buyer. The proposed system relates the category of items, not just specific items that may be interested in the users. We see the effectiveness of this approach in comparison with baseline methods in the area of recommendation system using three parameters precision, recall, and NDCG through online and offline evaluation studies with user data, and its performance is better than all other baseline systems in all aspects.

Key words: Recommendation System (RS), Collaborative Filtering (CF), Knowledgebase, Knowledge graph. e-Commerce, Subject predicate object (SPO), K- Nearest Neighborhood (KNN), Normalized Discounted Cumulative Gain (NDCG), Bayesian Personalized Ranking(BPR), Visual Bayesian Personalized Ranking (VBPR), Deep Convolutional Neural Network (DCNN), Joint Representation Learning (JRL).

AMS subject classifications. 68M11

1. Introduction. This is the era of I-way. With the development of high-speed Internet, smart computing technology, and artificial intelligence, e-Commerce is growing very rapidly. It tries to replace traditional selling and purchasing methods. It provides more choices to the user for purchasing items in easy and efficient ways. Many e-Commerce companies are available through the world like Amazon, Tokopedia, Flipcart, Bukalap, Netflix, Snapdeal, and so on are using the recommendation system to help their users. The choices provided by these e-Commerce companies influence the user's decision in buying the products. Since the number of users increases exponentially, hence it is the challenge of e-Commerce companies to handle those. Due to multiples choices of similar types of products user generally confused and irritated, to overcome these problems, companies are trying to develop or upgrade a system to guide the users in very efficient ways. The recommendation system is a technique by which the system suggests the relevant products to the customers in reputed E-commerce [1] companies, where there is a large verity of products are available. The recommendation system can be personalized or non-personalized. The non-personalized system commonly used in a physical store that processed the common features of the products like the popularity of the product for example actors in a movie. On the other hand, the recommendation system that uses the navigation details and the activities in the form of rating/review of the products given by the customers is called a personalized [2] recommendation system. The product can be recommended to a customer if the calculated preference score is high. If it is to consider that each product or service has a different page then page rank provides the rating of the similar product of different brand names by using page ranking concept, we can calculate the popularity of the product or service. Nowadays every E-commerce site wants to use the recommendation system that is used to serve the millions

* Department of Computer Science and Informatics. University of Kota, Rajasthan, India. (maheshkrsg@gmail.com).

† Department of Computer Science and Informatics. University of Kota, Rajasthan, India. (omprakashrishi@yahoo.com).

Table 1.1
List of some popular websites that are using recommendation system to facilitate their users.

Sr. No.	Name of site	User	Item	Description
1	Linkedin	Member	Members or Jobs	Members are interested in other members or jobs
2	Facebook	Member	Members	Members are interested in another member
3	Amazon	Member	Products or books	Members are interested in products
4	Netflix	Member	Movies or Story series	Members are interested in watching movies or story series.
5	Flipcart	Member	Products	Members are interested in products.

of customers. It increases the customers' faithfulness and directs them to new items in the product catalog. Very popular E-commerce site Amazon attracts large community using user-centric recommendation system by accessing the user history data, while small scale company faces the problems extensive shopping history record since many customers are one-time visitors. Some popular sites where the recommendation is working based on user and item as shown in Table 1.1.

Recommendation system [3] processes user's information stored in the system during navigation of the e-Commerce site and provide the most relevant product to the user and solve his/her problems. Let U be the set of users and I be the set of items, then prediction function P is defined as $P : UXI- > R$ where R be the list of items for a specific user. There are so many techniques that are used in designing recommendation systems, including popularity, based, content-based filtering [4], collaborative filtering [5], graph-based filtering [6], and demographic filtering [7]. In this study, we use a knowledge-based recommendation collaborative filtering technique that processes the personalized heterogeneous information in the form of triplets (u,r,p) that are generated when user u interact with the product p with r activity, stored in the system.

Whole paper is divided into four sections. The first section contains a literature survey that discusses the latest research paper in this area, motivation, and objective of this paper. Second section deals model or method which is used in this research work. The third section is the result and discussion section that deals with model implementation and comparison with baseline systems. The fourth section includes the conclusion and future scope of the paper.

2. Related Work. The recommendation system tries to process the historical data of users or features of items and provides the best relevant result for a particular user. It is used in almost various fields to help users to decide like a restaurant recommendation system provides certain preferences [8] of restaurants, book recommendations [9] provides a list of most preferred books, social network [10] most relevant friends, movie recommendation [11] and widely used product recommendation [12] in e-commerce recommendation system to recommend the item to the user based on ranking. The recommendation for publication [13, 14] that help authors decide where they should submit their manuscripts in Content-based Journals and Conferences. This system recommends suitable journals or conferences with a priority order based on the abstract of a manuscript. Proposed a three-step hybrid ranking [15] order system for finding the top-N list for the multi-criteria recommendation system this approach is good for both the traditional no-ranking item-based collaborative recommendation and single-criteria-ranking approach that uses two popular learning-to-rank methods. Recommender System on computer science. The recommendation system techniques are mainly classified [16] into two categories personalized and non-personalized. Since a recommendation system is created for any customer for his/her likes and dislikes, hence non-personalized recommendation system is negligible, there is no or very limited scope for the researchers. Personalized recommendation systems [17] are used user's profile that is user-based and provides lists of different items while non- personalized recommendation system uses the popularity of the items and provides a similar list for all users.

Several techniques are used in the development and design of a recommendation system which includes, collaborative filtering, content-based filtering, and hybrid filtering technique (see fig 2.1). Collaborative filtering is a widely used tool in the development of a recommendation system based on user's action and profiling to generate similarities in data, there are two types of collaborative filters, memory-based and model-based. The similarity in collaborative filtering is computed in two different ways user-user similarity and item-item similarity using classifiers [18] and clustering algorithms. Content-based filtering [19] technique used features of items to find similarity, hybrid filtering technique uses both features content and collaborative. Similarity descriptions

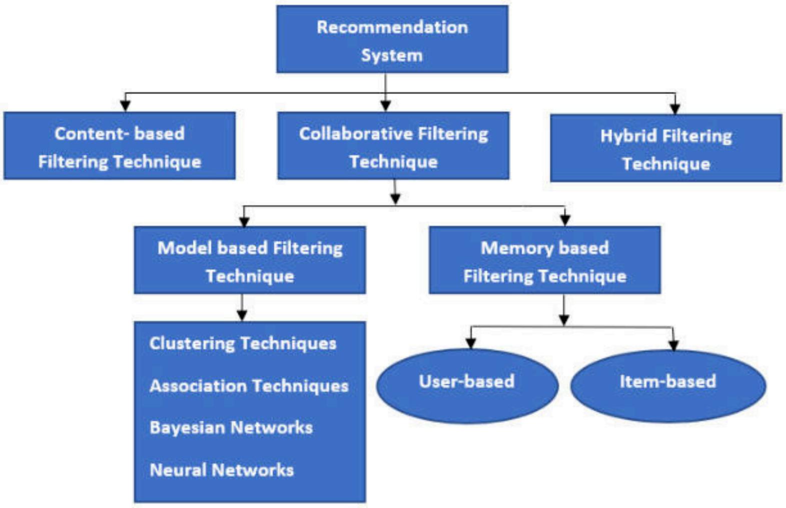

Recommendation System
Content- based Filtering Technique
Collaborative Filtering Technique
Hybrid Filtering Technique
Model based Filtering Technique
Memory based Filtering Technique
Clustering Techniques
Association Techniques
Bayesian Networks
Neural Networks
User-based
Item-based

Fig. 2.1. *Classification of Recommendation Systems*

and factors can influence the decision of product purchasing. Most of the recommendation system techniques are used homogeneous data in the form of a review or rating [18, 20] given by the users to the items, but there are very few users that are ready to rate or review the items while they are bought and used the items. Hence, a recommendation system based on review and ratings [18] is not producing an accurate list of items as he/she required. Image-based recommendation [21] used the appearance of the items to find the similarity. One more problem of the recommendation system is that product is bought or returned by the user.

2.1. Motivation. Artificial intelligence, machine learning, and deep learning are the current growing technologies. There are so many opportunities in these areas for the researchers. E-commerce is rapidly growing technology nowadays due to the availability of high-performance computing devices, and efficient systems. Information overload is one of the main problems for both enterprise as well as users, but recommendation system (RS) play a vital role to minimize it. Almost every E-Commerce company is using it, so efficient recommendation system design and development is one of the key features for researchers nowadays. There are two approaches used to design and implementation of RS, classical programming approach, and machine learning approach. The classical approach is used for small data set while the machine learning approach is used in a large data set. Here we use the classical programming approach. There are so many RS based on popularity, rating, and review but, we know that only 20 to 30 percent of users are participating in review and rating.

The proposed system uses maximum user involvement by the navigation of web site in the form of product view, selection, and purchasing, these heterogeneous data play a vital role in creating a recommendation system.

3. Methodology. The methodology used in the proposed system is shown in figure 3.1. Data is the backbone of any predictive system. The primary aim of any system is the collection of quality data. Data can be collected in two ways, implicit and explicit. Explicit are those data which are provided by the user required by e-Commerce web sites in the form of registration, reviews comments, and rating, etc. Implicit data can be stored by e-Commerce sites automatically using cookies or any other methods implemented by the system, that contains the user's navigation details and activities performed by the user during surfing the e-Commerce website. Knowledge base [22] is a type of implicit data that is collected by the e-Commerce website. Every web site is a collection of web pages connected and formed liked a directed graph, each page is reserve for each product that shows the description of that product, types of links describe the types of activities. Domain knowledge monitors the activities performed by users like purchase transactions, clicks, etc.

The knowledge base is created with the help of a knowledge graph which is, a directed graph that shows relationships among items-users and items-items. In the given data set, set of items and set of users denoted

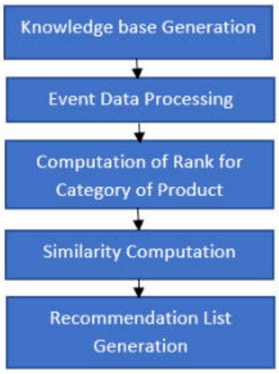

Knowledge base Generation
Event Data Processing
Computation of Rank for Category of Product
Similarity Computation
Recommendation List Generation

Fig. 3.1. *Methodology*

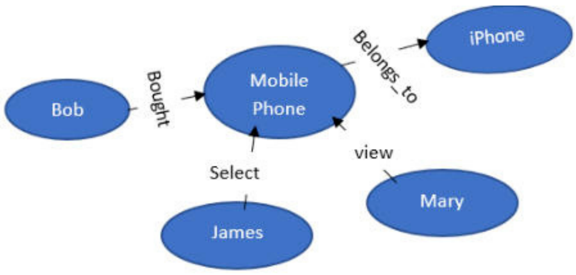


Fig. 3.2. *Knowledge Graph*

by vertices and set of relationships by edges.

Suppose there is a sequence of events as Bob bought a mobile phone of the brand iPhone, James selects the same mobile and Mary only views that mobile phone then this information can be represented by a knowledge graph. The facts of the knowledge base for these events are written as Buys(Bob, Mobile_Phone), belongs_to(Mobile_Phone, iPhone), Selects(James, Mobile_Phone), Views(Mary, Mobile_Phone), as shown in figure 3.2.

Event Data Preprocessing is a module that processes the knowledge base as per the requirement of the system. The event attribute contains three types of entries view, select, and buy. Transform data as separate columns as buy, select and view as the tuple (buy, select, view), if any user buys then it is considered as (1,1,1), select (0,1,1), and view as (0,0,1) respectively. A user i can interact with item j in many ways like view, select, buy, also_buy, etc., only three events are considered for the computation of user interaction score or preference rank of the item. The weight of events as $buy > select > view$ is considered.

The calculation of preference score or rank of each class of items is formulated with the equation (3.1):

$$C_{up} = X\frac{C^v_{up}}{\sqrt{\sum_{p=1}^{n}(C^v_{up})^2}} + Y\frac{C^s_{up}}{\sqrt{\sum_{p=1}^{n}(C^s_{up})^2}} + Z\frac{C^b_{up}}{\sqrt{\sum_{p=1}^{n}(C^b_{up})^2}} \tag{3.1}$$

where C^v_{up} denotes user u viewed item/product p, C^s_{up} user u select item/product p and C^b_{up} user u bought item/product p, X, Y and Z are weight adjusting constants, for better result $X = 0.25, Y = 0.5$ and $Z = 1.0$.

Algorithm:

Input: Entity set E (Set of products P and Set of users U), Relation Set R, triplet set (u,r,p) is in S (ordered set).

Output: List top N products recommendation to the corresponding user as per user's information.

Begin

Step1: Let N is the number product classes and M are the number of users compute preference score of each product class by formula (3.1).
Step2: Calculate similarity of the users by the formula (4.1).
Step3: Calculate prediction score Pre(u,p) of product p with respect to user u using formula (4.2).
Step4: Select top N class of products based for user u on the basis of value of Pre(u,p) with preference score .
Step4: If any product is already bought by the user then replace it with other product and form the list of recommendation.
Step5: Prepare most recommended products for enterprise. .
End

4. Experiments and Results. In our study, we develop our method working with retail e-Shop that sells mobile phones and home appliances only electronics items, about 200 users visit daily but 50 percent of them only visit or view the items, hence user-item bought or select data are very sparse that are ignored. Only 50 percentage data are considered in which at least one event must be bought or select that show their interest. Consider a data set that stores events of users, contains 20680 events (i.e. Buy, select, view) performed by 1240 users with unique user_IDs on 114 class products with unique product_IDs with the multiple numbers of products, of two categories (i.e. mobile phones and electronics items) and brands (i.e. Nokia, iPhone, Sony, Vivo and RedMe, Tata). The event data set contains five attributes User_ID, event, product_ID, Brand, and Category as shown in table 4.1. We use Jupiter notebook for providing coding environment,python 3.0 for programming, pandas [23] python library for data wrangling or cleansing and analysis, matplotlib for visualization, NumPy for numerical computation, sci-kit-learn, scipy.sparse for sparse matrices. We use a memory-based collaborative filtering technique.

4.1. Wrangling or Cleansing of Data. In the selected data set event attribute contains different types of activities performed by users during navigation of web site. We have to categorize these events into three main categories buy, select and view, because they have different weights in preference computation, and binary formation of these entries in the data set. After expansion and transformation result shown in table 4.2.

4.2. Computation of Preference Rank. Computation of preference rank of each class of products, sub-setting of data-frame applied on product_ID. The preference rank of the product depends on the number of users interact with the product. It is computed by using equation (3.1). This shows the performance of the product. Product with product_ID 11278 is the highest performing and product_ID 33000 is least performing products as shown in table 4.3.

4.3. Similarity Computation. Similarity computation can be done in two ways, user-user similarity and item-item similarity. The similarity between target user u and a neighbor v can be computed using Pearson's Correlation coefficient [24] as equation (4.1). Ignoring all least performing items that are less than 0.75 preference score.

$$sim(u,v) = \frac{\sum_{p\varepsilon P}(r_{u,p} - \overline{r_u})X(r_{v,p} - \overline{r_v})}{\sqrt{\sum_{i\varepsilon I}(r_{u,i} - \overline{r_u})^2}X\sqrt{\sum_{p\varepsilon P}(r_{v,p} - \overline{r_v})^2}} \tag{4.1}$$

Table 4.1
Event table

UserID	Event	ProductID	Brand	Category
14893	Buy	11277	sony	Electronics_item
48958	view	11378	RedMe	Mobile
48357	Buy	11278	iPhone	Mobile
55674	select	11378	RedMe	Mobile
48357	View	21277	Sony	Electronics_item
48360	View	11278	iPhone	Mobile
48730	select	21277	Sony	Electronics_item
48351	View	11278	iPhone	Mobile
.	.	.	.	.
.	.	.	.	.

Table 4.2
After Expansion and Binary formation table in preprocessing

UserID	Buy	Select	View	ProductID	Brand	Category
14893	1	1	1	11277	sony	Electronics_item
48958	0	0	1	11378	RedMe	Mobile
48357	1	1	1	11278	iPhone	Mobile
55674	0	1	1	11378	RedMe	Mobile
48357	0	0	1	21277	Sony	Electronics_item
48360	0	0	1	11278	iPhone	Mobile
48730	0	1	1	21277	Sony	Electronics_item
48351	0	0	1	11278	iPhone	Mobile
4853	1	1	1	21309	Tata	Electronics_item
464056	0	1	1	21303	Nokia	mobile
483575	0	0	1	21301	Nokia	mobile
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Table 4.3
Product with its Preference Rank

ProductID	Brand	Category	Preference Rank
11278	iPhone	Mobile	6.1013572
11277	sony	Electronics_item	5.4464
44575	iPhone	Mobile	2.41421356
34275	Nokia	Mobile	2.25829
33488	RedMe	Mobile	3.07565139
12453	VIVO	Mobile	3.30039626
14575	iPhone	Mobile	3.0913
11378	RedMe	Mobile	6.0655
21266	Sony	Electronics_item	3.6670
21277	Sony	Electronics_item	5.49658159
22458	VIVO	Mobile	3.83711731
24555	iPhone	Mobile	1.75
33000	RedMe	Mobile	0.35355339
33277	Sony	Electronics_item	2.91421356
33479	RedMe	Mobile	1.75
21254	sony	Electronics_item	1.6771
11279	iPhone	Mobile	1.75
21278	sony	Electronics_item	5.7760
21280	sony	Electronics_item	4.65690
21284	sony	Electronics_item	5.6672
21286	sony	Electronics_item	4.9983
21309	Tata	Electronics_item	2.76862
.	.	.	.
.	.	.	.

where $r_{u,p}$ and $r_{v,p}$ be rank given by user u and v to item p, $\overline{r_u}$ and $\overline{r_v}$ be the average rank of user u and v. It is best suited for less number of users but the main problem with the user based is scalability when the number of users is very large then it very difficult to find the similarity among the users.

4.4. Recommendation List Generation. Using the equation (4.1), we can compute the ranking prediction of item or product i for the target user u. Association of products with user play a vital role in recommendation list generation for the users. User-user based collaborative filtering (equation 4.2) and top N recommendation technique [12] is used to generate the list of recommendation of items the user. Recommendation lists of each user are shown in table 4.4.

$$Pre(u,p) = \overline{r_u} + \frac{\sum_{v\varepsilon V} sim(u,v)X(r_{v,p} - \overline{r_v})}{\sum_{v\varepsilon V} |sim(u,v)|} \tag{4.2}$$

where $Pre(u,p)$ is the prediction score of item/product p with user u, V is the set of all users that are similar to u, $r_{v,p}$ is ranking of user v given to the item p. After computing similarity between items, we select as k

Table 4.4
List of Products Recommended to User

UserID	**ProductID**	**Brand**	**Category**	**Preference Rank**
4830	11278	iPhone	Mobile	6.101357
.	21277	Sony	Electronics_item	5.496582
.	12453	VIVO	Mobile	3.300396
.	33277	Sony	Electronics_item	2.914214
.	34275	Nokia	Mobile	2.25829
.	21254	Sony	Electronics_item	1.866025
4890	21277	sony	Electronics_item	5.22129
.	21266	Tata	Electronics_item	3.6670
.	14575	iPhone	Mobile	3.0913
.	33523	iPhone	Mobile	2.9352
4852	11278	iPhone	Mobile	6.101357
.	36000	RedMe	Mobile	2.914214
.	44575	iPhone	Mobile	2.414214
4893	11278	iPhone	Mobile	6.101357
.	21277	Sony	Electronics_item	5.496582
.	12453	VIVO	Mobile	3.300396
.	33277	Sony	Electronics_item	2.914214
.	34275	Nokia	Mobile	2.25829
.	21254	Sony	Electronics_item	1.866025
.	.	.	.	.
.	.	.	.	.

most similar items to the target item and generate prediction value of user u. Some users' recommendation lists are shown in table 4.4. The recommendation list for a user contains those products which are not bought by the user. High performing products always on the top of the list.

4.5. Evaluation Matrices. To evaluate the quality of the proposed system we assess the general performance using three most commonly used evaluation matrices [25] in the retrieval system, precision (Eq 4.3), recall (Eq 4.4), and Normalized Discounted Cumulative Gain (NDCG) (Eq 4.6). The recommendation is viewed as an information retrieval task i.e. retrieve all products which are predicted to be ”good”.

Precision is a measure of exactness, determines the fraction of relevant products retrieved out of all products retrieved.

$$\text{Precision} = \frac{|\text{Good products recommended}|}{|\text{Total number of recommended products}|} \tag{4.3}$$

It measures the system's ability to reject any non-relevant products in the retrieved set

Recall is a measure of completeness, determines the fraction of relevant products retrieved out of all relevant products.

$$\text{Recall} = \frac{|\text{Good products recommended}|}{|\text{Total good products}|} \tag{4.4}$$

It measures the system's ability to find all the relevant products in the retrieved set.

Discounted Cumulative Gain (DCG) is measured the usefulness or gain of a product based on its position in the result list. The DCG accumulated at a particular rank position p of any product.

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2 (i+1)} \tag{4.5}$$

Normalized Discounted Cumulative Gain (NDCG): performance from one query to next can not be consistently achieved using DCG only hence NDCG is used. It is done by sorting all relevant products by their relative relevance producing the maximum possible DCG through position p of any product. NDCG is computed as

$$NDCG_p = \frac{DCG_p}{IDCG_p} \tag{4.6}$$

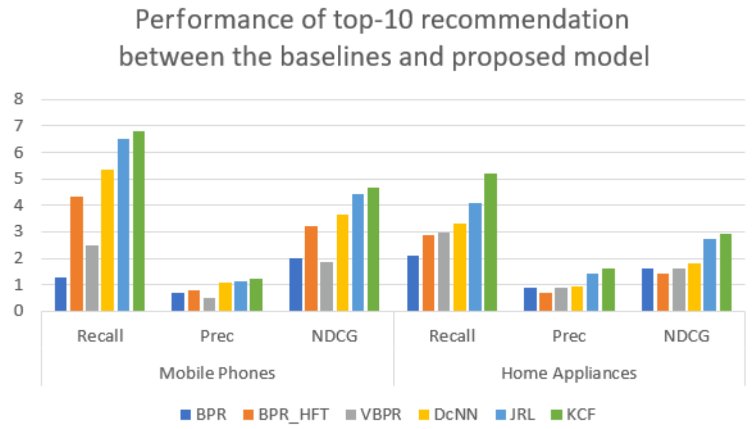


Fig. 4.1. *Performance of top-10 recommendation between the baselines and Knowledge based collaborative filtering(KCF)*

4.6. Comparisons with Baseline Methods. We choose commonly used representative and state-of-the-art methods as the baselines for the performance comparisons.

BPR: The Bayesian Personalized Ranking [26] is a popular model used in top-N recommendation system, it is based on matrix factorization. It gives good results small data-set.

BPR_HFT: Hidden Factors and Topics (BPR_HFT) [27] is a method used for textual reviews; we use HFT under BPR pair wise ranking framework for fair comparison.

VBPR: Visual Bayesian Personalized Ranking (VBPR) [28] is used for recommendation with images.

DeepCoNN: Deep Convolution Neural Network (DCNN) [29] is a review based deep recommendation method to jointly model the users and the products.

JRL: Joint Representation Learning (JRL) [30] is a model which can leverage multi-model information for Top-N recommendation.

The performance of the proposed system as well as the baseline models are shown in the performance graph (see fig. 4.1). The baseline models are categorized based on the input source of information which is review based (HFT and DeepCoNN), rating based (BPR), Image-based (VBPR), and heterogeneous information modeling JRL. The information is modeled as rating though buy relation, review though select relation and knowledge about the product through view, also_view, belong_to_category, belong_to_brand relations.

From the experimental result, it is clear that both review based and rating based models enhanced the performance of recommendation system as compare to popularity based, but heterogeneous information source-based model like JRL performs much better than baseline systems, which gives the idea about the heterogeneity. Knowledge-based collaborative filtering (KCF) performs much better than that of JRL consistently overall evaluation measures that verify the proposed system. This improvement in the performance, due to more information sources, and better-structured knowledge graph to model the heterogeneous information. Figure 4.1 shows the comparison of KCF with the other baselines models from the comparison graph it is clear that KCF performs better as the selected baseline system models in the area of recommendation system.

Conclusion and Future Scope. Choices made by the e-Commerce web site must be considered, it influences the decision of the customer to buy the products. Hence it is the main focus of the recommendation system. The experimental results on the e-Shop simulated dataset show the promising performance of our proposed approaches in terms of the accuracy of the recommendation in comparison with the traditional baseline approaches. The proposed system focuses on users' event-driven approach, hence the participation of user increases as compare to those methods which are based on reviews or ratings because about 20 to 40 percent of users participated in review and rating while others are not interested but they still interact with the products. Therefore, the proposed system generates a more closed list as a recommendation as compared to other such systems. This research very useful for small scale enterprises like retailers since it requires storing only user navigation details to generate the ranking score of the item, and item details to recommend the list of items in the future. For future work, several interesting directions can be explored and experimented, we may extend this technique on a large real-time dataset in the future.

Academic Contribution. All most every recommendation system is based on anyone's method from popularity, review, rating, etc. nowadays, but in some popular, and large enterprises choose hybrid method due to a huge number of users and budget allocation. It requires a powerful and efficient computing environment which is very costly, about to unaffordable by the small scale of retail sales, Users involvement in all the above methods is very low in ratio as compare to the number of users associated with the system. The proposed system is very useful for small scale enterprise and retail shops, where the number of users is low. Every user is important so we can handle it very efficient manner due to consideration of view events in the generation of the ranking score of any item.

Acknowledgments. The first author express the profound sense of gratitude and deep sentiments to the supervisor Dr. O.P.Rishi, Associate Professor, Department of Computer science and Informatics, University of Kota, Kota, for their generous help, perpetual encouragement and abundant interest towards the paper. It was their benign direction that saw the initiation, conduction, and completion of the paper. The first author also thanks the publisher for the INFO-TeX mailing list for the valuable indirect assistance he received, which is very helpful for writing this paper.

REFERENCES

[1] Z. Zeng,*An Intelligent E-Commerce Recommender System Based on Web Mining*, International Journal of Business and Management (IJBM), Vol. 4, No. 7, pp 10-14, July 2009.

[2] Pitkow et al,*PERSONALIZED SEARCH A contextual computing approach may prove a breakthrough in personalized search efficiency*, COMMUNICATIONS OF THE ACM Vol. 45, No. 9,2002.

[3] C. C. Aggarwar, *Recommender Systems: Text Book*, Springer,(2016) DOI 10.1007/978-3-319-29659-3.13.

[4] M. Balabanovic and Y. Shoham,*Content-Based Collaborative Recommendation* Comm. ACM, Mar. 1997, pp. 66-72.

[5] S. Panagiotis, A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos,*Collaborative recommender systems: Combining effectiveness and efficiency* Expert Systems with Applications 34.4 (2008) pp 2995-3013.

[6] H. Sadreazami, A. Asif and A. Mohammadi, *Iterative Graph-Based Filtering for Image Abstraction and Stylization*, in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 2, pp. 251-255, Feb. 2018.

[7] Y. Dai, H. Ye, and S. Gong, *Personalized Recommendation Algorithm Using User Demography Information*, 2009 Second International Workshop8on Knowledge Discovery and Data Mining, Moscow, 2009, pp. 100-103.

[8] A.U. Martliong and N.M.S. Iswari, *Rancang Bangun System Rekomendasi Restoran Menggunakan Metode AHP dan VIKOR pada platform Line*, J.Ultim. Comput. Vol 10 no. 1,(2018), pp 27-33.

[9] Z. Wang, J. Liao, Q. Cao and H. Q, *Friendbook A Semantic based friend recommendation for Social Networks.*, IEEE Transaction, Mob. Comput., vol. 14 no. 3, (2015) pp. 538-551.

[10] S.Chen, S. Owusu and L. Zhou , *Social Network Based Recommendation Systems: A short Survey*, doi:10.1109/SocialCom.2013.134,2013

[11] C.A. Gomez Uribe and N. Hunt, LU-*The Netflix Recommender System: Algorithms, Business Value, and Innovation*, ACM Trans. Manag. Inf. Syst. 2015

[12] N.M.S. Iswari, Wella and A. Rusli, *Product Recommendation for e-Commerce System Based on Ontology*, IEEE Explore, ICORIS -2019

[13] D. Wang, Y. Lianga, D. Xua, Xi. Feng, R. Guan,*A content-based recommender system for computer science publications*, Knowledge-Based Systems 157 (2018) 1–9.

[14] A. Kanakia, Z. Shen, D. Eide, K. Wang,*A Scalable Hybrid Research Paper Recommender System for Microsoft Academic*, ACM ISBN 978-1-4503-6674-8/19/05. (2019) https://doi.org/10.1145/3308558.3313700

[15] A. Kouadria, O. Nouali, M. Yahya, H. Al-Shamri,*A Multi-criteria Collaborative Filtering Recommender System Using Learning-to-Rank and Rank Aggregation*, Arabian Journal for Science and Engineering September 2019 https://doi.org/10.1007/s13369-019-04180-3

[16] M. K. Singh, O. P. Rishi, S. Awasthi, A. P. Srivastava and S. Wadhwa, *Classification and Comparison of Web Recommendation Systems used in Online Business*, 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 2020, pp. 471-480.

[17] L. Yan, *Personalized Recommendation Method for E-commerce Platform based on Data Mining Technology*, Proc.-2017Int. Conf. Smart Grid Electr. Autom ICSGEA 2017, Vol. 2017- Jan. (2017), pp 514-517.

[18] D.Yu, Y. Mu, and Y. Jin,*Rating Prediction using review text with underlying sentiment*,Information Processing Letters,117, (2015), pp 10-18.

[19] P.Lops, M.D.Gemmis, and G.Semeraro,, *Content-based recommender systems: state of the art and trends,* ", in Recommender Systems Handbook, Springer, (2011), pp. 73–105.

[20] S.S. Li and E. Karahanna,*A review and future directions online recommendation system in B2C E-commerce Context.*Journal of Association for Information Systems Online Recommendation Systems in B2C E-commerce Context. 16(2),(2015), pp 72-107.

[21] J. McAuley, C. Targett, Javen and A.V.D Hengel,*Image-based Recommendations on styles and substitutes*, SIGIR-15, August 09-13,(2015), Santiago, Chile.

[22] Y. Zhang, Q. Ai, X. Chen and P. Wang, *Machine Learning and Knowledge Discovery in Databases-Part III*, Springer (2018)

[23] |online|,*Available: https//pandas.pydata.org*, Accessed September 21,2018.

[24] Hong, Bo, Yu, Mengchen ,*A collaborative filtering algorithm based on correlation coefficient*, Neural Computing and Applications, November 2018, DOI 10.1007/s00521-018-3857-7.

[25] A. Gunawardana, G. Shani,*Evaluating recommender systems*, in Recommender Systems Handbook, Springer, 2015, pp. 265–308

[26] S Rendle. C Freudenthaler, Z. Gantner, L. Schmidt-Thieme, *BPR: Bayesian personalized ranking from implicit feedback*, In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 18–21 June 2009; pp. 452–461.

[27] J.McAuley, J. Leskovec,*Hidden: factors and hidden topics: understanding rating dimensions with review text*, In Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China,12–16 October 2013; pp. 165–172.

[28] R. He, J. McAuley ,*VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback*, In Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 144–150.

[29] L. Zheng, V. Noroozi, P.S. Yu,*Joint deep modeling of users and items using reviews for recommendation*, In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 425–434.

[30] Y. Zhang, Q. Ai, X.Chen, w.B. Croft,*Joint representation learning for top-n recommendation with heterogeneous information sources*, In Proceedings of the 2017 ACM Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 1449–1458.

Edited by: Anand Nayyar
Received: Mar 10, 2020
Accepted: Jul 6, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 379–389, DOI 10.12694:/scpe.v21i3.1725

DEEP CONVOLUTIONAL NEURAL NETWORK WITH TENSORFLOW AND KERAS TO CLASSIFY SKIN CANCER IMAGES

HOUSSAM BENBRAHIM[*], HANAÂ HACHIMI[†], AND AOUATIF AMINE[‡]

Abstract. Skin cancer is a dangerous disease causing a high proportion of deaths around the world. Any diagnosis of cancer begins with a careful clinical examination, followed by a blood test and medical imaging examinations. Medical imaging is today one of the main tools for diagnosing cancers. It allows us to obtain precise images, internal organs and thus to visualize the possible tumours that they present. These images provide information on the location, size and evolutionary stage of tumour lesions. Automatic classification of skin tumours using images is an important task that can help doctors, laboratory technologists, and researchers to make the best decisions. This work has developed a classification model of skin tumours in images using Deep Learning with a Convolutional Neural Network based on TensorFlow and Keras model. This architecture is tested in the HAM10000 dataset consists of 10,015 dermatoscopic images. The results of the classification of the experiment show that the accuracy was achieved by our model, which is in order of 94.06% in the validation set and 93.93% in the test set.

Key words: Skin Cancer, Image Classification, Deep Learning, Convolutional Neural Network, TensorFlow, Keras, HAM10000 Dataset.

AMS subject classifications. 68U10, 68T05

1. Introduction. Skin cancer is the most frequently diagnosed disease in the world, it has an impact on the quality of life and it is can be deadly [1]. Skin cancer is considered as an important problem in public health, however, most cases are avoidable if detected early with a better prediction [2]. In Morocco, cancer is a health major problem requiring a comprehensive policy of care [3]. The situation in Morocco is very disturbing, there are approximately 30,000 new cases of cancer per year. Cancer is responsible for 7.2% of all deaths in Morocco, whereas skin cancer has been growing rapidly and it achieves ascending numbers of cases per year [4]. The diagnosis of skin cancer is necessary to reduce the negative effects of the dangerous development of this disease [5], which is usually performed with a doctor using visual inspection. This inspection is dependent on the experience and training of the doctor. In general, doctors can identify skin cancer with a sensitivity of 75% and specificity of 87% [6]. New techniques of image recognition such as Deep Learning technology can help doctors to better diagnose skin cancer disease. This method uses powerful and very advanced algorithms that achieve effective results [7]. Convolutional Neural Networks are the most frequently used Deep Learning algorithms, which are based on the human visual cortex. Its methods have been favourably admitted for numerous imaging classification [8]. In Morocco, the use of new technologies is very low in the health sector, which also negatively affects the detection of cancer in general and specifically skin cancer [9]. Morocco needs a national electronic system for the health sector based on powerful and advanced technologies [10], that can allow doctors to better analyze and diagnose all diseases and especially skin cancer. The general objective of this paper is to develop a system based on Deep Convolutional Neural Networks with TensorFlow and Keras that will be able to detect skin cancer. For this reason, this work used a test database named HAM10000, consists of 10,015 dermatoscopic images, and it has 7 different classes of skin cancer. The use of an intelligent model, which is based on powerful and innovative technologies, will be able to help doctors make a better diagnosis of skin cancer. This system

[*] BOSS-Team, GS-Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco (houssam.benbrahim@uit.ac.ma).

[†] BOSS-Team, GS-Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco (hanaa.hachimi@univ-ibntofail.ac.ma).

[‡] BOSS-Team, GS-Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco (aouatif.amine@uit.ac.ma).

will aim to avoid the consequences of delay or bad diagnosis, as well as, to sensibilize Moroccan doctors for the importance and the necessity to employ its advanced techniques in medical imaging examinations.

2. Deep Learning. Deep Learning represents a class of Machine Learning and Artificial Intelligence (AI) techniques where the machine can learn on its own, unlike traditional programming where the program simply executes predetermined rules [11]. What is first the Machine Learning concept?
Machine Learning (ML) is an innovative technology that gives the possibility of computers to learn through experience [12]. In ML programming, the learning process is supervised and the programmer is the person responsible for extracting the characteristics of an image, for example [7]. With ML, computers need data to analyze and to make predictions or decisions by themselves. In recent decades, various ML techniques were developed to create algorithms that could learn and improve independently [13]. Deep Learning (DL) is, therefore, a set of automatic learning techniques where the machine learns to recognize patterns by training on qualified data models. The algorithms that make its operations possible use mathematical concepts based essentially on Artificial Neural Network (ANN) [14]. The advantage of DL is that the program itself extracts the features, which makes the treatment faster, and more accurate.

2.1. Performance. In the human brain, each neuron receives about 100,000 electrical signals from other neurons to detect for example an object, on this concept and more specifically on the ANN that based the DL [15]. This network is composed of several ”layers” of neurons, each one receiving and interpreting the information of the previous layer. The higher the number of neurons is, the more the result is deep [16]. For a high level of accuracy, DL programs must have access to big amounts of training data and it uses powerful processing methods, that were difficult for programmers before the advent of Big Data and Cloud Computing technologies. With its two latest advanced techniques, DL programming has become fast, simple and very effective [17].

2.2. Application subjects. DL was applied to several problems over the last few years. It allows us to perform all kinds of tasks at incredible speed. There are many different applications of DL, here are some examples of using this technique [18].

- Health care: DL can detect symptoms of diseases such as cancer detection and drug discovery, relying on image data for example.
- Speech recognition: DL is used for various applications for speech recognition, such as Google Voice, Cortana, Amazon Echo, etc.
- Autonomous cars: DL is employed by researchers to detect white lines, pedestrians crossing, signs, other vehicles on the road, etc.
- Security: DL is applied for face recognition and video surveillance.

This is not an exhaustive list, but there are many other applications for example in natural language processing, customer relationship management, recommendation systems, and bioinformatics [19].

3. Convolutional Neural Network. A Convolutional Neural Network (CNN) is a type of ANN, in which the connection pattern between neurons is inspired by the animal’s visual cortex. However, CNN’s are specifically designed to process input images. It gets automatically and strongly the principal features to use in classifying images, that means it learns directly from samples [27]. Through several specific layers of sophisticated mathematical operations, CNN performs this type of automatic selection and classification of features. This technique has marked great importance on the DL committee in recent years especially in the field of image recognition or classification, machine vision projects, as well as skin cancer detection [28].

3.1. CNN for Classification. Image classification refers to the task of retrieving information classes from an image, which is one of the most principal tasks in computer vision and pattern recognition[29]. However, the task of choosing the best characteristics is enormously lengthy and is frequently unsuccessful. Also, increasing the features of other images types become big trouble [30]. CNN can solve an image classification problem by choosing features that could be the pixel location, the colour, object edges, or any other characteristics that could be taken from the images. This process is performed automatically by CNN. The better and more effective the feature sets extracted, the more accurate and efficient the image classification we can obtain. CNN’s try to find a solution to this problem by employing more hidden layers [31]. To classify images and to construct a

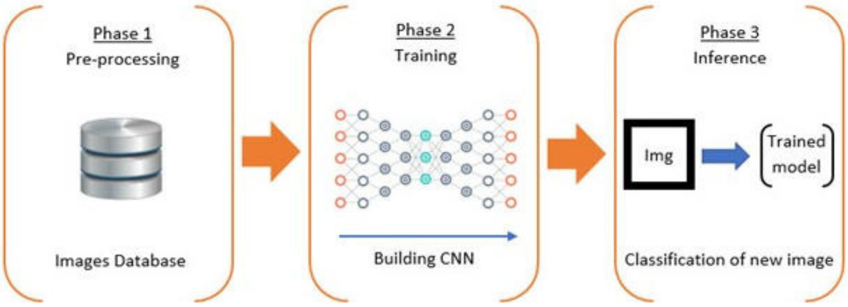


Fig. 3.1. *The Deep Learning process.*

digit identification system using CNN, the DL process consists of the following phases [32], the first step is the preprocessing of the input data, second, training the DL model, and finally the inference of the model. Firstly, the transformation of the images into a legible format. Secondly, the creation of a CNN and train it with several images. Finally, the validation of the model by the classification of new images. Figure 3.1 summarizes its phases.

3.2. CNN Architecture. In the training process, a deep neural network has multiple hidden layers, which are different in the type and their connectivity. CNN has multiple Convolutional layers, Non-Linearity layers, pooling layers, and fully connected layers [33]. The main purpose of the first layer is to apply a convolution operation to the input data, which means to detect different patterns or features from an input image, after that, it will transmit the result to the next layer. So, this layer defines a filter, which is also called a kernel. If the program slides the filter over the image and applies the dot product of the filter to the image pixels, the result would be a new image with all the edges. The second (ReLu layer) is integrated to replace all negative pixel values in the feature map by zero. The third layer is called Max-Pooling. This last integrates the outputs of a cluster of neurons in the precedent layer into one neuron in the next layer, which is an operation that finds the maximum values and simplifies the inputs. In other words, it reduces the number of parameters within the model. It turns the low-level data into higher-level information. And the fourth layer links every neuron in the precedents layer to every neurone in the next layer. Fully-Connected layers take the high-level filtered images from the previous layer and convert them into a vector [34].

4. TensorFlow. Created by the Google Brain team in 2011, TensorFlow is a dedicated system for tasks that require heavy numerical calculations. Originally called DistBelief, DistBelief's source code was changed and this tool became an application-based library. In 2015, it was renamed TensorFlow and Google made it open source. This system has an interface for Python and C++ and it went into version 1.0 in February 2017 [20]. Simply put, TensorFlow is a library of ML and deep neural networks, it is a toolbox for solving extremely complex mathematical problems. It can be conceived as a programming system in which the calculations are represented in the form of a data flow graph, which means we can create first a program and then execute it in a session [21]. TensorFlow offers several advantages for an application. It's able to run faster than pure Python code due to a C/C++ backend. It should be noted, that TensorFlow supports parallel computing, CPU and GPU [22].

4.1. TensorFlow structure. TensorFlow's structure is built on the running of a data flow graph. This last has two basic units: the nodes and the edges. The first one represents a mathematical operation, and the second defines the multidimensional arrays, known as tensors. As cited before the standard use of this structure is to run a session after the creation of the graph. The session translates and passes the computations represented into the graphs to the desired environment to run them on a GPU or CPU [23]. Figure 4.1 represents a graph in a session with TensorFlow. For example, Z and Y are tensors. The program can call MatMul (an operation over the tensors Z and Y), after that, the operation Add is used, and the program can add the result with the tensor X. To get the desired result W, the outcoming tensors of each working pass through

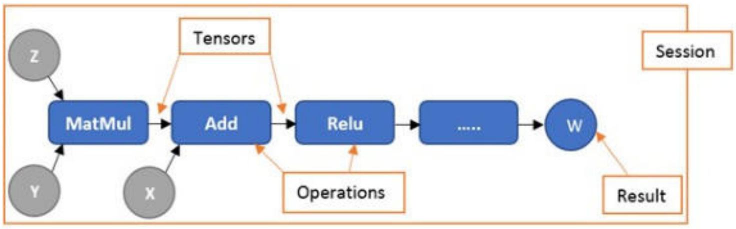


Fig. 4.1. *TensorFlow work example.*

Table 4.1
Tensor rank.

Dimension	Mathematical Object	In Code
Zero	Scalar	[1]
One	Vector	[1,2,3,4,...]
Two	Matrix	[[1,2,3,4,...],[1,2,3,4,...],[1,2,3,4,...],...]
Three	Tensor	[[[1,2,...],...],[[1,2,...],...],[[1,2,...],...],...]

the next one up to the end. Finally, the program can build a session to execute the graph and implement the calculations. So, what means a tensor? In mathematics, more precisely in multilinear algebra and in differential geometry, a tensor designates a very general object whose value is expressed in a vector space. It can be used to represent multilinear applications or multivectors. A Tensor is a multidimensional array, it may be in different dimensions, zero-dimensional, one dimensional, 2-dimensional, etc. Table 4.1 summarize its information [24]. The Tensors giving programmers the freedom to form a dataset the way they want it, especially it's helpful when dealing with images, because of the nature of how the information contained in the images is encoded.

4.2. Architecture of TensorFlow. TensorFlow is built to support and execute computational graphs with new ML models and system-level optimization. TensorFlow supports programming language interfaces C++, Python, and others. When we create and compile a program with the TensorFlow system, the language binding will invoke the TensorFlow library which includes the Core Execution System. It allowed programmers to deploy very complex computations on CPUs, GPUs, a desktop, server, or a mobile device. This signifies that programmers can build their program once, and then they can run it easily on different devices [25]. TensorFlow architecture has three primary steps. The first one is the data processing, programmers can write the codes in several programming languages, and then TensorFlow converts that into hardware instruction sets for CPU, GPU, Android, etc. The second one is to build the model in layers. The last one for training and estimating the model, for this reason, programmers can use Estimator and Keras model. The canned Estimator allows TensorFlow to support Neural Network, Logistic Regression, and Linear Regression [26].

5. Material and methods. In this study, a Deep CNN with Keras TensorFlow model to classify skin cancer images into seven classes was created, tested and validated. This work used the HAM10000 dataset that has seven several classes of skin cancer which are: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions. In the test phase, the Anaconda environment (Python distribution) was used with TensorFlow and Keras model. The execution of our program is made on a CPU with Max RAM 16 GB. In this part, this paper presents a detailed description of the data set and it gives a general overview of the Keras TensorFlow model.

5.1. The HAM10000 dataset. The dataset "Human Against Machine with 10000 training images" is a great series of multi-source dermatoscopic images of common pigmented skin lesions [35]. Published and created by Philipp Tschandl (Medical University of Vienna), Cliff Rosendahl (University of Queensland) and Harald Kittler (Medical University of Vienna), Austria [36]. This database obtained from several populations acquired and stored by multiple modalities. The final dataset consists of 10,015 dermatoscopic images that are released as a training set for academic machine learning purposes and are publicly available through the ISIC archive. Response data are all encoded within a single CSV file (comma-separated value) file, with each

Table 5.1
Attributes information of the Dataset.

Column name	Description
lesion_id	Input image identifier of the form HAM_
image_id	Input image identifier of the form ISIC_
dx	1. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec). 2. Basal cell carcinoma (bcc). 3. Benign keratosis-like lesions (solar lentigines /seborrheic keratoses and lichen-planus like ker- atoses, bkl). 4. Dermatofibroma (df). 5. Melanoma (mel). 6. Melanocytic nevi (nv). 7. Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc).
dx_type	1. histo 2. follow_up 3. consensus 4. confocal
age	The age of the patient
sex	The gender of the patient
localization	Location of the tumour in the patient's body

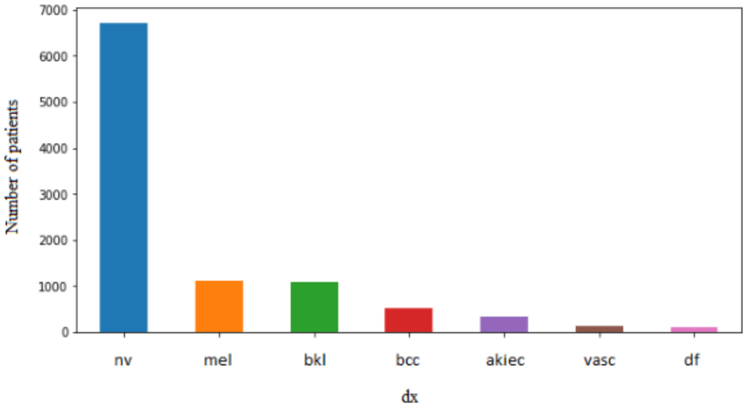


Fig. 6.1. *The distribution of different classes of cell type.*

classification response in a row. Table 5.1 shows the attribute information of the HAM10000 dataset.

5.2. Keras Model. Keras is the advanced model for the Python DL. Keras able to run on top of TensorFlow, Microsoft Cognitive, R, Toolkit, PlaidML, or Theano. It was created by François Chollet (Software Engineer at Google) to allow quick experiments. Its latest version is 2.3.1 (October 7, 2019). Keras is recommended for:

- Rapid and easy prototyping.
- Sustains convolutional networks and recurrent networks.
- Works on CPU and GPU.

For the test phase, the Keras sequential model is used by adding a list of layers instance to the constructor like the Convolutional layer, Pooling layer, Flatten and Dense layers.

6. Results.

6.1. Data Explore. After importing and cleaning the data, the exploratory data analysis was started. In this step, different attributes of the HAM10000 was explored. The database has seven different classes of cell type, Figure 6.1 visualizes its distribution. The maximum number of images are of type nv (6,705 images) and the minimum number is of type df (115 images), the others are classified as follows: mel 1,149 images, bkl 1,063 images, bcc 514 images, 327 images for akiec and 142 images for vasc. There are four categories of

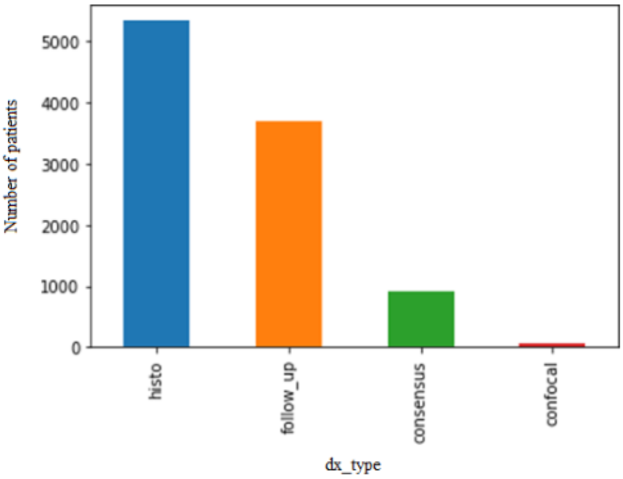


Fig. 6.2. *The distribution of four categories of the dx_type.*

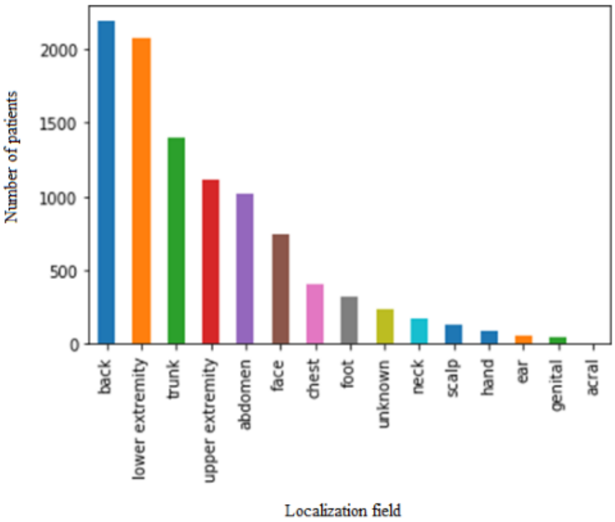


Fig. 6.3. *The distribution of the localization field of skin cancer.*

the dx_type which represent the technical validation field (ground truth), the distribution of this feature is plotted in Figure 6.2. More than 50% of lesions were confirmed through histopathology (histo) which means diagnosis has been performed by specialized dermatopathologists. The rest (follow_up) is more than 3,500 instances, (consensus) for a loan of 1,000 instances, and finally confirmation by in-vivo confocal microscopy (confocal). Figure 6.3 represents the distribution of the localization field of skin cancer in the database. The results confirm that back, the lower extremity was heavily compromised regions of skin cancer at the first level, in the second level the trunk, upper extremity, abdomen and face have been marked, other cases continue to spread throughout the body. The age distribution shows that there are large cases of patients between 30 and 60 years. Figure 6.4 illustrates the results. For the gender of patients, the results affirm that more than 50% are male, the distribution of males and females is illustrated in Figure 6.5.

6.2. Experiments. In this phase, before the implementation of the model, the program loaded the images. The input data are dermoscopic lesion images in JPEG format where there are 7 possible disease categories. Figure 6.6 presents one sample for each disease category. The reel size of the loaded images is 450x600x3, which

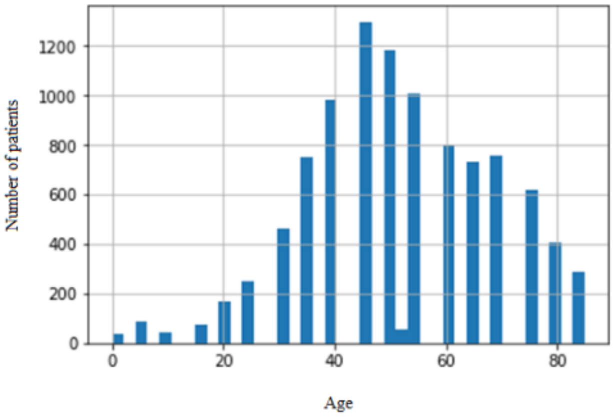


Fig. 6.4. *The age distribution.*

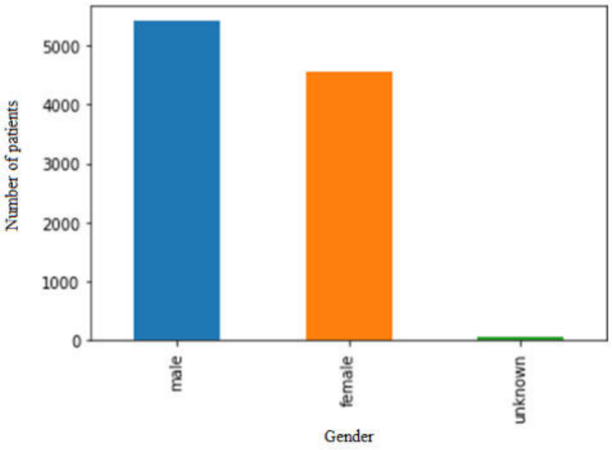


Fig. 6.5. *The sex distribution.*

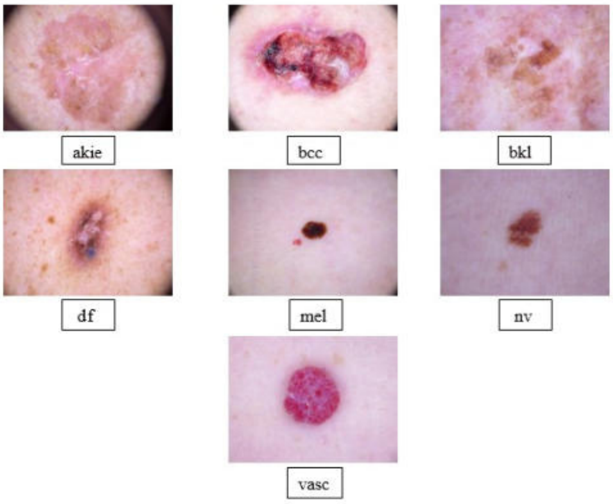


Fig. 6.6. *An example of each skin cancer disease.*

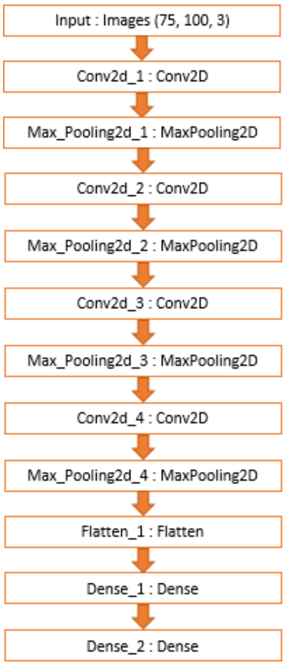

Input : Images (75, 100, 3)
Conv2d_1 : Conv2D
Max_Pooling2d_1 : MaxPooling2D
Conv2d_2 : Conv2D
Max_Pooling2d_2 : MaxPooling2D
Conv2d_3 : Conv2D
Max_Pooling2d_3 : MaxPooling2D
Conv2d_4 : Conv2D
Max_Pooling2d_4 : MaxPooling2D
Flatten_1 : Flatten
Dense_1 : Dense
Dense_2 : Dense

Fig. 6.7. *The general architecture of our model.*

Table 6.1
Model summary.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 75, 100, 64)	1792
max_pooling2d_1 (MaxPooling2D) (Conv2D)	(None, 36, 49, 64)	0
conv2d_2 (Conv2D)	(None, 34, 47, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 17, 23, 64)	0
conv2d_3 (Conv2D)	(None, 15, 21, 128)	73856
max_pooling2d_3 (MaxPooling2D)	(None, 7, 10, 128)	0
conv2d_4 (Conv2D)	(None, 5, 8, 128)	147584
max_pooling2d_4 (MaxPooling2D)	(None, 2, 4, 128)	0
flatten (Flatten)	(None, 1024)	0
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 512)	524800
dense_2 (Dense)	(None, 7)	3591

TensorFlow can't handle, that's why the program resize all the images of the database HAM10000 (10,015 images) into 75x100x3. For the implementation of the model, the database was partitioned into 3 sets (80% train, 10% validation and 10% test). In this step, the model receives 75x100 images as input, then has a sequence of four convolutional and pooling layers as feature extractors, followed by a flatten operation and two dense layers. A plot of the model is created in Figure 6.7. In the model, the first imported layer is the convolutional (Conv2D), the program uses 64 filters for the two firsts ones and 128 for the two last ones. Its layers are employed to extract features from the images. The second employed layer in the architecture is the pooling (MaxPool2D). This last is used four times to reduce the dimensionality of the images. After that, the program added a Flatten layer to transform the two-dimensional matrix of features into a vector. Also, the Dropout layer integrated to improve overfit on neural networks. Finally, the program integrated the Dense layer that regroups the results and generates a prediction. The CNN architecture is detailed in Table 6.1. In the first step, the performance of the model is calculated by measuring the accuracy index, it refers to the proximity of the measurements to giving value. Figures 6.8 shows the model accuracy of the architecture. Secondly,

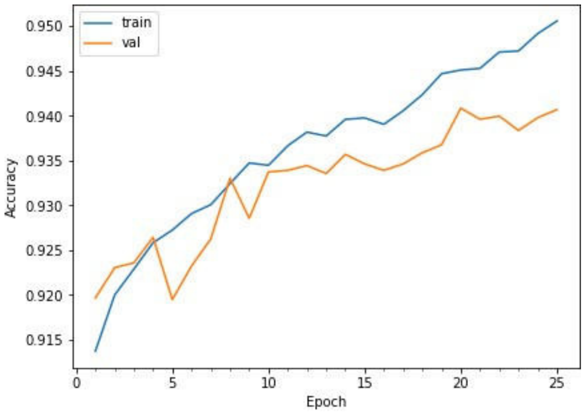


Fig. 6.8. *Model accuracy.*

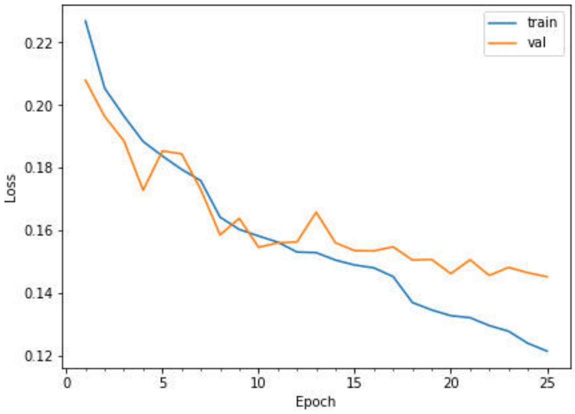


Fig. 6.9. *Model loss.*

establishing the model loss, It defines how much the modelling of the problem, which is an approximation of reality, loses information compared to the reality observed through the example data. Figure 6.9 plots the model loss of the architecture. This work confirms that the model achieves the accuracy of 94.06% in the validation set and 93.93% in the test set, as well as 14.52% for loss validation and 14.63% for test loss with epochs equal to 25. Epoch means an instant of time marked by an event that begins a new period, it is usually characterized by a distinctive development. These results show the importance of automatic image classification and especially the detection of skin cancer using its advanced technologies such as Deep Learning and more precisely CNN, TensorFlow, and Keras model. Morocco needs to use its techniques in the field of health in general and that Moroccan doctors in addition to their experiences and their training can be familiar with its approaches, methods, and techniques to increase the rate of a best and effective diagnosis.

7. Discussion. The overall objective of this work is to create an automatic classification of skin cancer images, for accurate identification of the tumour class. Fast, efficient and accurate detection of this disease can positively affect the next steps of the diagnosis. In this paper, a method based on deep CNNs with TensorFlow and Keras model was proposed for the extraction of the skin cancer in JPEG images. In this article, the input of the architecture is 75x100 images and the output is a layer of seven neurons because there are seven classes of skin cancer in the HAM10000 database. In this model, four layers of convolutions were created each with MaxPooling layer. Before feeding into a Dense layer the Dropout is a little trick to improve the efficiency of a neural network by throwing away some of the neurons. All of this is programmed in a sequential model based on TensorFlow with Keras. The experimental results proved that the proposed method can reach a very high accuracy of 94.06% in the validation set and 93.93% in the test set with epochs equal to 25. In a similar study created by [37], the authors developed a classification of images from the HAM10000 dataset, they used convolution, centring and full connection layers in the model created with the VGGNET architecture. In their research, the test phase of the educated model, class validation was obtained at 85.62%. In another work [38], the authors have trained a CNN based on the ResNet50 architecture to classify dermoscopy images of skin lesions. With their custom model, they obtained a balanced accuracy of 91% on the validation dataset. In [39], the authors created a skins cancer identification system of HAMl0000 dataset using CNN, the accuracy of training and testing of skins cancer identification system was 80% and 78%. In [40], the authors proposed an automated system for skin lesion classification through transfer learning-based deep neural network (DCNN). The experimental results achieved an accuracy of 89.8%. After all that we can say that the model proposed has achieved very important results and the architecture is succeed to classify dermoscopic images of skin lesions in one of the seven categories.

8. Conclusion. In this study, an architecture on Deep CNN using TensorFlow framework and Keras model was developed to classify 7 types of skin cancer. The implementation of the model was realized by applying an image classification structure on HAM10000 Database. The latter contains a significant number of skin cancer images in JPEG format knowing that there are seven possible disease classes. In Morocco, the number of cases of skin cancers is growing, and it can become dangerous due to a delay in the diagnosis phase or the failing prediction of the disease. The goal is to build a solution that able to classify skin tumours in the images. This application can help doctors, especially in Morocco to better diagnose, detect and quickly identify attacked patients by skin cancer and speed up the workflow. In this work, we design a CNN sequential model to classify the seven skin diseases. The results of the classification experiment show that the accuracy of the model was achieved, which had, in its best configuration, 94.06% and 93.93% in the validation set and test set, respectively. These results can convince doctors in Morocco to share digital images of their patients in private mode to create a common database that can be used as a reference for scientific research in DL context, also to enhance the use of advanced technologies for doctors. Our future work aims to implement the architecture proposed in this work into a Big Data environment, to prove the feasibility, the reliability of this model, and to improve their performance with several techniques like the use of Tensorflow on Spark.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, Cancer statistics, 2015, CA: a cancer journal for clinicians, 2015, vol. 65, no 1, pp. 5-29.

[2] C. Wild, World cancer report, World Health Organization, 2014, pp. 482–494.

[3] R. Bekkali, Lutte Contre Le Cancer Au Maroc- L'apport De La Fondation Lalla Salma, International Journal of Medicine and Surgery , 2017, vol. 4, no 1, pp. 55-59.

[4] Ministry of Health in Morocco and Lalla Salma Association of Fight Against Cancer, Plan national de prévention 2010-2019 Axes Stratégiques et Mesures, 2009.

[5] M. Harte and G. Knepil, Skin cancer detection, British dental journal 227, 2019, pp. 539–539.

[6] Y. A. Glickman, O. Filo, M. David, A. Yayon, M. Topaz, B. Zamir, A. Ginzburg, D. Rozenman, and G. Kenan, Electrical impedance scanning: a new approach to skin cancer diagnosis, Skin Research and Technology, 2003, vol. 9, no 3, pp. 262-268.

[7] L. Deng, and D. Yu, Deep learning: methods and applications, Foundations and Trends in Signal Processing, 2014, vol. 7, no 3–4, pp. 197-387.

[8] R. Zhao, W. Ouyang, H. Li, and X. Wang, Saliency detection by multi-context deep learning, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1265-1274.

[9] H. Benbrahim, H. Hachimi, and A. Amine, Survey on the Use of Health Information Systems in Morocco: Current Situation, Obstacles and Proposed Solutions, in : International Conference on Advanced Intelligent Systems for Sustainable Development, Springer, Cham, 2018, pp. 197-204.

[10] H. Benbrahim, H. Hachimi, and A. Amine, Moroccan Electronic Health Record System, in: International Conference on Industrial Engineering and Operations Management, 2018, Paris, France.

[11] Y. Bengio, A. Courville, and P. Vincent, Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence, 2013, vol. 35, no 8, pp. 1798-1828.

[12] S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh, and W. C. Hong, Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward, IEEE Access, 2019, vol. 8, pp. 474-488.

[13] C. M. Bishop, Representation learning: Pattern recognition and machine learning, springer, 2006.

[14] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks, 2015, vol. 61, pp. 85-117.

[15] S.A. Kalogirou, Applications of artificial neural-networks for energy systems, Applied energy, 2000, vol. 67, no 1-2, pp. 17-35.

[16] J. G. Lee, S. Jun, Y. W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, Deep learning in medical imaging: general overview, Korean journal of radiology, 2017, vol. 18, no 4, pp. 570-584.

[17] Q. Zhang, L.T. Yang, Z. Chen, and P. Li, A survey on deep learning for big data, Information Fusion, 2018, vol. 42, pp. 146-157.

[18] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, Deep learning applications and challenges in big data analytics, Journal of Big Data, 2015, vol. 2, no 1, pp. 1.

[19] N. F. Hordri, S. S. Yuhaniz, and S. M. Shamsuddin, Deep learning and its applications: a review, in : Conference on Postgraduate Annual Research on Informatics Seminar, 2016.

[20] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S Corrado, A. Davis, J. Dean, M. Devin, and others, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04467, 2016.

[21] M. Abadi, P. Barham, and J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and others, Tensorflow: A system for large-scale machine learning, in : 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016, pp. 265-283.

[22] M. T. Martinez, An Overview of Google's Machine Intelligence Software TensorFlow, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2016.

[23] G. Zaccone, Getting Started with TensorFlow, Packt Publishing Ltd, 2016.

[24] R. Bonnin, Building Machine Learning Projects with TensorFlow, Packt Publishing Ltd, 2016.

[25] P. Goldsborough, A tour of tensorflow, arXiv preprint arXiv:1610.01178, 2016.

[26] G. Zaccone, M.R. Karim, and A. Menshawy, Deep Learning with TensorFlow, Packt Publishing Ltd, 2017.

[27] Y. Kim, Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882, 2014.

[28] A. Khan, A. Sohail,U. Zahoora, and A. S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, arXiv preprint arXiv:1901.06032, 2019.

[29] W. Wang, D. Liang, Q. Chen, Y. Iwamoto, X. H. Han, Q. Zhang, H. Hu, L. Lin, and Y. W. Chen, Medical Image Classification Using Deep Learning, in : Deep Learning in Healthcare, Springer, Cham, 2020, pp. 33-51.

[30] W. Rawat, and Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, Neural computation, 2017, vol. 29, no 9, pp. 2352-2449.

[31] Q. Li, W. Cai, X. Wang, Y. Zhou, , D. D. Feng, and M. Chen, Medical image classification with convolutional neural network, in 2014 13th international conference on control automation robotics & vision (ICARCV). IEEE, 2014. pp. 844-848.

[32] I. Arel, D. C Rose, and T. P. Karnowski, Deep machine learning-a new frontier in artificial intelligence research [research frontier], IEEE computational intelligence magazine, 2010, vol. 5, no 4, pp. 13-18.

[33] A. Hidaka and T. Kurita, Consecutive dimensionality reduction by canonical correlation analysis for visualization of convolutional neural networks, in Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications, the ISCIE Symposium on Stochastic Systems Theory and Its Applications, 2017. pp. 160-167.

[34] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights into imaging, 2018, vol. 9, no 4, pp. 611-629.

[35] P. Tschandl, C. Rosendahl, and H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Scientific data, 2018, vol. 5, pp. 180161.

[36] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, and others, Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), arXiv preprint arXiv:1902.03368, 2019.

[37] E. Çevik and K. Zengin, Classification of Skin Lesions in Dermatoscopic Images with Deep Convolution Network, Avrupa Bilim ve Teknoloji Dergisi, 2019, pp. 309-318.

[38] N. D. Reddy, Classification of Dermoscopy Images using Deep Learning, arXiv preprint arXiv:1808.01607, 2018.

[39] A. A. Nugroho, I. Slamet, and Sugiyanto, Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network, in : AIP Conference Proceedings, AIP Publishing LLC, 2019, pp. 020039.

[40] M. A. Khan, M. Y. Javed, M. Sharif, T. Saba, and A. Rehman, Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification, in : 2019 international conference on computer and information sciences (ICCIS), IEEE, 2019, pp. 1-7.

Edited by: Anand Nayyar
Received: Mar 31, 2020
Accepted: Jul 17, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 391–406, DOI 10.12694:/scpe.v21i3.1734

DEADLOCK FREE RESOURCE MANAGEMENT TECHNIQUE FOR IOT-BASED POST DISASTER RECOVERY SYSTEMS

MADHAVI DEVI B[*], SMRITI AGRAWAL[†] AND R. RAJESHWARA RAO[‡]

Abstract. Disasters are inevitable, but their impact can be mitigated with careful planning. An IoT-based network with limited resources can be used in the post-disaster recovery. However, the resource of common interest creates contention among its contenders. This contention leads to tussle which in turn may lead to a deadlock. Some of the existing techniques prevent or avoid deadlock by performing stringent testing with significant testing overhead. While others propose recovery action after the deadlock is detected with significant overhead. A deadlock leads to a breakdown of the post-disaster recovery system while testing overhead implies delayed response either case can lead to catastrophic losses. This paper presents a new class of techniques that do not perform stringent testing before allocating the resources but still ensure that the system is deadlock-free and the overhead is also minimal. The proposed technique suggests reserving a portion of the resources to ensure no deadlock would occur. The correctness of the technique is proved in the form of theorems. The average turnaround time is approximately 18 % lower for the proposed technique over Banker's algorithm and also an optimal overhead of $O(m)$.

Key words: Banker's Algorithm, Deadlock, Deadlock avoidance, Deadlock Recovery, Operating systems, Safety sequence, post disaster management, IoT network, IoT resource scheduling

AMS subject classifications. 68M14

1. Introduction.

The UN Office for Disaster Reduction (UNISDR) [1] survey reveals that natural disasters such as floods, storms, heat-waves, etc., have claimed thousands of human lives, brought misery to millions of people and profoundly affected the economy. The recent pandemic of the corona virus outbreak is no exception. Other than natural disasters caused due to accidents in factories, homes, trains, roads, etc. also have catastrophic effects. In these circumstances, response to the disaster by effective utilization of the available resources is critical for the mitigation of their effects. The management of the limited available resources is essential for various post-disaster activities, such as evacuating or quarantining the people to safe places, providing medical help, etc. A network can connect these resources to find their whereabouts, availability, etc. This network can be established using the Internet of Things (IoT) [2, 3, 4, 5, 6, 7, 8, 9]. The IoT utilizes the internet and can be used in post-disaster recovery [10, 11, 12, 13, 14, 15, 16, 17, 18]. Authors [10, 11, 12, 13, 14, 15], represented the IoT for post-disaster system as an operating system with the set of activities as processes. This paper also considers the same.

Multiple post-disaster recovery activities (processes) of all types and sizes need to share resources for their effective utilization [11].Contention for a resource occurs when two or more processes request for it. This contention increases as the number of processes contending and the set of resources of interest increases. Eventually, some processes hold some resources and wait for others, which are held by another set of processes waiting for some other resources, forming a cycle of wait for. Hence, everybody is waiting for somebody to complete and no one has sufficient resources to complete, hence no process completes. The system utilization tends to become zero as none of the processes has the resources it needs to continue its execution. This situation is referred to as Deadlock [19, 20, 21, 22, 23].

Resource sharing systems such as single processor, multiprocessor systems, IoT based systems, parallel computing,cloud computing and distributed systems [15, 16, 17] with applications such as post-disaster management, distributed systems,automated manufacturing systems, etc. can be designed without keeping deadlock

[*] Research scholar JNTU Kakinada, India (madhavi.polagangu@gmail.com)

[†] CBIT, Hyderabad, India (agrawal.smriti@gmail.com)

[‡] Professor, JNTUUCEV, Vizianagaram, India

in account. The deadlock brings the system into a standby state which may have catastrophic effects. Further, it cannot be resolved on its own and requires external intervention. This is because no process will release the resources it has acquired without completing itself. Thus, efficient resource management[24, 25, 26, 27] is needed to manage the resources for avoiding the deadlock as well as improve the system performance by effective resource utilization and reduce the average turnaround time and increase the throughput of the system.

Coffman [22] suggested that a deadlock can occur only if, necessary conditions namely; Mutual exclusion, Hold and wait, No preemption and Circular wait.Thus, the solution for a deadlock problem lies around these conditions. The first set of the solutions suggests ensuring that one of the necessary conditions does not hold such techniques are referred to as *Deadlock Prevention (DP).* The second set of techniques is called *Deadlock Avoidance (DA)*, it is a look forward technique to avoid deadlock in future. The next set of solutions advocates reactive approach called *Deadlock Detection and Recovery (DDR)*, where the system is recovered after the deadlock is detected. Most recently, authors [24, 25, 26, 27],suggested a new class of techniques called *Resource Reservation Techniques (RR)* for deadlock handling.

The Resource Reservation (RR) techniques handle deadlock with lower overhead. They supported *reserving a set of resources,* in a reserve pool such that these resources can be used for avoiding deadlock. The remaining unreserved resources were allocated injudiciously to any process requesting it. All these resource reservation techniques[24, 25, 26, 27], reduce the cost for testing the avoidance condition in the deadlock avoidance techniques. They reserve the resources and hence are not completely blind as the reactive technique of deadlock detection and recovery while allocating the resources. Thus, these techniques reduce the overhead of the avoidance techniques and deadlock frequency as compared to deadlock detection and recovery techniques.

In post-disaster recovery systems, deadlock may imply that the recovery activities are held up which may have catastrophic effects. At the same time,a delayed response in resource allocation can also lead to the unavailability of the critical resources when they are most needed. The existing resource reservation techniques present a promising reduction in the computing overhead for resource management. However, none of them is 100 % deadlock-free.Thus, this paper aims to propose a resource handling policy to ensure that the deadlock never occurs while maintaining a low response time. This work aims to re investigate the resource reservation policies to answer the question of;*how many resources must be reserved and how they must be allocated to avoid deadlock completely?* The correctness of the proposed technique is proved in the form of theorems.

The rest of the paper is organized as follows. Section 2 provides a further look at the existing literature. While Section 3 describes the system model, motivational examples are given in Section 4 to illustrate the various existing techniques for deadlock management. Section 5 presents the proposed Deadlock Free Resource Reservation (DFRR) technique, followed by results and analysis in Section 6. Finally, the paper concludes with Section 7.

2. Related Work. The post-disaster response needs dynamic and instantaneous updates for which IoT is one of the best solutions [10, 11, 12, 13]. Zhang et. al. [10] proposed an agent based resource allocation in emergency management systems considering the severity of the disasters at various levels. Authors [11, 12] studied an IoT- based post-disaster response system and estimated the response time of the resources as they wait to be scheduled using Banker's Algorithm. In [33] uses priority basis stable matching algorithm for effective allocation of resources during disasters using IoT. In a post-disaster response system, the response time is critical and can impact lives and property. In such a system there are limited resources and deadlock of these resources can have catastrophic effects. This paper aims to develop a deadlock-free system with a reduced response time of the resource management system and also estimate the resources needed for establishing an optimal emergency response center.

Deadlock is a well-known problem in computing as well as in all those fields where resources are shared. It attributes to major overhead in the system in both cases: disregarded or managed. In case the deadlock is completely disregarded, it may seize the system. However, the deadlock management has the corresponding overhead, it has been studied over the years. One of the earliest studies used a graph based model proposed by Coffman et. al. [22] for inferring the conditions that must simultaneously hold for the deadlock to occur. There are four strategies to handle deadlocks; DDR (deadlock detection and recovery), DA (deadlock avoidance), DP (deadlock prevention) and RR (Resource Reservation).

DDR is the simplest technique suggesting to 'do nothing', that is till the time the deadlock actually occur do nothing. The DDR technique advocates detecting the deadlock and suggests recovering from it. Holt [21] suggested one of the oldest technique for deadlock detection. This technique represented the resource hold and request as a graph and reduced it for detecting a deadlock.Cheng Xin et.al [30], used threads for deadlock detection in distributed systems. Xiao and Lee [39], suggested a parallel algorithm for deadlock detection on Multi-Processor System-on-a-Chip (MPSoC). Shiu et. al. [40], presented a low cost hardware solution for deadlock detection.DDR techniques though simple yet have the overhead of deadlock detection algorithm running periodically, and may need to restart one or more process for recovering from the deadlock which may not be possible in all the systems. The DP techniques target to prevent one of the necessary conditions from occurring so that the deadlock will not occur.

Deadlock avoidance (DA) methods perform forward calculation based on the knowledge of the resources required by the processes in future to avoid deadlock. Havender [25], suggested to allocate all the resources a process demands in the beginning to avoid deadlock. This method was though effective had serious limitation, because the resources remained underutilized. Other methods suggested by Havender [25] include i) ordering the resources and ii) preempting and re-requesting. The ordering technique ordered the resources in such a fashion that the deadlock would not occur. The resources were preempted and re-requested if sufficient resources were not available when an incremental request was made. Another one of the most renowned DA algorithm is Banker's Algorithm, derived by Dijkstra [19] for a single resource type. Habermann [20] extended this algorithm to multiple resource types. Banker's Algorithm tests all possible allocations and ensures that the request for resources is granted only when no deadlock is foreseen. Lang [28] extended the Banker's algorithm using a control flow graph to determine the resources are released before the control are transferred to the processes in the next region.

Yin et.al. [37] studied the deadlock problem in multithreaded program in a multicore architecture. The authors converted the program source code into a formal model which was used by the discrete control theory to automate the lock acquisition on the resources. They suggested to postpone the resource lock acquisition to avoid deadlock. Wang et. al. [38] used Banker's Algorithm for broadcasting in wireless network, where carrier frequency are one of the resources. Lee and Mooney [31] implemented the DA as a hardware, for a Multi-Processor System-on-a-Chip (MPSoC). Pyla and Varadarajan [42] reiterated the fact that deadlock can be avoided completely only when some prior knowledge of the resource requirement of the processes in the system is available and it is likely to occur if the resources are allocated arbitrarily. They suggested 'Sammati'a tool for threaded applications using POSIX for automatic deadlock detection and recovery. Youming Li et. al. [43] proposed a modification of Banker's algorithm (BA). Permutation matrices were used for each resources and the process was selected greedily. However, their space requirement was more than the original Banker's Algorithm.

Kawadkar et. al. [44] extended the Banker's algorithm by examining the processes in the waiting queue. A process entered the waiting queue when its requested resources are either unavailable or could not be granted.They suggested to pick a process from the waiting queue based on the resources it is holding and needing. However, authors do not mention how selected process from the waiting queue should be allocated the resources such that the deadlock is avoided. Dixit and Khuteta [45] suggested to change the resource requirement at the run-time to prevent a possible deadlock.

All the deadlock avoidance techniques have high computation time as they must perform some test before each allocation to avoid deadlock. The Resource Reservation (RR) techniques eliminated this test for deadlock avoidance and thus, reduce the turnaround time of the processes. The RR techniques suggest to reserve some resources that can be used to avoid the deadlock. Botlagunta et. al. [25] first introduced this idea and suggested to reserve resources based on a threshold. Agrawal et. al. [24] suggested to estimate the resources to be reserved based on the sum of the resources needed further.

Authors [26] suggested another way based on the shortest execution time process. The process with minimum worst case execution time reserved the resources as suggested in [26]. Shubham et.al [36] extended the technique suggested in [26] to reserve the resources based on the remaining resource need. Botlagunta et. al. [27] extended the RR techniques to dynamically allocate a budget of resources to each incoming process and perform resource reservation as per the technique suggested in [25]. These techniques are discussed in detail in the subsequent section.

Table 2.1
Deadlock Handling Methods

Deadlock detection and recovery (DDR)	Deadlock Prevention (DP)	Deadlock Avoidance (DA)	Resource Reservation(RR)
Reactive to deadlock, does nothing to handle it in advance	Prevent one of the four necessary conditions from occurring	Look forward to avoid any deadlock in future. Perform Safety Sequence Test	Reserve some resources for avoiding the deadlock

Both DA and RR techniques presume that the system resource requirement is known in advance. The DA techniques perform some kind of test for ensuring that the deadlock will not occur in the future. However, the RR techniques do not perform any test rather use the knowledge of the resource requirement for estimating the portion of resources that must be reserved to avoid deadlock. The resource reserved by different RR Techniques existing so far is not sufficient for completely avoiding the deadlock. The present work extends the idea of the RR techniques to use the knowledge of the resource requirement by the processes for estimating the reserve pool strength. The proposed technique ensures that all the processes have the opportunity to complete with the reserved resources.

3. System Model. The post-disaster management system resource pool is assumed as a finite set of $m$resource types represented as $R_1,\ R_2,\ R_3 \ldots R_m$, with instances as $\alpha_1,\ \alpha_2,\ \alpha_3 \ldots \alpha_m$ of each. Further, the post-disaster recovery activities are called processes $P_1,\ P_2,\ P_3 \ldots P_n$ with attributes arrival and worst-case execution time represented as $(a_i,\ e_i)$. The state of the system resources is represented in the data structures from Table 3.1.

4. Motivational Example. Motivational examples are presented in this section to analyze the existing techniques.

Table 3.1
Notations used in System

Notations	**Meaning**
P $_i$	ith process
R $_j$	jth resource
alpha$_j$	number of instances of resource type R $_j$
m	Number of resource types
n	Number of processes.
Request[i][j])	A process maximum resource requirement is indicated by Max[i][j]. However, a process does not need all the resources at once but incrementally. The resources requested by the processes at any time instance is represented as a two-dimensional array of size n by m. Thus, Request[i][j]= k, where k= 0, 1,. . . Max[i][j].
Allocation[i][j]	It is a two dimensional n by m array containing the resources assigned to each process.
Need[i][j]	It is a $n \times m$ array, containing the estimated resources further required by processes. Mathematically, $Need\,[i]\,[j] = Max\,[i]\,[j] - Allocation\,[i]\,[j]$
Available[j]	A subset of the resource pool consisting of resources that can be allocated to the requesting process. It is represented by an array of m elements. It is initializes as $Available\,[j] = \alpha_j - Reserve\,[j]\,, \forall R_j j = 1, 2, 3\ \ldots m$
Reserve[j]	A subset of the resource pool is marked as reserved. It contains the quantity of each resource type reserved. It is represented as an array with $m$elements. Thus, $Reserve[j]$ equals k where k indicate that k instance of the resource of type R_j are reserved.
Max[i][j]	Maximum resource requirement
Throughput	the number of processes completed per unit of time.
Turnaround time(TT)	the difference between submission of a process and its completion.
Average turnaround time(ATT)	The total turnaround time divided by the number of processes.
Safety Sequence (SS) [19]	A safety sequence was suggested by Banker's algorithm is an order in which the processes can be completed without deadlock.
Safe State and UnSafe State[19]	A system for which a safety sequence exists ensures that the deadlock will not occur is said to be in a Safe State otherwise in Unsafe state

Table 4.1
System initial state at start time t_0

	Allocation				Maximum				Request			
	R$_1$	R$_2$	R$_3$	R$_4$	R$_1$	R$_2$	R$_3$	R$_4$	R$_1$	R$_2$	R$_3$	R$_4$
P$_1$	0	0	0	0	0	0	1	2	0	0	0	0
P$_2$	0	0	0	0	2	7	5	0	0	2	1	0
P$_3$	0	0	0	0	6	6	5	6	4	3	1	2
P$_4$	0	0	0	0	4	3	5	6	1	0	2	4
P$_5$	0	0	0	0	0	6	5	2	0	2	2	0
Available												
					R$_1$	R$_2$	R$_3$	R$_4$				
					8	13	11	10				

Table 4.2
System state at time t_1 after allocation of resources requested at time t_0

	Allocation				Need				Request			
	R$_1$	R$_2$	R$_3$	R$_4$	R$_1$	R$_2$	R$_3$	R$_4$	R$_1$	R$_2$	R$_3$	R$_4$
P$_1$	0	0	0	0	0	0	1	2	0	0	0	0
P$_2$	0	2	1	0	2	5	4	0	0	2	1	0
P$_3$	4	3	1	2	2	3	4	4	0	0	0	0
P$_4$	1	0	2	4	3	3	3	2	1	0	0	0
P$_5$	0	2	2	0	0	4	3	2	0	0	0	0
Available												
					R$_1$	R$_2$	R$_3$	R$_4$				
					3	6	5	4				

Example 1 [9]: Consider a system consisting of resources as $\{R_1,\ R_2,\ R_3, R_4\} = \{8, 13, 11, 10\}$ is to implement processes P_1, P_2, P_3, P_4, P_5 with resource requests as illustrated in Table 4.1 for time t$_0$. Table 4.2 indicates the system state as the resources are allocated. Table 4.3 enlists a hypothetical sequence of resource requests made by the processes subsequently. The existing techniques discretion on these requests is as follows:

1. ***Deadlock Avoidance Banker's Algorithm(BA) [19]:***
   The requests made as per (refer Table 4.3) are considered by BA by estimating the safety sequence. Thus, request of P_2, $P_4$and P_3 are granted but the last request ofP_2 is not granted.
2. ***Deadlock Detection and Recovery (DDR) [23]:***
   This is a reactive technique where no pretesting is done to ensure that a deadlock will not occur. It will simply grants every possible request if sufficient resources are available as and when they are made. In the present example,all requests as listed in Table 4.3 are granted. The resulting state of the system is shown in Table 4.4. The system does not enter into deadlock immediately as the need of process P_1 can be satisfied. Once the process P_1 completes the state of the system is as shown in Table 4.5. This state is referred to as Unsafe state (defined in Sect.3), as it leads to a deadlock. Substantial overhead is incurred in detecting and recovering from it.
3. ***Worst-Case Execution Time Based Resource Reservation (ETRR) Technique [26]:***
   This technique is motivated by shortest job first scheduling assigning highest priority to a process with smallest the worst case execution time. In this example priorities are assigned as $\langle P_1,\ P_5,\ P_2,\ P_4,\ P_3\rangle$, indicating that the computation time of process P_1 is least. Hence, resources (0, 0, 1, 2) are placed in the reserve pool. All the requests as an when they arrive (refer Table 4.3) are granted by this technique also. Finally, at time t_4 when process P_1 completes, the system state can be seen in Table 4.5. Thus, this technique also fails to avoid the deadlock subsequently.
4. ***Threshold based Resource Allocation (TRA) Technique [25]:***
   The resources are reserved by this technique based on a threshold estimated as $Threshold[j] = \lceil \lfloor Need[i][j] \forall i = 1, 2 \ldots n \rfloor, 0 \rceil$. Accordingly, for this example, it reserves (2, 3, 1, 2) resources at the onset. At time t_1, when P_2 and P_4 request for additional resources, both the requests are granted. Thereafter, the resources in the reserve pool are (2, 3, 1, 2) while that in the available pool are (0, 1, 3, 2). However, at time t_2 when process P_3 requests for (2, 0, 0, 2) resources, sufficient resources are

Table 4.3
Sequence of requests

Time at which request is made	Process requesting for the resources	Request (R_1, R_2, R_3, R_4)
t_1	P_2	(0,2, 1, 0)
t_1	P_4	(1,0, 0, 0)
t_2	P_3	(2,0, 0, 2)
t_3	P_2	(0,2, 1, 0)

Table 4.4
Snapshot after allocation to P_2, resources (0, 2, 1, 0) at time t_3

	Allocation				Maximum				Need			
	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
P_1	0	0	0	0	0	0	1	2	0	0	1	2
P_2	0	6	3	0	2	7	5	0	2	1	2	0
P_3	6	3	1	4	6	6	5	6	0	3	4	2
P_4	2	0	2	4	4	3	5	6	2	3	3	2
P_5	0	2	2	0	0	6	5	2	0	4	3	2
Available												
				R_1	R_2	R_3	R_4					
				0	2	3	2					

Table 4.5
Snapshot at time t_4 after completion of process P_1

	Allocation				Maximum				Need			
	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
P_1	Execution Complete											
P_2	0	6	3	0	2	7	5	0	2	1	2	0
P_3	6	3	1	4	6	6	5	6	0	3	4	2
P_4	2	0	2	4	4	3	5	6	2	3	3	2
P_5	0	2	2	0	0	6	5	2	0	4	3	2
			Available				Reserve					
			R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4		
			0	2	3	2	0	0	0	0		

Table 4.6
System initial state at time t_0 for Example 2

	e_i	Allocation		Maximum		Request		Available	
		R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2
P_1	10	0	0	2	4	1	2	4	4
P_2	2	0	0	4	2	2	1		

Table 4.7
System state after request granted to process P_2

	Allocation		Maximum		Need		Available	
	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2
P_1	1	2	2	4	1	2	1	1
P_2	2	1	4	2	2	1		

Table 4.8
System state after request granted to P_1

	Allocation		Maximum		Need		Available		Reserve	
	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2
P_1	2	4	2	4	0	0	0	0	2	0
P_2	0	0	4	2	4	2				
Request of process P_1 is granted and P_2 is denied. safety sequence is $\langle$ P_1, P_2 $\rangle$										

Table 4.9

System state with reserve pool after request granted to process P_2

	Allocation		Maximum		Need		Available		Reserve	
	R$_1$	R$_2$	R$_1$	R$_2$	R$_1$	R$_2$	R$_1$	R$_2$	R$_1$	R$_2$
P$_1$	0	0	2	4	0	0	0	2	2	1
P$_2$	2	1	4	2	2	1				
Request of process P$_1$ is denied and P$_2$ is granted. Safety sequence is ⟨ P$_2$, P$_1$ ⟩										

Table 4.10

a) Process P$_1$ requests (1, 2) resources							
R$_j$	Request	Need	Available	Threshold	X	Y	Can_grant
R$_1$	1	2	4	2	1	1	1
R$_2$	2	4	4	2	0	1	1
granted							
b) Process P$_2$ requests (2, 1) resources							
R$_j$	Request	Need	Available	Threshold	X	Y	Can_grant
R$_1$	2	4	3	1	0	1	1
R$_2$	1	2	2	2	1	0	1
granted							

not available in the pool. Hence, the reserve pool resources are used and (2, 3, 4, 4) are granted. No deadlock occurs for this example.

5. ***Total Need Based Resource Reservation (TNRR) Technique:[24]***
   The total number of resources neededby requesting processes is considered for reserving resources by this technique. Mathematically, $Reserve\,[j] = \lceil\frac{(Total\_Need - \alpha_j)}{n}\rceil$ where $Total\_Need\,[j] = \sum_{i=1}^{n} Max\,[i]\,[j]$, α_j and $Reserve\,[j]$ are defined in Sect. 3.

In the motivational example 1, $Total\_Need\,[j]$ is estimated as 12, 22, 21, 16 for the resources R_1, R_2, R_3 and R_4 respectively. Thus, (2, 3, 2, 2) resources are reserved at the on set. At time t_1, when P_2 and P_4 request for additional resources, both the requests are granted without using any of the reserved resources. However, at time t_2 when process P_3 requests for (2, 0, 0, 2) resources, only (0, 1, 2, 2) resources are in the available pool which are not sufficient. The resources from the reserve pool are used to satisfy need of the process. No deadlock occurs for this example.

This example illustrated scenario where the existing priority based techniques failed to reserve sufficient resources to avoid deadlock. The following example demonstrates the limitation of the other techniques that survived in this example.

Example 2: Suppose in system processes P_1, P_2 are have an upper bound of 4 instances of each resources R_1 and R_2 illustrated in the Table4.6 request for (1, 2) and (2, 1) resources are made by these processes respectively. The assessment on these requests of the existing techniques is as follows:

1. ***Deadlock Avoidance Banker's Algorithm (BA) [19]:***
   The Banker's Algorithm calculates the safety sequence for each request. It then grants P_1's request but denies P_2 's request, refer to Table 4.7, to avoid deadlock in future.
2. ***Deadlock Detection and Recovery (DDR) [23]:***
   DDR grants every request indiscreetly and does the same in this case too. The resulting state is as in the Table 4.7, which leads to a ***Deadlock eventually occurs***.
3. ***Worst-Case Execution Time Based Resource Reservation (ETRR) Technique[26]***
   The ETRR technique reserves the resources for process P_2 , because it is the shortest job(cf.Table 4.6). The reserve and available pools thus, contain (4, 2) and (0, 2) resources respectively. When the process P_1 requests for the resources (1, 2), it is denied by the system. However, the request by process P_2 is granted, refer Table 4.9. The system is in Safe state.

Table 4.11
Analysis of the example 1 and 2

Technique	**Occurrence of deadlock in Example**	**Occurrence of deadlock in Example 2**
BA	Deadlock does not occur	Deadlock does not occur
DDR	Deadlock occurs	Deadlock occurs
ETRR	Deadlock occurs	Deadlock does not occur
TRA	Deadlock does not occur	Deadlock occurs
TNRR	Deadlock does not occur	Deadlock occurs
DFRR(proposed method)	Deadlock does not occur	Deadlock does not occurs

4. ***Threshold based Resource Allocation (TRA) Technique [25]:*** At the inception the threshold is (2, 2) based on which the request of process P_1 is granted shown in4.10 a). It is then updated as (1, 2) (refer to table 10 b)) and the request of P_2 is also granted. The system reaches the state in the 4.7 which indicate that the ***Deadlock will occur***.
5. ***Total Need Based Resource Reservation (TNRR) Technique[24]:*** The resources to be reserved are estimated as $Reserve\,[j] = \lceil\frac{(Total\_Need - \alpha_j)}{n}\rceil$ as $Reserve\ [1] = \lceil\frac{(6-4)}{2}\rceil$ and $Reserve\,[2] = \lceil\frac{(6-4)}{2}\rceil$, i.e., (1, 1). The available pool will contain the remaining (3, 3) resources, sufficient for the process P_1 and P_2 request ((1, 2) and (2, 1) respectively). Thus, the requests are granted leading the system into the unsafe state as shown in 4.7.

Table 4.11 summarizes the effectiveness of the existing techniques in maintaining a deadlock free system for the motivational examples.The DDR technique performs no testing before granting any resource to a requesting process and hence, ends up into a deadlock most often. The overhead saved by eliminating the test before granting of resources to a requesting process is consumed by the periodic testing required for deadlock and recovery whenever it occurs. On the other hand, the BA algorithm which ensures that deadlock will never occur is too costly. The sub optimal techniques (ETRR, TRA and TNRR) are able to prevent the deadlock in some case (table 4.11) but are not full proof. This paper proposes a resources reservation technique which is deadlock free (proved in the form of a theorem). It also lowers the overhead incurred by the BA.

The following section presents the proposed technique for a deadlock free system using resource reservation strategy.

5. Proposed Deadlock Free Resource Reservation Techniques (DFRR). The above section analyzed the performance of the prevailing techniques and summarizes them in the table 4.11. The table 4.11 reveals that the Banker's algorithm calculates the safety sequence before allocating the resources on every request. The cost of such an estimation is $O\left(mn^2\right)$ where m is the different type of resources and n is the number of processes in the system. This overhead is considerable for even a small system. On the other hand, newer RR techniques such as ETRR, TRA and TNRR are cost effective but not 100% deadlock free. A closer look at the motivational example 1, illustrates that ETRR is process oriented, i.e., reserve resources for a process P_1 and guarantee that it will complete successfully and it does. However, process P_1reserves only (0,0, 1, 2) resources, i.e., it reserves only one and two instances of resource type R_3 and R_4 respectively. While all the instances of resources R_1 and $R_2$are unreserved. The remaining resources are claimed by the rest of the processes leading the system into an Unsafe state. The resources released by the process P_1on its completion are not sufficient for any other process and the deadlock will eventually occur. In other words, ETRR although reserve resources for one of the process but they do not restrict the remaining processes from entering into Unsafe state.

The TRA and TNRR techniques are resource oriented, where reservation of resources is done such that there is at least one process requiring that resource. However, a process may require other resources as well that may not be available in sufficient quantity for its successful completion. The example 2, Sect.4, illustrates that sufficient instances of R_1 are available for processes P_1, while for process P_2 ample instances of resource

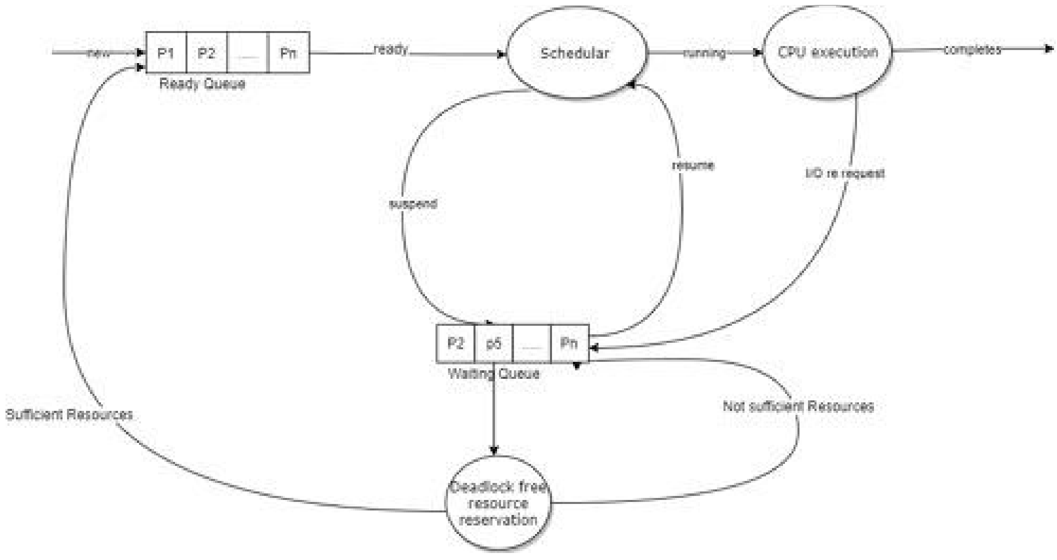


FIG. 5.1. *Proposed System Model.*

R_2 are available. However, enter into a circular wait as they wait for the other to release the resources they are holding.

This paper proposes a deadlock free resource reservation (DFRR) technique which ensures that deadlock will never occur. The system model for this DFRR can be seen in the Fig.5.1. It suggests to reserve some resources in the reservation pool as per the theorem 5.2. The remaining resources are made available in the available pool. Whenever a process requests for a set of resources, available pool resources are granted (as per theorem 5.2). In case the request cannot be granted only from the available pool then the reservation pool are also used such that the maximum resources this process can ever want are granted to give this process the chance to complete.

The theorem 5.1,claims that if a process is allocated all the resources it will ever need then the system will be deadlock free again if it was deadlock free before this allocation. Theorem 5.2, proves the deadlock free of the proposed technique.

Theorem 5.1 is stated as follows.

THEOREM 5.1. *A system consisting of set of resource types as $R_1, \quad R_2, \quad R_3 \ldots R_m$ with $\alpha_1, \quad \alpha_2, \quad \alpha_3 \ldots \alpha_m$ instances of each type, when scheduling a set of independent processes $P_1, \quad P_2, \quad P_3 \ldots P_n$, if is in a safe state will remain in safe state even after allocating all the resources ever needed by a requesting process P_{req} from the set.*

Proof. Suppose at any time t the system is in safe state, then the safety sequence can be estimated as

$$\langle P_b, \quad P_{req}, \quad P_a \rangle \tag{5.1}$$

where $P_b = P_{b1}, \ P_{b2} \ldots P_{bx}$ and $P_a = P_{a1}, \ P_{a2} \ldots P_{ay}$ are the set of processes in the safety sequence before and after the requesting process P_{req} .

Let, the resources available at this time t be

$$\omega\left(t\right) = Available\left[j\right] + Reserve\left[j\right] \quad \forall j = 1 \ldots m \tag{5.2}$$

Processes are in the safety sequence, this implies:

$$Need\left[b1\right]\left[j\right] \leq \omega\left(t\right); \quad \forall j = 1 \ldots m \tag{5.3}$$

It will complete and release all it's resources. Similarly, the remaining processes will complete. Reconsider if at this time t a process P_{req} needs

$$Need\,[req]\,[j] = Max\,[req]\,[j] - Allocation\,[req]\,[j] \tag{5.4}$$

resources to complete.

In worst case, initially, no resources are allocated to this requesting process P_{req}, i.e.,

$$Allocation\,[req]\,[j] = 0 \quad \forall j = 1 \ldots m \tag{5.5}$$

Therefore, substituting (Eq.5.5) in (Eq.5.4)
$Need\,[req]\,[j] = Max\,[req]\,[j] \quad \forall j = 1 \ldots m$

$$Need\,[req]\,[j] \leq \omega\,(t) \quad \forall j = 1 \ldots m \tag{5.6}$$

If Eq.5.6 holds, then all the needed resources can be granted to this requesting P_{req} process. In worst case (Eq.5.6 will be)

$$Need\,[req]\,[j] = \omega\,(t) \quad \forall j = 1 \ldots m \tag{5.7}$$

indicates system has just sufficient resources needed by the process to complete. These resources are thus allocated to the requesting process P_{req}; mathematically,

$$Allocation\,[req]\,[j] = Allocation\,[req]\,[j] + Need\,[req]\,[j]$$

Substituting (Eq.5.5) and (Eq.5.7)

$$\Rightarrow Allocation\,[req]\,[j] = \omega\,(t) \quad \forall j = 1 \ldots m \tag{5.8}$$

implying that all needed resources are allocated, thus, the need can be updated as

$$Need\,[req]\,[j] = 0, \quad \forall j = 1 \ldots m.$$

The process P_{req} can thus, complete at time t1. It then releases

$$Allocation\,[req]\,[j] = \omega\,(t)\,, \quad \forall j = 1 \ldots m$$

all the resources allocated to it as per equation (Eq.5.8).

Therefore, the resources in the system will again be restored to be $\omega\,(t)$. These are sufficient for process P_{b1} (Eq.5.1) as suggested by equation (Eq.5.3).

The new safety sequence is $\langle P_{req},\ P_b,\ P_a \rangle$.
Thus, system remains in the safe state. Hence proved. □

The theorem 5.1, ensures that if a process P_i is granted all the resources it may further need to complete, i.e. $Need\,[i]\,[j] = Max\,[i]\,[j] - Allocation\,[i]\,[j]$, then the system will remain/return to safe state after it's completion. Now, the question is:*does the system has the Need[i][j] number or resources?* The theorem 5.2, ensures that every process in the system can acquire the needed resources, and ensure the system is deadlock free. The theorem 5.2 can be stated as follows.

Theorem 5.2. *A set of independent processes $P_1,\ P_2,\ P_3 \ldots P_n$ when scheduled on a system consisting of set of resource types as $R_1,\ R_2,\ R_3 \ldots R_m$ with $\alpha_1,\ \alpha_2,\ \alpha_3 \ldots \alpha_m$ instances of each type, with Max[i][j] as the maximum number of resources of type R_j demanded by the i^{th} process P_i during its execution time can be done without deadlock if the following conditions hold:*

1. $\forall R_j$ system reserves

$$Reserve[j] = \begin{cases} 0 & \eta_j \leq 1 \\ 0 & \sum_{i=1}^{n} Max[i][j] \leq \alpha_j \\ \left\lceil \frac{\left(\sum_{i=1}^{n} Max[i][j]\right) - \alpha_j}{\eta_j - 1} \right\rceil & \sum_{i=1}^{n} Max[i][j] \geq \alpha_j \end{cases}$$

instance of its type forming a reserve pool, where $\eta_j = count\,(Max\,[i]\,[j] \neq 0, \forall i = 1, 2 \ldots n)$ is the number of processes requesting for the resource R_j .Leaving the resources in the available pool as $Available\,[j] = \alpha_j - Reserve\,[j]$

2. $\forall P_i$, $Max\,[i]\,[j] - Reserve\,[j] \geq 0$

3. A process P_i can acquire either

(a) $Max\,[i]\,[j] - Reserve\,[j]$ resources from the available pool,

(b) $Max\,[i]\,[j]\,\forall R_j$ from available and reserve pool put together such that the available pool is consumed first. That is a process must acquire all the resource it might need.

4. A process P_i on completion relinquishes all the resources, $Max\,[i]\,[j]\,\forall R_j$, that are returned to the reserve and available pools as per condition 1.

Proof. A process P_i can acquire resources either partially as per condition 3.a. or completely as per condition 3.b. Thus, some processes may acquire partial resources ($Allocation\,[i]\,[j] = Max\,[i]\,[j] - Reserve\,[j]$) and some complete set ($Allocation\,[i]\,[j] = Max\,[i]\,[j]$) .

Processes that acquire resources as per condition 3.b.

$$(Allocation\,[i]\,[j] = Max\,[i]\,[j]\,;\ Need\,[i]\,[j] = Max\,[i]\,[j] - Allocation\,[i]\,[j] = 0)$$

have all the requisite resources and will complete and relinquish the resources as per condition 4. The system remains in safe state as per theorem 5.1.

In worst case all the processes acquire resources partially as per the condition 3. a. and none of the process can complete. Mathematically,

$$Allocation\,[i]\,[j] = Max\,[i]\,[j] - Reserve\,[j] \quad \forall P_i \tag{5.9}$$

Applying summation to equation (cf.Eq.5.9); the total number of instances of type R_j , allocated to any process is thus, $\sum_{i=1}^{n} Allocation\,[i]\,[j] = \sum_{i=1}^{n} (Max\,[i]\,[j] - Reserve\,[j])$ where $Max\,[i]\,[j] -$ $Reserve$ $[j]$ is the number of resource type R_j allocated to process P_i .

Since, $Reserve\,[j]$ is independent of the process, the total number of instance already allocated are

$$\sum_{i=1}^{n} Allocation\,[i]\,[j] = \left(\sum_{i=1}^{n} Max\,[i]\,[j]\right) - n * Reserve\,[j] \tag{5.10}$$

Consider $Reserve\,[j] = \lceil \frac{(\sum_{i=1}^{n} Max[i][j]) - \alpha_j}{\eta_j - 1} \rceil$ as per condition 1., here, $\eta_j = n$ in worst case, indicating a resource is demanded by all the processes in the system, leading to higher conflicts. Therefore, $(n-1) * Reserve\,[j] = (\sum_{i=1}^{n} Max\,[i]\,[j]) - \alpha_j$ implies

$$n * Reserve\,[j] = \left(\sum_{i=1}^{n} Max\,[i]\,[j]\right) - \alpha_j + Reserve\,[j] \tag{5.11}$$

Substituting, $Available\,[j] = \alpha_j - Reserve\,[j]$, as per condition 1., in equation (Eq.5.11) implies

$$n * Reserve\,[j] = \left(\sum_{i=1}^{n} Max\,[i]\,[j]\right) - Available\,[j] \tag{5.12}$$

Substituting (Eq.5.12) in (Eq.5.10), the number of instance allocated to all the processes put together is $\sum_{i=1}^{n} Allocation\,[i]\,[j] = (\sum_{i=1}^{n} Max\,[i]\,[j]) - (\sum_{i=1}^{n} Max\,[i]\,[j]) - Available\,[j]$, implying:

$$\sum_{i=1}^{n} Allocation\,[i]\,[j] = Available\,[j]\,, \forall R_j$$

That is the number of instance allocated to all the processes put together are equal to $Available\,[j]\,, \forall R_j$. In other words, at most available pool is emptied by the processes, however, reserve pool resources are still unused. Since, every process has acquired, as per equation (Eq.5.9), $Allocation\,[i]\,[j] = Max\,[i]\,[j] - Reserve\,[j]$, hence,

$$Need\,[i]\,[j] = Max\,[i]\,[j] - Allocation\,[i]\,[j]$$

$$\Rightarrow Need\,[i]\,[j] = Max\,[i]\,[j] - (Max\,[i]\,[j] - Reserve\,[j])$$

$$\Rightarrow Need\,[i]\,[j] = Reserve\,[j]\,, \quad \forall R_j$$

Therefore, every process needs only $Reserve\,[j]\,, \forall R_j$ more resources to complete which are available in the reserve pool. Thus, every process has equal opportunity to complete.Once any process completes it will release any resources it was holding (at least $Reserve\,[j]\,\forall R_j$), which are sufficient for the subsequent processes to complete (Theorem 5.1). Hence, there is no deadlock or starvation. Hence proved. □

The theorem 5.2 guarantees that the deadlock will never occur if the system follows it. The deadlock free resource reservation (DFRR) technique is proposed based on these theorems it is stated in the form of the DFRR algorithm in cf.Alg.1. The time complexity for resource management to decide on the resource allocation as per the request of the proposed algorithm is $O\,(m)$. The resource manager only refers the available pool for the requested resources incurring $O\,(m)$ overhead. In case the sufficient resources are not available in the Available pool the same are granted from the Reserve pool. This resource management policy is optimal and is proved in the form of lemma 5.3.

Lemma 5.3. *No resource management policy can assign resources to a requesting process from a set of independent processes* $P_1, P_2, P_3 \ldots P_n$ *when scheduled on a system consisting of set of resource types as* $R_1, R_2, R_3 \ldots R_m$ *with* $\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_m$ *instances of each type, with an overhead lower that* $O\,(m)$.

Proof. Each Process can request any resource $R_1,\ \ R_2,\ \ R_3 \ldots R_m$ its request vector is thus of length 'm'. Hence, any resource management technique will check for the request and availability of all the 'm' resources at least once for a process. Hence, the overhead will not be less than $O\,(m)$. □

The motivational examples in Sect. 2 are revisited to show the effectiveness of the proposed technique. The resources reserved for the motivational example 1 (Table 4.1) are (2, 3, 3, 2). Requests in Table 4.1 are granted followed by the request made by the processes P_2 and P_4 as at time t_1 as stated in table using resources from the available pool. However, the request of process P_3 and P_2 at time t_2 and t_3 cannot be granted as per the condition 3.a theorem 5.2. The system remains in the safe state.

Deadlock Free Resource Reservation Technique, reserves (2, 2) resources, for the motivational example 2, cf. Table 4.6. Hence, either process P_1 acquires the necessary resources and completes as per Table 4.8 or as per Table 4.9, P_2 completes and relinquishes the resources. Hence, no deadlock.

The following section presents the results obtained on implementation of the proposed technique.

6. Simulations Results. Simulations were performed on process sets to analyze the behavior of the proposed DFRR technique as compared to the existing techniques. The average time of the each process spends in the system from its submission to its completion also known as*Average Turnaround time* is the key parameter of the system.

Algorithm 1: DFRR Algorithm

```
   input: Process Priority Queue
 1 begin
 2     for $j$=1 to $m$ do
 3         Reserve[j]=$\left\lceil \frac{\left(\sum_{i=1}^{n} Max[i][j]\right)-\alpha_j}{\eta_j-1} \right\rceil$
 4         Available[j]=$\alpha_j$-Reserve[j]
 5     end
 6     while No new Request do
 7         Execute a ready process;
 8     end
 9     A process P$_i$ request for Request[i][j] resources do
10         if (Allocate[i][j]+Request[i][j])$\leq$(Max[i][j]-Reserve[j])$\forall j$=1,2...m) then
11             Allocate[i][j] = Allocate[i][j] + Request[i][j];
12             Available[j ]= Available[j] - Request[i][j];
13             Need[i][j]=Need[i][j]-Request[i][j];
14         end
15         else if Need[i][j]$\leq$(Available[j]+ Reserve[j])$\forall j$=1,2...m then
16             Allocate[i][j]=Allocate[i][j]+ Need[i][j];
17             if Need[i][j]Available[j] then
18                 Available[j]= Available[j]-Need[i][j];
19             end
20             else
21                 Reserve[j]=Reserve[j]+ Available[j]-Need[i][j];
22                 Available[j]=0;
23             end
24             Need[i][j]=0;
25         end
26     end
27     If a process $P_i$ completes then return the resources and goto to step 5.
28 end
```

All simulations are performed on a 2.0 GHz Processor. With a Resource pool of 10 resources with instances as { 11, 14, 7, 2, 16, 8, 15, 17, 13, 5}. 100 process sets for each point in the graph were generated with each process randomly selects a execution time between 0 to 50.

Fig.5.1, illustrate the effect on average turnaround time as the process load increases. As process load increases the chances for the resource conflict also increases and hence, the average turnaround time. The analysis of the proposed DFRR reveals that the turnaround time of the processes is relatively better than the existing for all process loads. The improvement in the performance is even more profound for the higher loads. This is because as the load increases so does the resource demand by the processes, leading to frequent resource requests. The Banker's algorithm has a high overhead for serving each request and hence the turnaround time increases as the load increases for it. On the other hand, the existing RR and DDR technique face more frequent deadlocks and spend time in recovering from the deadlock, which in turns increases the turn around time. The proposed DFRR technique has lower overhead for resource management and is deadlock free. The average turnaround time is approximately 18 % lower for the proposed DFRR over Banker's algorithm.

A process can request all the resources it could need at a time or incrementally as its execution processed. Fig.6.1 illustrates the effect of the number of incremental resource requests (Steps) made by the process set in its execution on the average turnaround time of the processes. The process load is fixed to be 0.8 for this figure. The Steps ranges from 1 to 10, where Steps =1 indicates that the processes requests all the resources at the

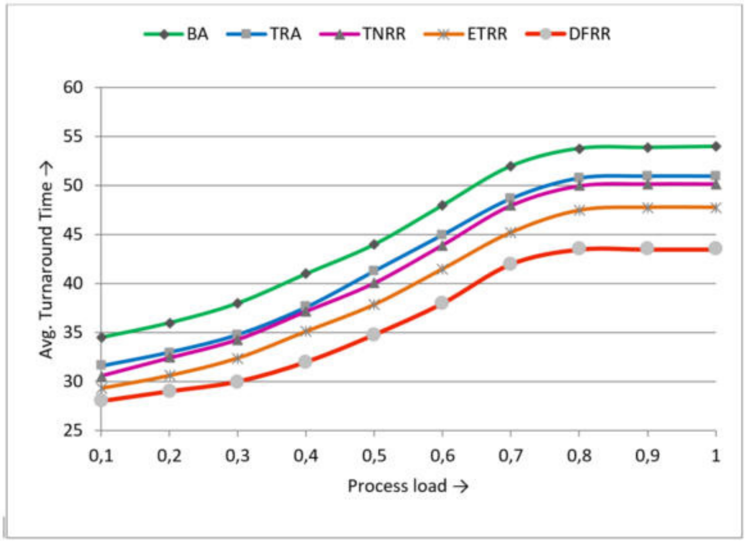


Fig. 6.1. *Process load vs. average turnaround time.*

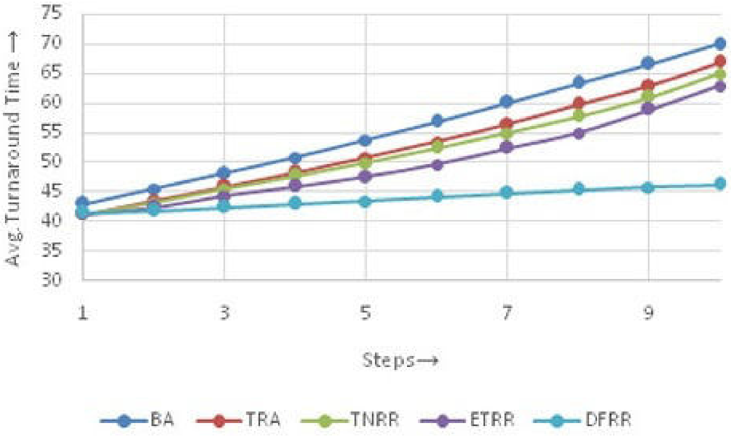


Fig. 6.2. *Steps vs average turnaround time.*

inception while 2, 3. . . .10 indicate that the processes made 2 or 3 . . . 10 incremental resource requests. BA has high overhead for resource management. Thus, as the processes make multiple incremental requests the BA spends more and more time in deciding on the resource grant. Moreover, the process whose request is denied by the BA enter into pending state which increases their wait time and hence, the average turnaround time of the set increases. The other TRA, TNRR and ETRR techniques though have lower overhead for resource management are not 100 % deadlock free. As the Steps increases the number of deadlocks encountered by them also increases incurring deadlock resolution overhead. The proposed DFRR technique has low overhead for resource management as well as it is deadlock free, hence the average turnaround time for this technique remains fairly constant.

7. Conclusion. This paper presented an IoT based post-disaster recovery technique for managing the limited resources available such that the system is deadlock free. A new class of deadlock handling technique which handle deadlock by reserving a portion of the resources was proposed. This class eliminates the excessive overhead incurred by the Deadlock Prevention and Avoidance techniques as well as the uncertainty of deadlock occurrence in the deadlock detection and recovery techniques. The proposed Deadlock Free Resource Reservation (DFRR) technique ensures that deadlock will never occur. The correctness of the proposed technique is proved in the form of the theorems. Its effectiveness is shown through motivational examples. The simulation analysis of the proposed DFRR indicate that it has approximately 18 % lower turnaround time than the existing Banker's algorithm. Thus, the proposed technique is a deadlock free technique with optimal overhead.

REFERENCES

[1] M.Wahlstromand D. Guha-Sapir, The Human Cost of Weather-Related Disasters 1995-2015, UNISDR, Geneva, Switzerland,2015.

[2] Solanki, A., Nayyar, A, Green internet of things (G-IoT): ICT technologies, principles, applications, projects, and challenges. In Handbook of Research on Big Data and the IoT (pp. 379-405), IGI Global, 2019

[3] Singh, S. P., Nayyar, A., Kumar, R., Sharma, A. Fog computing: from architecture to edge computing and big data processing. The Journal of Super computing, 75(4), 2019

[4] D.Rathee, K.Ahuja, A.Nayyar, Sustainable future IoT services with touch-enabled handheld devices. Security and Privacy of Electronic Healthcare Records: Concepts, Paradigms and Solutions, 131,2019

[5] Krishnamurthi. R, Nayyar. A, Solanki. A, Innovation Opportunities through Internet of Things (IoT) for Smart Cities. In Green and Smart Technologies for Smart Cities (pp. 261-292). CRC Press,2019

[6] Nayyar, A., Rameshwar, R., Solanki, A. Internet of Things (IoT) and the Digital Business Environment: A Standpoint Inclusive Cyber Space, Cyber Crimes, and Cybersecurity, The Evolution of Business in the Cyber Age, pp.111-152. 2020

[7] Pramanik. P. K. D, Solanki. A, Debnath. A, Nayyar. A, El-Sappagh. S, Kwak. K. S, Advancing Modern Healthcare With Nanotechnology, Nanobiosensors, and Internet of Nano Things: Taxonomies, Applications, Architecture, and Challenges. IEEE Access,vol 8,2020

[8] Balaji. B. S, Raja.P.V, Nayyar.A, Sanjeevikumar. P, Pandiyan.S, Enhancement of Security and Handling the In conspicuousness in IoT Using a Simple Size Extensible Blockchain. Energies, 13(7), 2020

[9] G. M. Lee, N. Crespi, J. K. Choi, and M. Boussard, Internet of things, in Evolution of Telecommunication Services, pp. 257–282,Springer, 2013.

[10] J. Zhang, M. Zhang, F. Ren, and J. Liu, An innovation approach for optimal resource allocation in emergency management, IEEE Transactions on Computers, 2016.

[11] J. Satishkumar, Mukesh A.Zaveri, Resource Scheduling for Postdisaster Management in IoT Environment, Hindawi Wireless Communications and Mobile Computing,2019.

[12] L. Yang, S. H. Yang, and L. Plotnick, How the internet of things technology enhances emergency response operations, Technological Forecasting & Social Change, vol. 80, no. 9, pp. 1854–1867, 2013.

[13] L.Khubnani, MS thesis, https://www.cs.rit.edu/usr/local/pub/GraduateProjects/2161/lhk3416/ Report.pdf

[14] A. Meissner, T. Luckenbach, T. Risse, T. Kirste, and H. Kirchner, Design challenges for an integrated disaster management communication and information system, in Proceedings of the First IEEE Workshop on Disaster Recovery Networks (DIREN '02), vol. 24, pp. 1–7, 2002.

[15] J Satishkumar, Mukesh A.Zaveri, Graph-based Resource Allocation for Disaster Management in IoT Environment, ACM, Proc. Advanced Wireless Information, Data, and Communication Technologies, 2017

[16] Sujoy Sahal,Nitin Agarwal, Priyam Dhanuka, Subrata Nandi, Google map based user interface for network resource planning in post disaster management ACM proc. computings of development, 2013

[17] Kaur M, Mohana R.Static load balancing technique for geographically partitioned public cloud. Scalable Computing: Practice and Experience. 2019 May 2;20(2):299-316.

[18] A. Musaddiq, Y. Bin Zikria, O. Hahm, H. Yu, A. K. Bashir, and S. W. Kim, A Survey on Resource Management in IoT Operating Systems, IEEE Access. 2018, doi: 10.1109/ACCESS.2018.2808324.

[19] E.W. Dijkstra, Cooperating Sequential Processes, Programming Languages, F. Genuys, ed., pp. 103-110, New York: Academic Press,1968.

[20] A.N. Habermann, Prevention of System Deadlocks, Comm. ACM, vol. 12, no. 7, pp. 373-377, 385, July 1969.

[21] R.C. Holt, Some Deadlock Properties of Computer Systems , ACM Computing Surveys, vol. 4, no. 3, pp. 179-196, Sept. 1972.

[22] Coffman, Edward G., Jr.; Elphick, Michael J.; Shoshani, Arie, System Deadlocks . ACM Computing Surveys 3 (2): 67–78. 1971

[23] A. Silberschatz, P. B. Galvin and G. Gagne, Operating System Principle , Seventh Edition, Wiley India.

[24] Agrawal, Smriti, Madhavi Devi Botlagunta, and Chennupalli Srinivasulu. A total need based resource reservation technique for effective resource management. International Journal of Computer Applications 68, no. 18 (2013).

[25] Devi, Botlagunta Madhavi, Smriti Agrawal, and Chennupalli Srinivasulu. An Efficient Resource Allocation Technique for UniProcessor System. International Journal of Advances in Engineering Technology 6, no. 1 (2013): 353.

[26] Botlagunta, Madhavi Devi, Smriti Agrawal, and R. Rajeshwara Rao. Effective resource management technique using reservation pool. In International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), pp. 1-7. IEEE, 2014.

[27] Botlagunta, Madhavi Devi, Smriti Agrawal, and R. Rajeshwara Rao. Dynamic Budget-Thershold based Resource Reservation Technique. Compusoft 8, no. 7 (2019): 3242-3249.

[28] Sheau-Dong Lang, An Extended Banker's Algorithm for Deadlock Avoidance ,IEEE Transactions On Software Engineering,

vol. 25, no. 3, May/June 1999

[29] Odun-Ayo, Isaac, Sanjay Misra, Nicholas A. Omoregbe, Emmanuel Onibere, Yusuf Bulama, and Robertas Damasevicius. Cloud-Based Security Driven Human Resource Management System. In ICADIWT, pp. 96-106. 2017.

[30] Cheng Xin, Xiaozong Yang,A concurrent distributed deadlock detection/resolution algorithm for distributed systems, Proceedings of the 5th WSEAS/IASME International Conference on Systems Theory and Scientific Computation, 2005, pp. 336-341.

[31] Jaehwan Lee, Vincent John Mooney,A novel deadlock avoidance algorithm and its hardware implementation, Proceedings of the 2nd IEEE/ACM/IFIP international conference on Hardware/software code sign and system synthesis, 2004, pp. 200-205.

[32] A. Chowdhury and S. A. Raut, A survey study on Internet of Things resource management, Journal of Network and Computer Applications. 2018, doi: 10.1016/j.jnca.2018.07.007.

[33] Kumar J.S, Zaveri M.A, Choksi M, Activity Based Resource Allocation in IoT for Disaster Management, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 220. Springer, 2018

[34] H.Wu,WN.Chin,J.Jaffar, An efficient distributed deadlock avoidance for the AND model ,IEEE Trans. Software Engg 28 (2002)18–29.

[35] Y.B.Ling, S.G.Chen, C.Y.Chiang, On optimal deadlock detection scheduling , IEEE Trans on Computers55(9)(2006)1178–1187.

[36] Shubham Kumar and SaravananChandran, Modified Execution Time based Resource Reservation (METRR) Algorithm), 3rd International Conference on Business and Information Management (ICBIM) 2016.

[37] Yin, W., Stephane, L., Terence, K., Manjunath, K., Scott, M.,The Theory of Deadlock Avoidance via Discrete Control , 36th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages, PoPL,202- 214(2009).

[38] Hongwei Wang; JinfengTian; Mingqi Li; Weixin Mu; XiangchuanGao,Banker's algorithm based resource allocation in next generation broadcasting wireless systems . IEE Proc Int'l Conf Communications and Networking in China (2015)

[39] Xiang Xiao, Jaehwan John Lee. A parallel multi-unit resource deadlock detection algorithm with O(log2(min(m, n))) overall run-time complexity. J. Parallel Distrib. Comput. 71. pp. 938–954. (2011)

[40] P. H. Shiu, Yudong Tan and V. J. Mooney, A novel parallel deadlock detection algorithm and architecture, Ninth International Symposium on Hardware/Software Codesign. CODES 2001 (IEEE Cat. No.01TH8571), Copenhagen, Denmark, 2001, pp. 73-78.

[41] E. E. Ugwuanyi, S. Ghosh, M. Iqbal and T. Dagiuklas, Reliable Resource Provisioning Using Bankers' Deadlock Avoidance Algorithm in MEC for Industrial IoT, in IEEE Access, vol. 6, pp. 43327-43335, 2018, doi: 10.1109/ACCESS.2018.2857726.

[42] H. K. Pyla and S. Varadarajan. Avoiding deadlock avoidance. PACT '10, pages 75–86, New York, NY, USA, 2010. ACM.

[43] Youming Li, A Modified Banker's Algorithm, in Springer Innovations and Advances in Computer, Information, Systems Sciences, and Engineering, pp 277-2819 (2012).

[44] Pankaj Kawadkar, Shiv Prasad, Amiya DharDwivedi, Deadlock Avoidance based on Banker's Algorithm for Waiting State Processes in International Journal of Innovative Science and Modern Engineering (IJISME) ISSN: 2319-6386, Volume-2 Issue-12, November 2014

[45] Dixit, Kshipra and Ajay Khuteta. A Dynamic and Improved Implementation of Banker's Algorithm. International Journal on Recent and Innovation Trends in Computing and Communication 5, no. 8 (2017): 45-49.

Edited by: Anand Nayyar
Received: Apr 4, 2020
Accepted: Jul 1, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 407–424, DOI 10.12694:/scpe.v21i3.1737

ACCESS MANAGEMENT OF USER AND CYBER-PHYSICAL DEVICE IN DBAAS ACCORDING TO INDIAN IT LAWS USING BLOCKCHAIN

GAURAV DEEP [*] JAGPREET SIDHU [†] AND RAJNI MOHANA [‡]

Abstract. Computing on the cloud has changed the working of mankind in every manner, from storing to fetching every information on the cloud. To protect data on the cloud various access procedures and policies are used such as authentication and authorization. Authentication means the intended user is access data on the cloud and authorization means the user is accessing only that data for which he is allowed. The intended user now also includes Cyber-Physical Devices. Cyber-Physical Devices share data between them, fetch data from cloud. Cloud data is managed by employees of cloud Companies. Persons sitting on the cloud managing companies data is always doubtful as so many insider attacks have happened in the past affecting the company Image in the market. Data Related to Cyber-Physical Space may come under Insider attack. Companies managing user data are also liable to protect user data from any type of attack under various sections of the Indian IT act. Work in this paper has proposed blockchain as a possible solution to track the activities of employees managing cloud. Employee authentication and authorization are managed through the blockchain server. User authentication related data is stored in blockchain. Authorization rules are written in any Role/Attribute-based access language. These authorization rules stores the data related to user requests allowed access to data in blockchain. Proposed work will help cloud companies to have better control over their employee's activities, thus help in preventing insider attack on User and Cyber-Physical Devices.

Key words: Cryptography, Transmission Control Protocol, Single sign-on, Internet of Thing, Policy enforcement Point, Timestamp, Nonce, Policy Decision Point, Internal Threat Detection Unit.

AMS subject classifications. 94A60

1. Introduction.

Cloud Computing Revolutionized Data storage to Processing in every area of science and technology. International Data Corporation (IDC)[21] released a study on the Global Data sphere, which says it will grow to 175 Zettabytes by 2025. Data is shared and accessed every moment throughout the world. In data storage and sharing Cloud plays an important role. Every cloud model through the world Follows Standards laid down by the National Institute of Standards and Technology (NIST). Cloud data is stored as well as shared on-demand with the help of configurable computing resources [41]. Digital Data on the Cloud is stored, which may represent any type of form of Information such as Images, Sound, Video, Database, etc.

Cloud Database means Database stored over the cloud, which offers various services to the users such as storing, modifying and making it available anywhere in the world. To maintain the Privacy of data, It Is Important to Protect Cloud Databases [53]. According to the CIA Principle, security concerns mainly deal with Confidentiality, Integrity, and Availability. Security Concerns covers Attacks from within and from outside the Organization, Issues related to Consistency Management, Access Control, Network Breaches, Resource Exhaustion, etc. are also covered [7].

Insider threat means threats originating from Employees of the organizations, These Employees have been provided Access rights to access the internal system, thus violating the Internal System organization security policy. Outsider threats try to release the confidential information out in the real world, which defaces the organization. When the Care Taker of Various Services of Cloud Computing tries to Steal Users Data it becomes more difficult to prevent it [38]. Cloud Database is no different from such a scenario where its numerous advantages supersede its disadvantages/Loopholes, these Loopholes cannot be ignored when it has affected in the past so much to the working of many Organizations whether it maybe Yahoo, Facebook, and Google.

[*] Research Scholar, Department of CSE & IT, JUIT, Solan, India
[†] Assistant Professor(Senior Grade), Department of CSE & IT, JUIT,Solan, India
[‡] Associate Professor,Department of CSE & IT, JUIT, Solan, India (rajni.mohana@juit.ac.in).

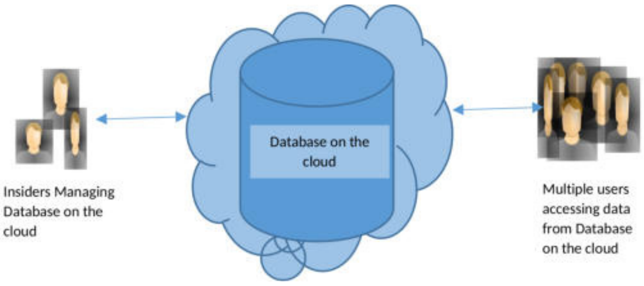

Insiders Managing Database on the cloud
Database on the cloud
Multiple users accessing data from Database on the cloud

Fig. 1.1. *Role of Insider and users on the Database on the cloud*

Intensions of these Staff members to monitor User and its data activity is always doubtful due to many insider attacks happened in the past. A Survey in this regard From U.S. State of Cybercrime in the year of 2016 [8] represented, Electronic crimes suspected to be caused by insiders is of 27%. According to this survey, 33 Percent of the Respondents Agreed that Insider attack is more Dangerous than Outsider Attack.

The number of Insiders depends on the amount of data they are manging over the cloud as shown in Figure 1.1. Insiders are also Employees of the Organization. They can take advantage of the information of how and where data is protected to do insider attacks.

Existing User Authentication Techniques Suffers from Various Attacks and Threats. User Password Can be guessed and Two Factor Authentication Using Short Message Service (SMS) also suffers from Attacks, as Code sent on SMS can be tracked by the Attacker on the network [49].Authentication codes from Google were sent by Google Authenticator to respective users via SMS as these Codes are difficult to Cracked but Security Breach to Google could lose all User Authentication Codes.

Organization of this paper. The rest of the paper is as follows, Authentication techniques available for outsiders and insiders to the cloud are covered in Section 2. It also covers various authorization policies available for users. Section 3 covers various sections available under the Indian IT Act. This section also covers details of sections dealing with various aspects of the Cloud. Section 4 covers Detail and working of Blockchain. This section also covers details of application areas of Blockchain. Section 5 covers Need of Transaction Authentication Mechanism for Access Management in Database as a Service (DBaaS) according to Indian IT Laws. Section 6 Covers Transaction Authentication Mechanism using Blockchain to store Every Transaction Detail according to Indian IT Laws. Experiment Results of Proposed Work were shown in Section 7. Finally in Section 8 Paper is concluded.

2. Authentication and Authorization of User.

Importance of Access Control Can be understood from the fact that many researchers tried to explore this field so that only legitimate users can access his data. User access control consists of two main components, Authentication and Authorization as shown in Figure 2.1.

This section discusses primarily Authentication and Authorization of Insiders and Outsiders on the cloud.

2.1. Authentication of the user.

2.1.1. Authentication techniques of the outsider users. The purpose of Authentication is to allow access only to the intended user. Work from Many researchers on Outsider threat is compared and shown in

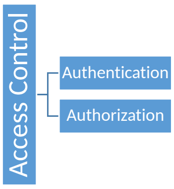


Fig. 2.1. *Access Control Components*

Table 2.1
Contrasting Authentication Techniques for Outsider User

	Authentication Type	Single sign-on	Technique used	Suitable for Resource constraint IOT	Mutual Authentication	Multi-Owner Authentication
Tsai et al.[49]	Three factor	Yes	Elliptic curve cryptography	No	Yes	No
Kalra et al.[31]	Two Factor	No	Elliptic curve cryptography	Yes	Yes	No
Amin et al. [3]	Multi-Factor	No	Bio Hashing	No	Yes	No
Yang et al. [55]	Two Factor	Yes	Delffie-Hellman	No	No	No
Kumari et al. [33]	Multi-Factor	No	Elliptic curve cryptography	Yes	Yes	No
Shajina and Varalakshmi [45]	Two Factor	Yes	Triple DES	No	Yes	Yes
Anakath et al.[4]	Multi-Factor	No	Simple-Homomorphic Encryption	No	No	No
Chaudhary et al. [15]	Three factor	Yes	Elliptic curve cryptography	No	Yes	No
Kumar et al.[32]	Biometric	No	Elliptic curve cryptography	No	Yes	No
Chatterjee et al.[13]	Biometric	No	Clustering	No	Yes	No

Table 2.1. Authentication Process for a user means that allows only legitimate users to access their data and restricts various hackers to attacks.

Tsai et al. [49] had proposed Authentication Scheme for mobile users. This scheme allows to use single private key for accessing cloud services from multiple service providers. Public and Private keys for the user as well as for the service provider are generated by Smart card generator so that they can authenticate each other . Smart card allows users to access services from the service providers. This scheme benefits in providing user anonymity, user un-traceability, mutual authentication, and key exchange.

Kalra et al.[31] had proposed mutual authentication protocol For Internet of Thing (IoT) devices and cloud servers This scheme uses Hypertext Transfer Protocol (HTTP) cookies to implement Secure Elliptic curve cryptography (ECC). Internet of thing Devices uses Cloud Services to enhance their Processing Capabilities. To get Authenticate with the server, embedded devices should work as HTTP Client. Transmission Control Protocol/Internet Protocol (TCP/IP) Protocol Stack is used to configure embedded devices. Three phases are available in this Protocol from Registering Devices in First Phase,Getting Log-in them in second Phase and Authenticate them in the last Phase. This Scheme Provides Resistance to various types of attacks such as brute force attack, eavesdropping, man-in-the-middle attack, offline dictionary attack, cookie theft attack, replay attack and provides forward secrecy, anonymity, confidentiality, and mutual authentication.

Amin et al. [3] had proposed User Authentication for a multi medical server system using User-id, Password, Biometric template like fingerprint and smart card. First, the user chooses his desired identity, Password and Biometric template like fingerprint and sends them through a Secure channel for user Registration. After receiving User details and Applying Bio Hashing on them, data is stored in the User Smart card as well as in User Registration Medical Server. Once the user is authenticated he is allowed to fetch data from the desired Medical Server according to his requirement. This scheme Prevents Session key discloser attack, User Impersonation Attack, replay attack, enables early wrong password detection and provides Mutual authentication, Resists off-line password guessing attack.

Yang et al.[55] had proposed a Protocol for allowing access to multimedia data on the multimedia cloud. This Protocol uses Two-factor authentication with Open ID which requires smart cards along with user Login details. Various Cloud models were used in this scheme for achieving Authentication of Smart cards and Users. User Authorization policies were based on Role-Based Access Control (RBAC). To validate this work various analysis was done like on security, Functionality, and Efficiency.

Kumari et al. [33] had proposed Authentication scheme for IoT and Cloud Servers. This scheme is uses Multi-Factor like login details, Cookies and device details along with Temper resistant device, Elliptic curve cryptography which helps in preventing various types of attacks like the absence of device anonymity, Insider attack, Offline Password guessing, and No session key computation. This Authentication Protocol is suitable for Resource constraint IOTs where mutual authentication is required.

Shajina and varalakshmi [45] had proposed protocol for Multi Owner Authentication, this protocol works on multiple owners, Group manager and service manager in a cloud for authentication and increases the Security requirement of Single sign-on. The Proposed mechanism allows the main owner to create a group and is allowed to add other members along with their access permissions. Owners are provided with a valid token after certified by a Certification authority with all required parameters .A Valid token Consist of all the details of User credentials, Token expiration time, services granted, etc., session tokens from session manager will allow these services to access.

Anakath et al. [4] proposed Trust Model for authentication, in this device identity is identified and authentication Protocol is selected. For authentication purposes, Knowledge, Possession and Inherence factors can be used. This protocol uses Possession factors, One Time Passwords, and passwords that users know only. A user profile is created on Big Data Multi-Factor Cloud Authentication (MACA) with user details and user Parameters in encrypted form.

Chaudhary et al. [15] had proposed an Improved User Authentication Scheme. This scheme uses a single private key for authenticating mobile users, allowing them to use services from multiple cloud service providers. This work is an improvement work of Tsai[49]. This paper prevents the Server Forgery attack at the time of the Authentication phase. Proposed work is much more secure, robust and is validated in ProVerif.

Kumar et al. [32] had proposed authenticating cloud users using Face features i.e by Bio-metrics based recognition. Encryption is used on Bio-metric Database which stores Facial features of cloud users. Facial Features of users are extracted from pre-processed face images. These Facial Features helps in recognition of users. Scores of Facial features are calculated and it is matched with the stored similarity scores of facial features on the cloud.

Chatterjee et al. [13] proposed a re-authentication system based on Bio-metrics, This system is better and enhances security level by using keystroke dynamics over Password based authentication mechanism. In this Scheme User is asked to enter Log-in details for authentication Purposes. When User enter's his credential for authentication , his Keystroke Dynamics are stored in Database. These details will help in identification and verification,which are extracted by a k-means clustering algorithm. Tests were conducted on Heterogeneous, Homogeneous, and Aggregate feature sets.

2.1.2. Authentication techniques of insider user. To safeguard user data on the cloud and accessed by only legitimate user various user authentication techniques have been discussed, there is a need to protect user data from insiders on the cloud also. Behavioral analysis is the area where many researchers have worked for designing authorization policies for insiders. Table 2.2 shows a comparison of many approaches that have worked on Insider Threat.

In the paper of Wu et al. [53] have proposed making the understanding of User Data difficult for insiders by applying Encryption. Data is Decrypted first before applying query on the user data, in the end encryption is done again. Feature index extraction is applied to user data before encryption proposed by authors. It also helps in making a query on the cloud. Encryption was done with Index Generator, Query translator helps in making Feature Index of user data and Query Executor executes the query.

Moon et al. [38] introduced two-tier architecture for analyzing the behavior of Insider . They have also proposed In-Memory Database (IMDB) for a database protection system. Work done by Insiders are saved in Log Files known as Change Audit Log. Database log Pre-processor pre-processes the log File.This File is further sent to Insider Behavior Analysis Server. This insider behavior analysis Server analysis and detects any availability of Attack. Cloud Capability is also incorporated in this.

Yaseen and Panda [57, 58, 56] have contributed three papers on the detection and prevention of Insider

Table 2.2
Contrasting Authentication Techniques against Insider Attack

	Insider Action Monitoring	Authorization rules Modification based on Insider Action Monitoring	User-Machine probity	Authentication of Insider	Availability of encryption on User Data before querying on Cloud
Wu et al. [53]	No	No	No	No	Yes
Moon et al. [38]	Yes	Yes	No	No	No
Yaseen et al. [57][58][56]	Yes	Yes	No	No	No
Dou et al. [23]	Yes	Yes	Yes	No	No
Shaghaghi et al.[44]	Yes	Yes	No	No	No
Chattopadhyay et al. [14]	Yes	No	No	No	No
Baracaldo et al. [6]	Yes	Yes	No	No	No
Meng et al. [36]	Yes	No	No	No	No
Babu et al.[5]	Yes	Yes	No	Yes	No
Eberz et al. [24]	Yes	No	No	Yes	No

attack. In their First work [57] Insider threat prediction and its prevention measures have been proposed, Insider knowledge is analyzed by using a knowledgebase Algorithm that also considers Constraint Dependency, Hot Cluster, Safe Cluster, and Dependency Matrix. By using this Algorithm Knowledge Graph is generated which helps in protecting Insider Attack. Threat Prediction Graph was proposed in the second work [58] by using the knowledgebase Algorithm. In third work, authors have proposed Architecture with Multiple Policy Enforcement Points (PEP's) and Single Policy Decision Point (PDP) to detect insider threats.In this architecture alogrithims proposed in previous works were used. This system is suitable to work when the number of PEPs is less in number.

Dou et al. [23] have proposed an authentication protocol for Hadoop with a Trusted platform. This protocol helps in removing the limitations of user authentications and insider attacks in Kerberos. Authentication keys and its operations were locally hidden in this Protocol. This Protocol is bounded with specific Systems. It stores current software and hardware details of the hosting machine in an internal set of platform configurations registers. This Proposed protocol helps in securing specific systems against insider attacks.

Shaghaghi et al. [44] have proposed Gargoyle Software Defined Network (SDN) architecture. The proposed work was designed to detect and deter suspicious activities of insider using SDN. It also analyzes Passive Network traffic and retrieves contextual information. Mainly three components were proposed in this architecture Network Context Analyser, Risk Management, and Advanced Enforcement Point. Various Risks were detected and actions can be planned accordingly based on insider activity details. This detail is extracted by monitoring network traffic.

Chattopadhyay et al. [14] have proposed Time-series classification of insider activities. In this work Insider behavior analysis was done on tracking single day activities. This analysis was done from a single day to over some time for detecting insider threats. Statistics were collected to detect malicious or non-malicious insider based on behavioral analysis. Classification technique two-layered deep autoencoder neural network is applied to improve the results.

Baracaldo et al. [6] proposed the Geo-Social Insider Threat Resilient Access Control Framework (G-SIR). In this proposed work Insider movement activities are monitored. This monitoring helps in classifying insiders into enablers, inhibitors or neutral. Risky users come in the category of Inhibitors. Trusted users come in the category of enablers and average users based on Risk level comes in the category of neutral. This framework uses PEP-PDP Model along with Monitoring, Context, Inference and Access control Module. Role-based access control (RBAC) is used to write Permissions and Roles.

Meng et al. [36] have proposed a technique to prevent Medical Smartphone Network from an Insider attack, where it can leak Patient information malicious devices are detected based on behavioral profiling. Nodes in the

Table 2.3
Comparison of Different Authorization Policies framework used in Distributed Environment.

	Authorization Policy framework used	Environment used	Application
Abomhara et al. [1]	WBAC (Team-based)	Distributed	Healthcare
Alam et al. [2]	GRBAC	Distributed	Cloud
Habiba et al. [27]	iCanCloud simulation platform	Distributed	Cloud
Sun-Moon Jo [30]	XML	Distributed	Mobile
Chen et al. [16]	RBAC	Distributed	Healthcare
Shin et al. [46]	Bilinear pairings, Strong Diffie–Hellman representation, Linear encryption	Distributed	Cloud
Gabillon et al. [26]	ABAC	Distributed	Pub-Sub Network For IoT
Rathore et al. [42]	Answer Set Programming	Distributed	Online Social Network

MSN are connected to the Central server, each node in the network sends its Statistics based on the user working on the Medical smartphone. Working Profile of each node is created on the central server. A malicious node on the network is detected by the difference in Euclidean distance between two behavioral profiles. Evaluations were carried out in Real-world MSN with the help of a Practical healthcare center.

Babu et al. [5] have proposed a technique to prevent Insider attack on the cloud by Analysing the Behaviour of the Insider and associating Risk-based access control. Behavior analyses are done by using Keystroke Dynamics. Risk analyses are done in an offline manner with the associated resource. Every object is assigned a value of risk. These Risk Values are stored in a database called resource repository. For Behavioral analyses, Support vector machine is used. If a Malicious user is detected in the system it's all Privileges are revoked.

Eberz et al. [24] have proposed a technique to prevent Insider Threat in an Organization by detecting Eye Movement biometrics. In this work, researchers have identified a set of 20 features of an eye by which user authentication can be done in a transparent continuous manner. Video-based gaze tracking is used to track eye movement. Experiments were conducted in a controlled manner in a Lab on 30 Persons from the general public. Persons were asked to perform various activities on the screen to study eye movement and various other parameters. Open set and closed set classifiers are applied to the retrieved data set. Experiments were repeated after some time to test the time stability of the proposed features.

2.2. Authorization of users. To allow a user to access data is decided under Authorization policy. Along with Authentication, Authorization policy plays a very important role. Various Authorization Policies framework used in Distributed Environment is discussed in Table 2.3 along with their application area.

Abomhara et al. [1] have proposed a work-based access control (WBAC) model with a team role classification based on the Belbin team role theory. Teams in WBAC models are segregated based on Thought, action, and management based on their contributions to collaborative work. Proposed work has been suggested for cooperative healthcare environments. In a Particular scenario Multiple Doctors from multiple Departments and hospitals working as a team on a patient.

Sharing and Access to the healthcare record of a patient with Multidisciplinary team consultants need Authorisation control, as leakage of sensitive data may happen by insiders. Authors have first formalized the model with basic elements and relations, defined various authorization constraints and access control decision functions for WBAC Model. This work reduced the complexity level of Permission reviews as compared to Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC).

Alam et al. [2] have proposed Garbled role-based access control (GRBAC) in the cloud. The concept of Garbled Circuits (GC) and Fully Homomorphic Encryption (FHE) are used in Garbled computation. This Garbled computation is used in Role-based access control (RBAC) making it into GRBAC. Permissions to user access are associated with their respective assigned roles. All the details of user roles are stored in the RBAC server. An attacker is not able to know the roles even if the GRBAC is compromised. It helps in Providing Strict Security in Cloud Environment.

Habiba et al. [27] have proposed the Dynamic Access control system in the cloud. This Dynamic Access control system consists of mainly four models namely the Data access right model, Policy model, Access control management model and Authorization model. The data access right model consists of access rights tree, in which Access rights are represented by access rights and the hierarchical relationship between two access rights is represented by edges.

Policy model classifies Policies into different categories obligations (O), conditions (C), primary rules (R), deadline (D) and user preferences (F). Policies must represent there Subject (S), a requested access right (A), a resource (Rs), a set of rules (R) and a preference (F). Every policy must be expressed with 8 –Tuple (S, A, Rs, R, C, O, D, F). Access control management model consists of Many Sub Models handling many key areas like Request Management, Communication, User Management, Data Management Module, Monitor, etc. Authorization model consists of three stages where decision making is performed as Pre, On-Going, and Post-stage of data access along with credit level checking of the user.

Sun-Moon Jo [30] have proposed a secure access policy method for Dynamic Extensible Markup Language (XML) data environment. The author is working on Resource Efficient Secure Access Policy in which access to data is allowed according to Privilege Information, Security Policy, Authorization Policy, and Propagation Policy. In this paper it focuses only on the accessible area as element units in the target document, on these element units very small access policies are applied. This Policy accesses parts of the target document allowing every small access policy to work on the whole target document.

Chen et al. [16] have proposed community medical Internet of things (CMIoT) for medical data. Privacy protection of medical data in healthcare is provided in terms of transmission protection, storage protection, and access control. Transmission protection is provided by asymmetric encryption, storage protection is provided by symmetric encryption and access control is provided by identity authentication and Dynamic authorization based on the role under which access is applied. In community medical Internet of things (CMIoT) data from various IoTs are collected at the gateways further Multi-Path fragmented, encrypted and sent to Cloud for storage. This cloud data is allowed to access by the user according to his predefined Role. Dynamic Authorization allows data to be fetched only from Third Party Cloud of community medical Internet of things.

Shin et al. [46] have proposed the Anonymous Authentication and Authorization (AAnA) scheme. There proposed work uses short traceable signatures. In this scheme, two authorities are simultaneously working one is a group manager and another is the authorization manager. This scheme achieves anonymity by having two different managers for Group membership and Authorization. The role of the Group Manager is to provide Group membership based on Short traceable signatures. Authorization manager provides Privileges to users based on their real identity, Authorization list of all users along with their Privileges is forwarded to Service provider. At any time the Service provider detects Illegal activity it will ask the signature from the user, which is passed to the Group manager and Authorization manager for further necessary action.

Gabillon et al. [26] have proposed a highly expressive attribute-based access control (ABAC) security model. This Model is used for the Message Queuing Telemetry Transport (MQTT) protocol. MQTT Protocol is used for Publisher-subscriber Network for the Internet of Things. Whenever publisher Publishes Messages in various topics, subscribers get messages under the topics for which they are subscribed. This Paper assumes only one Trusted Broker is working in MQTT Protocol. This paper also assumes to be working on TLS/SSL at the Transport layer between all nodes of the IoT network. Authors have used first-order logic with equality to define the proposed model. The access control enforcement system in the proposed model uses Policy Enforcement Point (PEP), a Policy Decision Point, a Policy Information Point (PIP—contextual database), and a Policy Administration Point (PAP). Logical Security policies for this model are defined in the Resource Description Framework (RDF).PEP intercepts all the MQTT requests and forwards them to PDP, PDP Takes help from PIP in deciding the access and saves the results in PIP.

Rathore et al. [42] have proposed an access control model for online social networks. This control model works on resources shared by Single or Multiple Parties on Online Social network. Privacy suffers on Social Network as Resources is shared by multiple times, some times without the consent of Owners. In this model Trust level is calculated Among Each Owner of the Resource. Access Policy among owners Depends upon trust level, as trust is higher among Family members and lower in Normal friends. The proposed model is logically represented by using Answer Set Programming.

3. Indian IT Laws for Privacy Threat. To prevent data threat on the cloud, many protocols were designed and implemented it can be seen From Tables 2.1-2.3. To prevent data from theft IT Laws plays a

Table 3.1
Indian IT Acts on Electronic Data and its Management

Sections based on concerned issues	Section According to Indian IT Act 2000	Amended Section According to Indian IT Act 2008	Rule According to Indian IT Act 2011
Retention of electronic records.	7	7 7A	
Secure electronic record	14	14	
Certifying Authority to follow certain procedures	30	30C 30CB	
Penalty for damage to the computer, computer system, etc.	43	43 43A	
Tampering with computer source documents.	65	65	
Hacking with the computer system.	66	66 66C 66F	
Protected system	70	70A 70B	
Confidentiality and Privacy breach	72	72A	
Responsibility of service provider and authorized agents			7
Account Audit and Information System details of the service provider and authorized agents			8

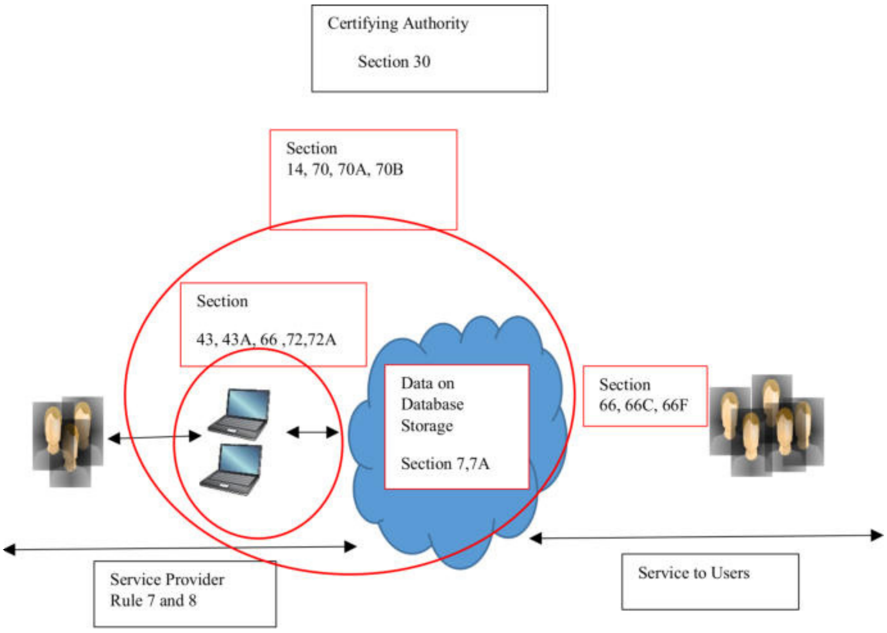


Fig. 3.1. *Role of various IT Act sections Data and IT Resources*

major role, but their role is limited when Laws differ from country to country. Indian government brought Indian IT ACT 2000 [9] in this regard in the year 2000, later amendments were done in 2008 [10] on Electronic Data and its Management is shown in Table 3.1.

Various points are provided related to the Retention of Electronic Records, Secure Electronic Records, and Certifying Authority Followed Procedure, etc. They are shown in Figure 3.1 and explained below in detail:

Section 7, Retention of electronic records in its original generated format is only allowed when it is required to keep electronic records for a certain period for subsequent reference with all document origin, destination date and time of dispatch or receipt details. This point was amended in IT act 2008 as Section 7A Audit of Documents etc. in Electronic form there is a provision for audit of documents, records or processed information /unprocessed information.

Section 14, Secure electronic record by this it means that any security procedure applied on the electronic record to keep it secure during at a specific point of time then that Electronic record is said to be Secure electronic record from such point of time to the time of verification.

Section 30, Certifying Authority to follow procedures from making use of IT Infrastructure to get secure from intrusion and misuse, Providing Reliability in various services and functions, To Follow all security procedures to ensure that the secrecy and privacy of the digital signatures are assured and to observe such other standards as may be specified by regulations. Some new points are added in this in IT Act 2008, a repository of all Electronic Signature Certificates issued under this Act are to be maintained and publish information regarding its practices, Electronic Signature Certificates and current status of such certificates.

Section 43, Penalty for damage to the computer, computer system, by any unauthorized person tries to access, retrieve any form information, corrupts the system with a virus, any sort of damage to the system he shall be liable to pay damages by way of compensation not exceeding one crore rupees to the person so affected. Manipulation or theft to data in any form is also considered as damage to the data in IT ACT 2008.

The new subsection to this is also added as “43 A Compensation for failure to protect data”. When a corporate body is unable to manage sensitive data which it owns making wrongful gains to someone then that Corporate is liable to pay damages by way of compensation, not exceeding five crore rupees, to the person so affected.

Section 65. Tampering, destroying, concealing or any form of damage to computer source documents, source code is done knowingly or intentionally, when it is required to be maintained by law for the time being in force, shall be punishable with imprisonment up to three years, or with fine which may extend up to two lakh rupees, or with both.

Section 66. Hacking with the computer system, Any Person who hacks the system Manipulates or deletes any information residing in a computer resource shall be punished with imprisonment up to three years, or with fine which may extend up to two lakh rupees, or with both. New sub Sections to this is also added in IT Act 2008 as:

Section 66C Punishment for identity theft. Under this point who so ever does any type of Identity theft, shall be punished with imprisonment of either description for a term which may extend to three years and shall also be liable to fine which may extend to rupees one lakh.

Section 66F Punishment for cyber terrorism. Under this point who so ever with intent to threaten the unity, integrity, security or sovereignty of India or to strike terror in the people or any section of the people by unauthorised access and any sort of damage to any computer resource or Computer database or Cybercrime using such conduct causes Losses to Human Beings or Property or adversely affect the critical information infrastructure in specified under section 70 shall be punishable with imprisonment which may extend to imprisonment for life.

Section 70, Protected system. The government may notify any Computer system/Computer Network as a Protected system accessed by authorized personnel only. Unauthorized access to the protected system shall be punished with imprisonment of either description for a term which may extend to ten years and shall also be liable to fine.

In IT Act 2008 this point Protected system is referred to as Critical Information Infrastructure means the computer resource, the incapacitation or destruction of which, shall have a debilitating impact on national security, economy, public health or safety. The government should follow various security procedures to secure Critical Information Infrastructure. New subsections to this is also added in IT Act 2008 as:

Section 70 A, National nodal agency. The government may designate any organization as National Nodal agency which shall be responsible for all measures including Research and Development relating to the protection of Critical Information Infrastructure.

Section 70 B, Indian Computer Emergency Response Team to serve as a national agency for incident response.

The government may designate any government agency as the Indian Computer Emergency Response Team with all the required staff. This team will work on various elements of Cyber incidents and Cybersecurity. Any service provider, intermediaries, data centers, body corporate or person who fails to provide the information called for shall be punishable with imprisonment for a term which may extend to one year or with fine which may extend to one lakh rupees or with both.

Section 72, Penalty for breach of confidentiality and privacy. Any person who has secured access to any electronic record, book, register, correspondence, information, document or other material without the consent of the person concerned and discloses them to any other person shall be punished with imprisonment for a term which may extend to two years, or with fine which may extend to one lakh rupees, or with both. In IT Act 2008 Definition of Any Person also consists Intermediary in Section 72 A. Imprisonment for a term which may be extended to three years, or with a fine which may extend to five lakh rupees, or with both.

In 2011 New IT rules were issued [11]:

Rule 7. Responsibility of service provider and authorized agents: The government may direct every service provider and authorized agent to keep every detail of electronic services delivered and the said records shall be produced for inspection and audit.

Rule 8. Account Audit and Information System details of the service provider and authorized agents:- Government may direct every service provider and authorized agents to go for Audit on regular intervals .To check various parameters of security, confidentiality and the privacy of information. The performance of any software application used in the electronic service delivery is also checked .The accuracy of accounts kept by the service providers and authorized agents is maintained . These service providers and authorized agents must rectify the defects and deficiencies pointed out by the audit agencies within the time limit specified by the audit agency.

Due to the declaration for protecting the data of every individual transaction and citizen is to be submitted by every service providers and the authorized agents. Protected data should not be disclosed to any one in unauthorized way without the written consent of either the individual or the appropriate Government. Otherwise, the Government is allowed to take necessary action under section 45, and can debar such service providers and the authorized agents.

4. Blockchain. Blockchain provides numerous benefits in terms of Security. Every technological area wants to take advantage of it. Blockchain provides the benefit of Immutability, Forgery Resistant, Democratic, Double-Spend Resistant, Consistent State of the Ledger, Resilient and Auditable [39, 34]. Immutability means once a transaction is done on the Blockchain it cannot be altered. Every Node in the Blockchain Cryptographic hash and digital signatures are used to make it Forgery Resistant. Every node in the Blockchain should have equal rights like in Democratic structure, no one is powerful than others. Preventing double spend in Blockchain is done by allowing to access every transaction up to the genesis block. All nodes in the Blockchain are Auditable, all previous nodes are accessible via a hash function.

Blockchains are of three types: Public, Private and Consortium Blockchain. In Public Blockchain anybody can Participate, whereas in Private Blockchain authorized user is allowed to participate in a controlled manner by a centralized authority. As in Private Blockchain Number of users are in Finite Numbers it is less complex as compared to Public Blockchain. In the Consortium Blockchain, the consensus process is controlled by a pre-selected set of nodes, these selected nodes control the authorization of nodes.

Each node in Blockchain is connected to its previous node, backward to the first node (Genesis Node) in the distributed network. Each node stores the Hash value of the previous node, by which it is checked membership of Blockchain. Various parameters are stored in each node of Blockchain like Index value, the Hash value of the previous node, Timestamp value, Merkle tree root hash, Data, Nonce value is shown in Figure 4.1.

It is getting difficult to change Blockchain Previous Node parameters as it grows. Blockchain has helped in numerous applications in achieving the Desired Security level. Table 4.1 shows various solutions proposed by Blockchain in various key areas.

Chen et al. [17] have proposed the use of Blockchain in the education sector. All academic details of a student are to be stored in the Blockchain including assignments, exam results and Degree details to prevent Fraud in education, which can be accessed by student ID. Blockchain can be used as learning as Earning, Digital

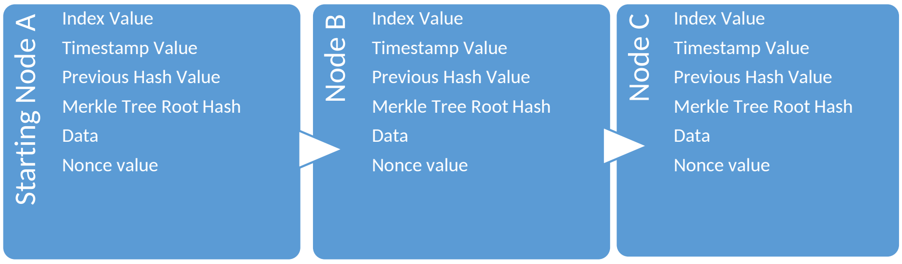


Fig. 4.1. *Blockchain Nodes with Stored Parameters*

Table 4.1
Research Issues with the use of Blockchain as the solution

	Deal with issues	Solution proposed
Chen et al. [17]	Fraud in Education	Student Education detail in Blockchain
Li et al. [35]	Bottleneck issue and chances of compromising centralized management server in VANET	Decentralized VANET with all data in Blockchain
Chung et al. [18]	Customized Product Process Management is difficult to maintain in Cognitive Manufacturing	Blockchain is used to maintain Data of Process Management in Cognitive Manufacturing
Sun et al. [48]	Trust issues in sharing based Smart cities	Blockchain can be used for trust-building in sharing based smart cities
B.Vinod [52]	Issue of interline charges, Bonus on genuine bookings to agents by airline and tracking of booking for a property becomes a tedious task when multiple sites are used for bookings	Blockchain can help in maintaining a single record for interline charges, the record of bookings and record of bookings for a property.
Han et al. [29]	Patient Health Data storage in a centralized system is prone to cyber-attacks.	Patient Health Data storage is done in Hybrid Blockchain, in Private Blockchain at the local Hospital and Consortium Blockchain at the Upper level. Hybrid Blockchain makes things difficult for an attacker.
Ryu et al. [43]	Digital forensics involves a very lengthy and difficult procedure in IoT for sent messages	IoT digital forensics is made simpler with the use of Blockchain

Currency can be rewarded for Smart contracts between students and teachers. The same can be applied to Teachers and schools where Teachers are rewarded with Digital currency based on their performance based on teaching activities.

Li et al. [35] have proposed the use of Blockchain in the area of Vehicle Adhoc network (VANET). Traditionally VANET works in a centralized system where it is controlled by the single management authority. The centralized system becomes a threat to its members once it is compromised by an attacker. Also, the Centralized system is prone to a single point of failure due to excessive load and bottleneck problems. To prevent centralized systems from the excessive load and bottleneck problem authors proposed to use a decentralized system with Blockchain. Vehicles are moving on the road in groups, their Parameters like speed, location, etc. are communicated to the Roadside unit by onboard unit installed in the vehicle. The roadside unit transfers these real-time vehicle parameters to the Certification Authority and other servers in the core network. All Data available in the core network is stored in Private Blockchain to make it more secure.

Chung et al. [18] have proposed the use of Blockchain in the area of Cognitive Manufacturing. Day by day competition in the market is increasing, Companies have started to attract customers by offering personalized customization on products. These customizations on products increase raw materials variety, so many changes in process management. All data generated from product customization to manufacturing to delivery is stored in a blockchain.it helps in understanding customer trends and demand. Sensors used in the manufacturing process are used for monitoring purposes. Data generated by these sensors are stored in Blockchain and can be accessed to detect any deviation in the required parameters.

Sun et al. [48] have proposed Blockchain in Sharing based smart cities. Blockchain can be used in basically Human, organization and technology in building trust in smart cities. To build trust in Sharing Transactions made by humans, Blockchain plays an important role. Data received from various IOT's in smart cities can be stored with the help of Blockchain builds trust in sharing based Services among businesses. Security provided by Blockchain builds trust in decentralized nodes, either may be used for transactions, IOT's or Services is sharing based smart cities.

B. Vinod [52] has proposed the use of Blockchain in Business related to Travelling. Loyalty bonuses for an airline can use digital tokens which can be accessed by using cryptocurrency. Interline charges are converted to cryptocurrency which can be taken by the next airline. Private Blockchain can be used for contracts between airlines and agents for tracking records of sales which helps in secure payments. The issue is raised when a property is booked by multiple sites. To eliminate this problem Blockchain can be used which helps in tracking the booking record of a property. Smart contracts can be generated using machine learning and stored in Blockchain.

Han et al. [29] have proposed the use of Blockchain storing medical records of patients in a hospital. In the Hospital chain, every Hospital stores Patient Health data in their centralized server, which is a soft target for cyber-attacks like WannaCry ransomware attack. To prevent Cyber-attacks patient data can be stored in de-centralized form using Hybrid Blockchain. Patient Health Data is stored in Private Blockchain at the Hospital level, if the patient allows it to share among other entities of the Hospital chain it is further stored in Consortium Blockchain. Two Blockchains are working one at the Hospital level and the other is at the Hospital chain level providing more security in the de-centralized Form.

Ryu et al. [43] have proposed to use Blockchain for IOT Digital Forensics. With the technological advancement with time, Exponential Growth of IoT's had happened. IoT can Communicate with each other as per requirement, in Cloud, on Network or directly. For Digital Forensics All three areas Cloud, Network and devices can be explored. Diversification of IoT's Type and usage has made Digital Forensics difficult. Authors have proposed Blockchain to store communication details of IOT's by which Digital Forensics is possible in a refined manner. Blockchain can be accessed by any one of the Participants Device user, Device Manufacturer, Service Provider and Investigator for Digital Forensics.

5. The need for User and Cyber-Physical Device Transaction Authentication Mechanism . User accesses its stored data from Cloud Databases, Performance of cloud database depends upon the architecture it is following. Policies to be implemented on Cloud Databases is decided by PDP, and are enforced by PEP. The best suitable architecture for Cloud databases is De-centralized (Distributed Network) based as shown in Figure 5.1.

Cloud Data storage to Manipulation related services is provided by Cloud service providers to various sections of users. According to the Indian IT act all the issues of Data and Resource management is to be managed by Service Provider. If any issue related to data privacy to data leakage comes then it is the responsibility of the Service provider to take care of. For better control of data, every organization tries to achieve better control in authentication and authorization by using various policies as shown in Table 5.1.

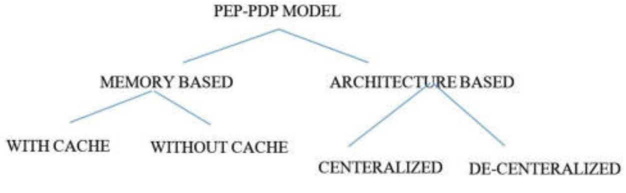


Fig. 5.1. *PEP-PDP Model types*

Table 5.1
Various Authentication and Authorization Policies Available

Under Indian IT Laws	
Various Authentication Policies available are	Various Authorization Policies available are
Elliptic curve cryptography	WBAC (Team-based)
Bio Hashing	GRBAC
Delffie-Hellman	iCanCloud simulation platform
Triple DES	XML
Simple-Homomorphic Encryption	RBAC
Clustering	ABAC

As the data Sharing between Cyber-Physical space [37, 28] is increasing day by day, there is a need to check the Privacy of data. The requirement to Modify existing Access and Authorization Policies is the need of time. Under the Indian IT act 2008 and its various amendments, it is the responsibility to keep track and maintain all transaction logs for any future need for this Section 7 and section 14 is provided, which says to keep track of all electronic transactions securely.

Whenever a request is received at DbaaS Cloud it should be able to differentiate between Request is from User or Cyber-Physical Device. Existing Rules Stored in PEPs are not applicable when requested data is from Cyber-Physical devices. Access Control Policies [12, 50] should be able to handle requests from Cyber-Physical Device. Every request to data is to be scanned and a new Rule is required to allow access to data. Whenever many Cyber-Physical Devices are requesting access to data, Existing systems are not able to handle requests at a large scale as proposed in [56]. There is a requirement to store every transaction done at DbaaS Cloud and the system should be able to handle a Large Number of Requests.

6. Proposed User and Cyber-Physical Device Transaction Authentication Mechanism (U & CPD TAM). Every authentication and authorization policy achieves its intended purpose up to a certain level when we see statistics of Data leakage and user Privacy crimes [25, 40, 47]. For better management of user data on Distributed networks in Cloud Databases, Management of user data is to be controlled by the technique which is based on Distributed networks. Blockchain can help in providing better control in terms of user authentication and authorization as shown in Table 6.1 as required in Section 7 and 14 Under Indian IT Act 2008 and its amendments.

Requests to access Data from User and Cyber-Physical Device is received at DbaaS Cloud, Insiders at the DbaaS cloud can Access to data illegally and can take benefits. The benefits of blockchain in this regard can be viewed in [51, 60, 59]. Blockchain server can keep an eye on Insider activities from Log-in to User Authorization. Insider Log-in Control Protocol using Blockchain is already published in our previous paper [22].Blockchain server also stores the data of user authorization as well as logs in detail. In any case, Insider tries to change Authorization Rules, its activity is stored in the Blockchain node as User Transaction data as shown in Figure 6.1.

Blockchain will help in monitoring the activities of Insider From authentication to authorization. Any uneven activity can be easily tracked and responsibility can be fixed with evidence. Each Request to access data from Insider goes to PEP which allows Access to data only if the rule for data access is available for that Insider. If Rule is not available at PEP, that request to access data from an insider is forwarded to PDPs, To

Table 6.1
Use of Blockchain at different levels

Request to Access Data from User / Cyber-Physical Device			
Authentication		Authorization	
Insider	Outsider	Insider	Outsider
Blockchain Technique For User Authentication		Authorization Technique with Blockchain for Authorization Policy Request Tracking	
Blockchain server for cloud database for Section 7 and 14 (Indian IT Act 2008 and its amendments)			
Under Indian IT Laws			

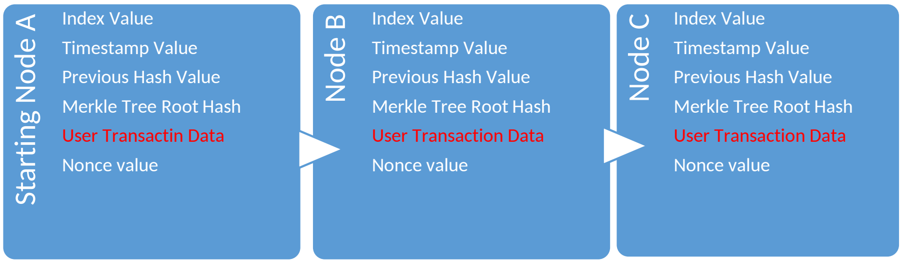


Fig. 6.1. *User Transaction Data in Blockchain Node*

Track each request for particular data, Blockchain can be used along with PDPs.

In our Proposed Algorithm 1, Working of the Dependency checkpoint and Internal threat Detection unit proposed in [57, 58, 56] in their research work is same we have introduced the concept of Distributed PEP-PDP Model along with Blockchain Servers for tracking every request forwarding between PEPs and PDPs For better control according to Section 7 and 14 Under Indian IT Act 2008 and its amendments.

Algorithm 1 Insider Threat Prevention Algorithm for Distributed PEP-PDP Environment with side caching Model

Input: An Insider Alice request Q for accessing a data item D, Q (Alice D), the PEP that receives Alice request, request may be forwarded to other other PEPs in the System S=PEP1, PEP2,.....PEPn, along with these common caches (CPEP1, CPEP2,...., CPEPn), Dependency checkpoint DCP available at designated PDP for each set of PEPs.

Output:Access Decision (Grant or Reject)

```
1: If Q (Alice, D) does not exist in corresponding PEP Cache then Go to Step 2 else Go to Step 5
2: If Q (Alice, D) does not exist in other PEP Cache then Go to Step 3 else Go to Step 4
3: Forward Alice request to Designated PDP by Calling Algorithm 2
4: Fetch Q (Alice, D) and Forward it to Corresponding CPEP Cache and Go to Step 5
5: Send Request D to Dependency-CheckPoint for checking Dependencies and Go to Step 6
6: If D can be combined with K to infer information then Go to Step 7  else Go to Step 9
7: If Check Alice has a cached value of K then it may be a possible threat then Forward Alice request to
   Designated PDP by calling Algorithm 2 else Go to Step 8
8: If Alice is not the cached value of K No threat found, re-issue the PEP cache response for Alice request to
   D then Go to Step 9
9: If D cannot be combined with K to infer information No threat found, then re-issue the PEP cache response
   for Alice request to D
```

Algorithm 2 PEP Request Authentication using Blockchain Mechanism.

Input:Request Q received at Blockchain Server of Designated PDP Server for each PEP, It checks for Q Request is from PEP.

Output:Access Granted or Rejected to‘ PEP Request by Bloackchain Server

```
1: If Request == PEP then Go to Step 2 else Go to step 5
2: If Login ID & Signature == Valid  then continue this step  else Go to step 5
3: If current index value > Last stored index vale & Hash value & Timestamp value & Nonce value== Valid
   then Create New Blockchain node with requested transaction details and Go to Step 4  else Go to step 5.
4: Grant Authentication with Update to a user record in Blockchain Database and Send Request to PDP
   Server by Calling Algorithm 3
5: Give error message and Exit
```

```
Algorithm 3 PEP Request is forwarded to Designated PDP Server.
    Input:Request Q received at Designated PDP Server for each PEP, It Checks for Possible threat at ITDU
    Output:Request Approved or Rejected
1: If Designated PDP, using the ITDU, decides that there are no threat exists then Alice is allowed to get his/her requested D, all CPEPs receive Designated PDP decision and Corresponding CPEP allows the user to get his requested D. else Go to Step 2
2: If Designated PDP, Using ITDU decides that a threat exists then Alice is not allowed to get his/her requested D,Request is rejected by Designated PDP.Response is updated in all CPEPs
```

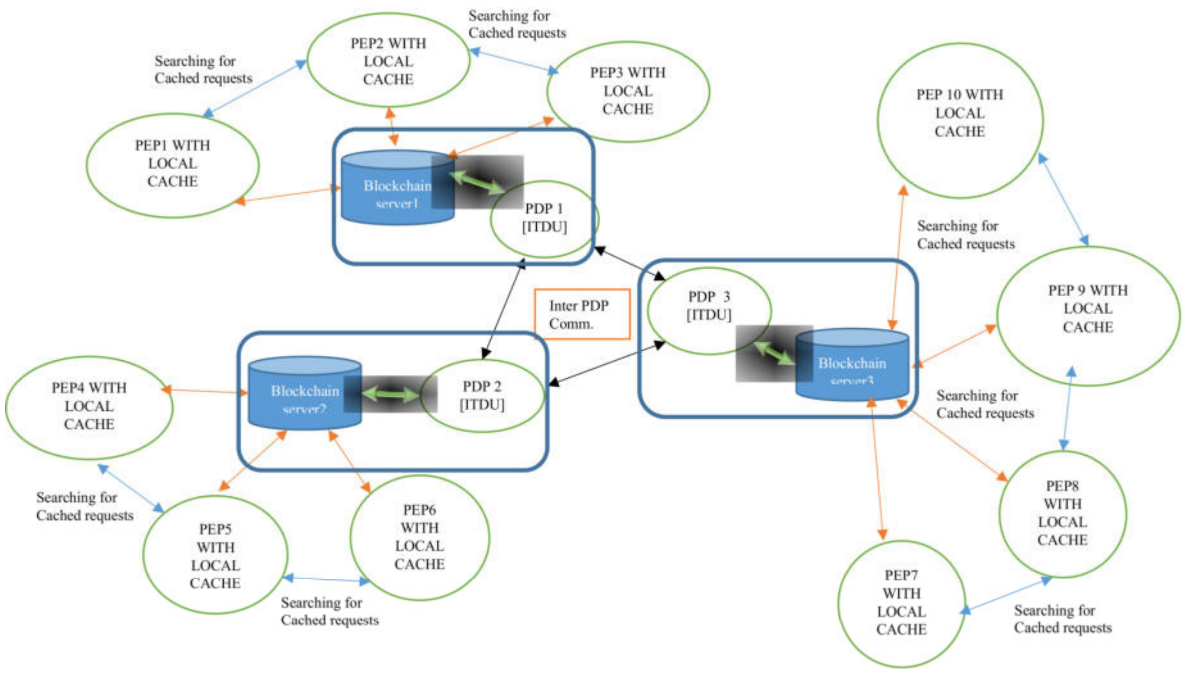
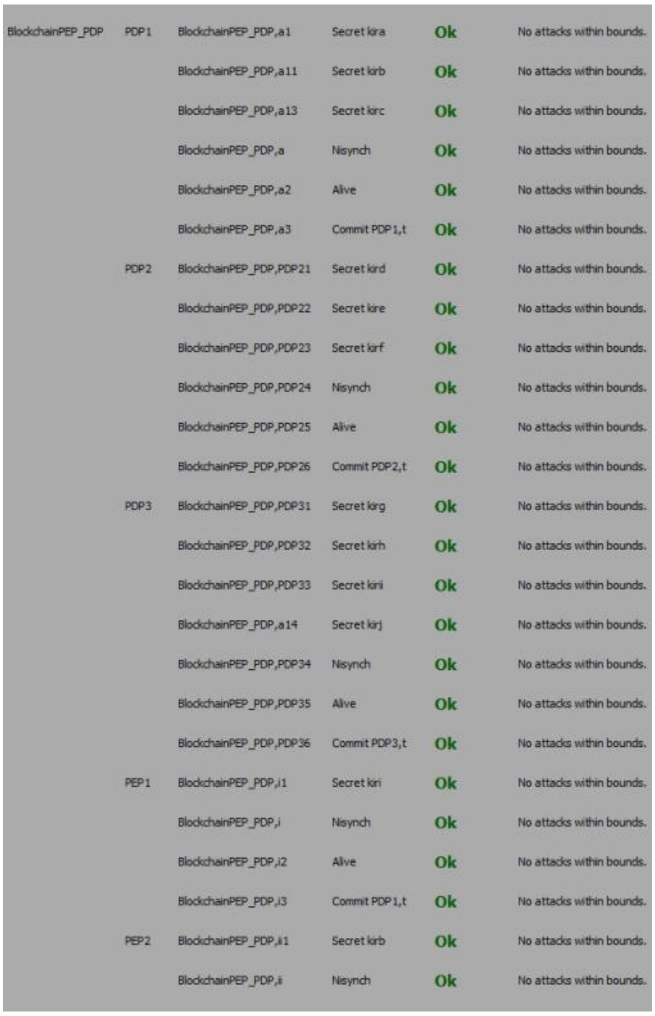


FIG. 6.2. *Working of Proposed Private Blockchain Servers PEP-PDP Architecture*

Blockchain will also help in tracking every request coming from PEPs to its designated PDP. The detailed proposed architecture is shown in Figure 6.2. For every Request from Insider to PEP, it is checked at the Dependency checkpoint for possible threat, if no threat was detected then Insider is allowed to access data according to cached Copy at PEP otherwise request to access data is forwarded to Designated PDP. First at PDP Side request is received by Blockchain Server where it was checked for authenticity by checking PEPs UID and Signature. If it was valid then a new node is created at Blockchain to track all details of the transaction and request is forwarded to Designated PDP. If it was Found Invalid, Request from PEP is rejected and an error message is conveyed.

The benefit of using Multiple PDPs is to have Better Scalability and better Response Time as compared to Existing architecture proposed in [56]. Better control of transactions can be achieved by using Blockchain Server where detail about any transaction can be fetched.

7. Experimentation Results. The proposed architecture is checked on the Verification and Validation Formal Tool Scyther by using Asymmetric keys. Each PEP to PDP communication is done by using Asymmetric keys. This Formal Tool verifies the proposed Protocol against all the Security Protocols. Whenever this tool detects any Attack in the proposed Protocol it creates an attack graph for better understanding. For having the best Security Requirements in place it uses Four Claims namely Alive, Nisynch, Secret and Commitment



BlockchainPEP_PDP	PDP1	BlockchainPEP_PDP,a1	Secret kira	Ok	No attacks within bounds.
		BlockchainPEP_PDP,a11	Secret kirb	Ok	No attacks within bounds.
		BlockchainPEP_PDP,a13	Secret kirc	Ok	No attacks within bounds.
		BlockchainPEP_PDP,a	Nisynch	Ok	No attacks within bounds.
		BlockchainPEP_PDP,a2	Alive	Ok	No attacks within bounds.
		BlockchainPEP_PDP,a3	Commit PDP1,t	Ok	No attacks within bounds.
	PDP2	BlockchainPEP_PDP,PDP21	Secret kird	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP22	Secret kire	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP23	Secret kirf	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP24	Nisynch	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP25	Alive	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP26	Commit PDP2,t	Ok	No attacks within bounds.
	PDP3	BlockchainPEP_PDP,PDP31	Secret kirg	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP32	Secret kirh	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP33	Secret kirii	Ok	No attacks within bounds.
		BlockchainPEP_PDP,a14	Secret kirj	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP34	Nisynch	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP35	Alive	Ok	No attacks within bounds.
		BlockchainPEP_PDP,PDP36	Commit PDP3,t	Ok	No attacks within bounds.
	PEP1	BlockchainPEP_PDP,i1	Secret kiri	Ok	No attacks within bounds.
		BlockchainPEP_PDP,i	Nisynch	Ok	No attacks within bounds.
		BlockchainPEP_PDP,i2	Alive	Ok	No attacks within bounds.
		BlockchainPEP_PDP,i3	Commit PDP1,t	Ok	No attacks within bounds.
	PEP2	BlockchainPEP_PDP,ii1	Secret kirb	Ok	No attacks within bounds.
		BlockchainPEP_PDP,ii	Nisynch	Ok	No attacks within bounds.

FIG. 7.1. *The Output For the Scyther Claim Test for Multiple PDPs and PEPs*

[19, 20, 54]. Intended Communication is achieved by having some events that are described as “Alive”. Nisynch stands for non-injective synchronization, it means that the receiver receives the messages from the sender in a synchronized manner. Commitment is a promise that is made by one party to the other. The confidentiality of data is achieved by using Claim Secret.

Results of the Proposed Protocol in Scyther are shown in Figure 7.1. It can be seen from the Result that Status is Ok which means there are No Attacks within Bounds. All four claims Alive, Nisynch, Secret and Commitment are achieved and verified. In this Proposed, Protocol 10 PEPs along with 3 PDPs and 3 Blockchain servers are tested against possible attacks.

8. Conclusion. Effective control of data on the cloud is the need of the hour. Many companies ran into losses due to data theft on the cloud. According to the Indian IT act Company managing Cloud is responsible for data theft occurred on the cloud. Employees working in companies may steal data from the cloud and put the company in a bad image. To control the activities of Employees in cloud managing companies, a strong mechanism is required. This Paper proposes Blockchain as a solution to control the activities of Employees from authentication to Authorisation. Request for Data Access from User and Cyber-Physical Device is Received at DbaasS Cloud. Policies to data Access are managed by Cloud Employees. Employees can take advantage, for performing Insider Attack.

To have better Control under section 7 and 14 under the Indian IT Act and its amendments this paper proposes the concept of Distributed PEP-PDP Model along with Blockchain Servers for tracking every request forwarding between PEPs and PDPs. Each Transaction between PEP and PDP is now tracked with the possible results. The proposed Protocol is tested on Scyther Formal Tool for possible attacks. From the results, it is concluded that the proposed system is highly efficient and robust. The proposed system is ready to be implemented in the actual scenario by providing better control of transactions between PEPs and PDPs. In future work, work will focus on better transaction control on the PEP side itself using Blockchain.

REFERENCES

[1] M. Abomhara, H. Yang, G. M. Køien, and M. B. Lazreg, *Work-based access control model for cooperative healthcare environments: Formal specification and verification*, Journal of Healthcare Informatics Research, 1 (2017), pp. 19–51.

[2] M. Alam, N. Emmanuel, T. Khan, Y. Xiang, and H. Hassan, *Garbled role-based access control in the cloud*, Journal of Ambient Intelligence and Humanized Computing, 9 (2018), pp. 1153–1166.

[3] R. Amin and G. Biswas, *A novel user authentication and key agreement protocol for accessing multi-medical server usable in tmis*, Journal of medical systems, 39 (2015), p. 33.

[4] A. Anakath, S. Rajakumar, and S. Ambika, *Privacy preserving multi factor authentication using trust management*, Cluster Computing, 22 (2019), pp. 10817–10823.

[5] B. M. Babu and M. S. Bhanu, *Prevention of insider attacks by integrating behavior analysis with risk based access control model to protect cloud*, Procedia Computer Science, 54 (2015), pp. 157–166.

[6] N. Baracaldo, B. Palanisamy, and J. Joshi, *G-sir: an insider attack resilient geo-social access control framework*, IEEE Transactions on Dependable and Secure Computing, 16 (2017), pp. 84–98.

[7] T. Bhatia and A. Verma, *Data security in mobile cloud computing paradigm: a survey, taxonomy and open research issues*, The Journal of Supercomputing, 73 (2017), pp. 2558–2631.

[8] H. Bleau, *Current state of cybercrime in 2016.* rsa.com, 2016.

[9] P. by Government of India, *The information technology act, 2000.* meity.gov.in, 2000.

[10] ———, *The information technology(amendment) act, 2008.* meity.gov.in, 2009.

[11] ———, *The information technology rules,2011.* meity.gov.in, 2011.

[12] Y. Cao, Z. Huang, Y. Yu, C. Ke, and Z. Wang, *A topology and risk-aware access control framework for cyber-physical space*, Frontiers of Computer Science, 14 (2020), pp. 1–16.

[13] K. Chatterjee et al., *Biometric re-authentication: An approach towards achieving transparency in user authentication*, Multimedia Tools and Applications, 78 (2019), pp. 6679–6700.

[14] P. Chattopadhyay, L. Wang, and Y.-P. Tan, *Scenario-based insider threat detection from cyber activities*, IEEE Transactions on Computational Social Systems, 5 (2018), pp. 660–675.

[15] S. A. Chaudhry, I. L. Kim, S. Rho, M. S. Farash, and T. Shon, *An improved anonymous authentication scheme for distributed mobile cloud computing services*, Cluster Computing, 22 (2019), pp. 1595–1609.

[16] F. Chen, Y. Luo, J. Zhang, J. Zhu, Z. Zhang, C. Zhao, and T. Wang, *An infrastructure framework for privacy protection of community medical internet of things*, World Wide Web, 21 (2018), pp. 33–57.

[17] G. Chen, B. Xu, M. Lu, and N.-S. Chen, *Exploring blockchain technology and its potential applications for education*, Smart Learning Environments, 5 (2018), p. 1.

[18] K. Chung, H. Yoo, D. Choe, and H. Jung, *Blockchain network based topic mining process for cognitive manufacturing*, Wireless Personal Communications, 105 (2019), pp. 583–597.

[19] C. J. Cremers, *The scyther tool: Verification, falsification, and analysis of security protocols*, in International conference on computer aided verification, Springer, 2008, pp. 414–418.

[20] C. J. F. Cremers, *Scyther: Semantics and verification of security protocols*, Eindhoven University of Technology Eindhoven, Netherlands, 2006.

[21] J. R. David Reinsel, John Gantz, *The digitization of the world from edge to core.* seagate.com, 2018.

[22] G. Deep, R. Mohana, A. Nayyar, P. Sanjeevikumar, and E. Hossain, *Authentication protocol for cloud databases using blockchain mechanism*, Sensors, 19 (2019), p. 4444.

[23] Z. Dou, I. Khalil, A. Khreishah, and A. Al-Fuqaha, *Robust insider attacks countermeasure for hadoop: Design and implementation*, IEEE Systems Journal, 12 (2017), pp. 1874–1885.

[24] S. Eberz, K. B. Rasmussen, V. Lenders, and I. Martinovic, *Looks like eve: Exposing insider threats using eye movement biometrics*, ACM Transactions on Privacy and Security (TOPS), 19 (2016), pp. 1–31.

[25] J. Foster, *1 terrifying cyber crime statistics.* dataconnectors.com, 2018.

[26] A. Gabillon, R. Gallier, and E. Bruno, *Access controls for iot networks*, SN Computer Science, 1 (2020), p. 24.

[27] M. Habiba, M. R. Islam, A. S. Ali, and M. Z. Islam, *A new approach to access control in cloud*, Arabian Journal for Science and Engineering, 41 (2016), pp. 1015–1030.

[28] V. Hahanov, *Cyber physical computing for IoT-driven services*, Springer, 2018.

[29] H. Han, M. Huang, Y. Zhang, and U. A. Bhatti, *An architecture of secure health information storage system based on blockchain technology*, in International Conference on Cloud Computing and Security, Springer, 2018, pp. 578–588.

[30] S.-M. Jo, *Secure access policy for efficient resource in mobile computing environment*, Journal of Computer Virology and Hacking Techniques, 13 (2017), pp. 297–303.

[31] S. Kalra and S. K. Sood, *Secure authentication scheme for iot and cloud servers*, Pervasive and Mobile Computing, 24 (2015), pp. 210–223.

[32] S. Kumar, S. K. Singh, A. K. Singh, S. Tiwari, and R. S. Singh, *Privacy preserving security using biometrics in cloud computing*, Multimedia Tools and Applications, 77 (2018), pp. 11017–11039.

[33] S. Kumari, M. Karuppiah, A. K. Das, X. Li, F. Wu, and N. Kumar, *A secure authentication scheme based on elliptic curve cryptography for iot and cloud servers*, The Journal of Supercomputing, 74 (2018), pp. 6428–6453.

[34] K. H. Kwak, J. T. Kong, S. I. Cho, H. T. Phuong, and G. Y. Gim, *A study on the design of efficient private blockchain*, in International Conference on Computational Science/Intelligence & Applied Informatics, Springer, 2018, pp. 93–121.

[35] H. Li, L. Pei, D. Liao, G. Sun, and D. Xu, *Blockchain meets vanet: An architecture for identity and location privacy protection in vanet*, Peer-to-Peer Networking and Applications, 12 (2019), pp. 1178–1193.

[36] W. Meng, W. Li, Y. Wang, and M. H. Au, *Detecting insider attacks in medical cyber–physical networks based on behavioral profiling*, Future Generation Computer Systems, (2018).

[37] D. P. Möller, *Introduction to cyber-physical systems*, in Guide to Computing Fundamentals in Cyber-Physical Systems, Springer, 2016, pp. 81–139.

[38] C. S. Moon, S. Chung, and B. Endicott-Popovsky, *A cloud and in-memory based two-tier architecture of a database protection system from insider attacks*, in International Workshop on Information Security Applications, Springer, 2013, pp. 260–271.

[39] M. Nofer, P. Gomber, O. Hinz, and D. Schiereck, *Blockchain. business & information systems engineering, 59 (3), 183–187*, DOI: http://dx. doi. org/10.1007/s12599-017-0467-3, (2017).

[40] M. Powell, *11 eye opening cyber security statistics for 2019.* cpomagazine.com, 2019.

[41] U. D. o. C. Published by NIST, *Nist cloud computing standards roadmap.* nist.gov, 2018.

[42] N. C. Rathore and S. Tripathy, *A trust-based collaborative access control model with policy aggregation for online social networks*, Social Network Analysis and Mining, 7 (2017), p. 7.

[43] J. H. Ryu, P. K. Sharma, J. H. Jo, and J. H. Park, *A blockchain-based decentralized efficient investigation framework for iot digital forensics*, The Journal of Supercomputing, 75 (2019), pp. 4372–4387.

[44] A. Shaghaghi, S. S. Kanhere, M. A. Kaafar, E. Bertino, and S. Jha, *Gargoyle: A network-based insider attack resilient framework for organizations*, in 2018 IEEE 43rd Conference on Local Computer Networks (LCN), IEEE, 2018, pp. 553–561.

[45] A. Shajina and P. Varalakshmi, *A novel dual authentication protocol (dap) for multi-owners in cloud computing*, Cluster Computing, 20 (2017), pp. 507–523.

[46] S. Shin and T. Kwon, *Aana: Anonymous authentication and authorization based on short traceable signatures*, International journal of information security, 13 (2014), pp. 477–495.

[47] R. SOBERS, *110 must-know cybersecurity statistics for 2020.* varonis.com, 2020.

[48] J. Sun, J. Yan, and K. Z. Zhang, *Blockchain-based sharing services: What blockchain technology can contribute to smart cities*, Financial Innovation, 2 (2016), pp. 1–9.

[49] J.-L. Tsai and N.-W. Lo, *A privacy-aware authentication scheme for distributed mobile cloud computing services*, IEEE systems journal, 9 (2015), pp. 805–815.

[50] M. Uriarte, J. Astorga, E. Jacob, M. Huarte, and O. López, *Survey on access control models feasible in cyber-physical systems*, in Cyber-Physical Systems: Architecture, Security and Application, Springer, 2019, pp. 103–152.

[51] B. van Lier, *Blockchain technology: The autonomy and self-organisation of cyber-physical systems*, in Business Transformation through Blockchain, Springer, 2019, pp. 145–167.

[52] B. Vinod, *Blockchain in travel*, Journal of Revenue and Pricing Management, 19 (2020), pp. 2–6.

[53] Z. Wu, G. Xu, C. Lu, E. Chen, F. Jiang, and G. Li, *An effective approach for the protection of privacy text data in the clouddb*, World Wide Web, 21 (2018), pp. 915–938.

[54] H. Yang, V. A. Oleshchuk, and A. Prinz, *Verifying group authentication protocols by scyther.*, JoWUA, 7 (2016), pp. 3–19.

[55] T.-C. Yang, N.-W. Lo, H.-T. Liaw, and W. C. Wu, *A secure smart card authentication and authorization framework using in multimedia cloud*, Multimedia Tools and Applications, 76 (2017), pp. 11715–11737.

[56] Q. Yaseen, Y. Jararweh, B. Panda, and Q. Althebyan, *An insider threat aware access control for cloud relational databases*, Cluster Computing, 20 (2017), pp. 2669–2685.

[57] Q. Yaseen and B. Panda, *Predicting and preventing insider threat in relational database systems*, in IFIP International Workshop on Information Security Theory and Practices, Springer, 2010, pp. 368–383.

[58] ———, *Insider threat mitigation: preventing unauthorized knowledge acquisition*, International Journal of Information Security, 11 (2012), pp. 269–280.

[59] Y. Zhao, Y. Li, Q. Mu, B. Yang, and Y. Yu, *Secure pub-sub: Blockchain-based fair payment with reputation for reliable cyber physical systems*, IEEE Access, 6 (2018), pp. 12295–12303.

[60] D. Zhaoyang, L. Fengji, and G. Liang, *Blockchain: a secure, decentralized, trusted cyber infrastructure solution for future energy systems*, Journal of Modern Power Systems and Clean Energy, 6 (2018), pp. 958–967.

Edited by: Anand Nayyar
Received: Apr 13, 2020
Accepted: May 21, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 425–440, DOI 10.12694:/scpe.v21i3.1741

INTERNET OF VEHICLES (IOV) OVER VANETS: SMART AND SECURE COMMUNICATION USING IOT

SUMIT KUMAR[*]AND JASPREET SINGH[†]

Abstract. The new age of the Internet of Things (IoT) is motivating the advancement of traditional Vehicular Ad-Hoc Networks (VANETs) into the Internet of Vehicles (IoV). This paper is an overview of smart and secure communications to reduce traffic congestion using IoT based VANETs, known as IoV networks. Studies and observations made in this paper suggest that the practice of combining IoT and VANET for a secure combination has rarely practiced. IoV uses real-time data communication between vehicles to everything (V2X) using wireless communication devices based on fog/edge computing; therefore, it has considered as an application of Cyber-physical systems (CPS). Various modes of V2X communication with their connecting technologies also discussed. This paper delivers a detailed introduction to the Internet of Vehicles (IoV) with current applications, discusses the architecture of IoV based on currently existing communication technologies and routing protocols, presenting different issues in detail, provides several open research challenges and the trade-off between security and privacy in the area of IoV has reviewed. From the analysis of previous work in the IoV network, we concluded the utilization of artificial intelligence and machine learning concept is a beneficial step toward the future of IoV model.

Key words: IoT, VANET, IoV, Secure communication, Traffic congestion, Artificial Intelligence techniques, Cyber-physical systems (CPS)

AMS subject classifications. 68M11

1. Introduction. The Internet of Things (IoT) refers to physical devices equipped with sensors, such as smart wearables, autonomous vehicles, mobile phones, home appliances, machines, and other electronic devices connected via an application programming interface (API) for data transmission over the Internet [1]. When vehicles are connected to the Internet and act as an ad-hoc network, it is known as the Internet of Vehicles (IoV). It is emerging as an innovative model in the wireless and mobile communications sectors with a resolution of new communication and connectivity technologies assisted by the development of IoT [2]. Vehicular Ad-hoc network (VANET) gave rise to the IoV and it refers to the network of dissimilar entities road transport, such as vehicles, foot-travelers, roads, parking lots and city infrastructure and offers real-time communication among them. The IoV is an IoT application that offers a solution for the flow control of traffic and secure communication in cities based on the technology [3]. The increment of the vehicle connectivity to IoT results in the formation of the IoV network. This is a developing field for the automotive industries and one of the significant aspects of the smart cities which helps to monitor the traffic. It is a scattered network that provisions the usage of data formed by linked vehicles and VANETs [4]. The increase in the people drives vehicles results in the corresponding increment of the fatality which occurs because of accidents. A significant objective of the IoV is to permit vehicles to communicate in real-time with their human drivers, foot-travelers, other vehicles, roadside set-up and fleet supervising systems[5]. The IoV supports different types of communication within the network as Vehicle-vehicle (V2V), Vehicle-sensors (V2S), Vehicle-infrastructure (V2I), Vehicle-road side (V2R), Vehicle-cloud (V2C), Vehicle-network (V2N), Vehicle-pedestrian (V2P), Vehicle-devices (V2D) communication. V2V wireless communication is the transfer of information regarding the position and speed of the surrounding vehicle. V2S technology enables sensor communication with neighbor vehicles using pre-installed On-Board Units (OBUs). The V2I is used as an IoT sensor to monitor vehicle internal performance through OBUs. V2R

[*] Research Scholar, Computer Science and Engineering Department, Chandigarh University, Mohali, Punjab, India, And Assistant Professor, Computer Science and Engineering Department, CGC College of Engineering, Mohali, Punjab, India (kumarsumit.cse@gmail.com).

[†] Associate Professor, Computer Science and Engineering Department, Chandigarh University, Mohali, Punjab, India (cec.jaspreet@gmail.com).

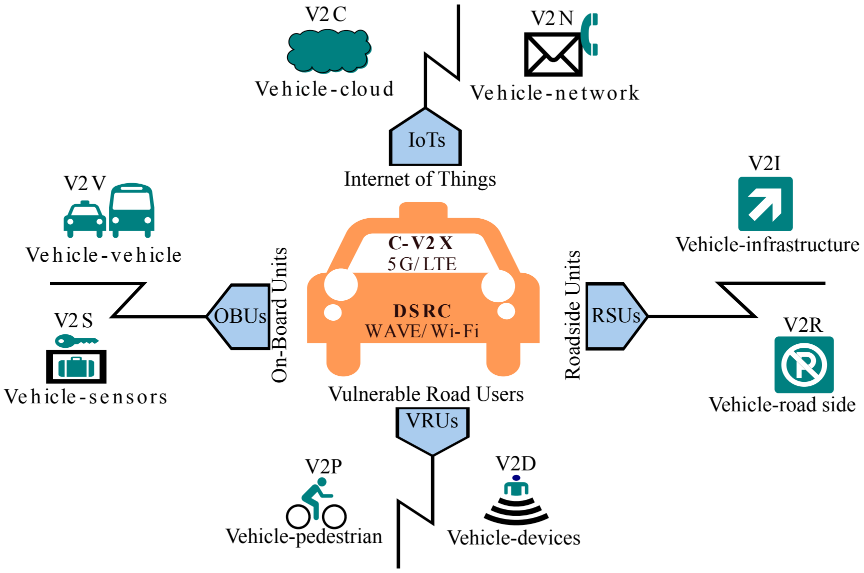


Fig. 1.1. *V2X Communication modes in IoV*

Table 1.1
Analysis of Short-range Communication Technologies in IoV

Parameters	**Bluetooth**	**UWB**	**Zigbee**	**Wi-Fi**
IEEE Specification	802.15.1	802.15.3a	802.15.4	80.2.11 a/b/g
Application	In-vehicle devices connectivity for infotainment	Enable real-time localization with other technologies	Interconnection of sensors with vehicles and RSUs	Enable V2X modes of communication
Domain	Telematics, Body	Telematics, Power train	Body	Telematics
Data-rate	01 Mbps	100 Mbps	250 Kbps	54 Mbps
Range	Up to 10 m	Up to 10 m	10 – 100 m	Up to 100 m
Power Consumption	Low	Ultra-Low	Very Low	High
Modulation Type	GFSK	BPSK, QPSK	BPSK, QPSK	BPSK, QPSK
Frequency Band	2.4 GHz	(3.1 to 10.6) GHz	(868, 915) MHz, 2.4 GHz	(2.4, 5) GHz
Max Bandwidth	Near to 1 Mbps	Near to 100 Mbps	Near to 50 Mbps	Near to 50 Mbps
Protection Techniques	CRC-16 bits	CRC-32 bits	CRC-16 bits	CRC-32 bits
Network Topologies	Spoke-hub	Peer-Peer (P2P)	Spoke-hub, Mesh	Spoke-hub

used to support the wireless exchange of information between a vehicle and supporting Roadside Units (RSUs). V2C and V2N allow the vehicle to access additional information from the cloud and fog server through APIs using the internet. V2P and V2D systems support awareness for Vulnerable Road Users (VRUs) or such as horse riders, pedestrians and cyclists having smart watches or mobile phones. Two or more than two objects communicate in real-time using many diverse technologies which makes IoV complex network. These basic types of communication in ad-hoc networks referred to as Vehicle-everything (V2X) communication that have proposed from a study on [1, 6, 4] in the Fig. 1.1.

V2X communication possible in two different ways, using Wi-Fi technology or 5G/LTE networking for IoT based Internet of Vehicles. The DSRC is responsible for communication between the RSU and the OBU by wireless connection based on WAVE standards. After considering these wireless technologies that used to maintain security and fast communication in the IoV network, some essential communication technologies based on various parameters has summarized in the Table 1.1.

This analysis has done based on work from [7, 8] in which radio technologies are superior to Zigbee and Bluetooth while considering in-vehicle applications depending on low bitrates and inadequate power source and also its low energy ingestion that could offer an extended lifetime. On the contrary, data at fast speeding-vehicle applications could help from the usage of Wi-Fi and Ultra-wideband (UWB) due to their less normalized energy consumption rate.

IoV uses technologies such as navigation systems, mobile communication, and sensor networks for data interchange and instruction systems. Cyber-physical system (CPS) is a combination of cyber (virtual) and physical (real) systems with networking and computation capabilities.

IoT applications typically include three basic layers:

- **Sensor (actuator) layer:** Use to understand (sense) road and traffic conditions.
- **Application (control) layer:** Analysis of data collected by the integration of big data with fog infrastructure in data centers.
- **Communication layer:** Smart wireless connectivity between sensors and fog servers.

In IoT, physical devices are connected over the Internet so that they can communicate with each other and make decisions intelligently and exchange information without or with little human intervention [9], for example, driverless cars or drones.

1.1. Issues and Challenges in IoV. The main focus of the Internet of Vehicles is to connect multiple users with vehicles, devices and networks, offering a safe and secure communication capability that is flexible, efficient and reliable. The construction of IoV with such multiple objects makes it a complex system [10]. Also, the use of IoV is less diverse compared to other networks and as a result, there are some particular requirements. Both of these issues add new technological challenges and test the research and development of IoV. During discussion related to the function and construction of IoV, here are some of the issues and challenges that researchers face:

- **Security and Privacy:** Since IoV combines a wide range of various services and standards, there is a requirement for the safety of information. As an open community network, IoV is aimed at cyber attacks and attacks that can cause physical loss and privacy leaks. Maintaining the balance between privacy and security is one of the key issues in IoV. The acceptance of reliable info from its sender to the recipient is important [11]. However, the senders' privacy requirement may be violated by this reliable information.
- **Vehicles Reliability:** Vehicles, sensors, and network sensors may fail sometime. The system has to deal with inaccurate data, plus malicious communications, for example denial of service (DOS) attacks. Some technologies can be deployed like Intrusion Detection Systems (IDS) to protect against attacks in traditional networks [12]. In general, car safety is very important compared to in-car entertainment.
- **Mobility and Dynamic Topology:** Compared to other vehicles in the network, cars can travel at much higher speeds, resulting in constantly changing network topologies. It requires a test to connect with consecutive nodes and transport the goods from one place to another. Therefore, the flexibility of network topology should be considered important for IoV development.
- **Open Standards:** The absence of standard can make successful V2X communication troublesome, so interoperability and standardization are required for quick selection. Receiving open standards will empower the flat-sharing of data. Governments ought to take an interest and urge enterprises to work together in the improvement of innovative prescribed procedures and open global standards.
- **Variable network load:** Network size is another big challenge, which can be very high or low due to changing traffic conditions. As the scale of the network in large urban areas can be high, for example, entries in urban areas, main highways, and metropolitan cities. In any case, if the network has severely broken can now remain fragmented, which cause road accidents. Therefore, a smart traffic surveillance system based on 5G technology will be required to solve this problem [13].
- **Geographical Communication:** Related to different networks that use multicast or unicast routing, where communication is reflected by a particular ID, car networks often have some form of transmission, which affects areas where bulk traffic should be sent.
- **Predictable Mobility:**Vehicular networks are different than different types of specially designated networks where nodes move randomly. Cars, in turn, are forced by topology and design, by the need to pay attention to traffic signals and traffic signals, and by the reaction of moving neighbour vehicles, which makes consistency as far as possible.
- **Sufficient Energy and Storage:** The general feature of nodes in Vehicular networks is that they contain more power to register (count, maintain and adjust), because nodes are in rows rather than

smaller portable gadgets.

- **Various Communication Environments:** Vehicular networks have generally used in two common communication areas. First, in the case of highways, the situation is generally basic and direct. In the case of cities, this has really problematic because the routes have isolated by divisions, such as sectors, trees, and other different obstacles [14].
- **Poor Network Connectivity:** As high mobility of devices and fast changes of network topology, leading to network partitioning and connection disappointment result in bad messages should be normal. At that time, the need for ways to joining back to the network is always being explored.
- **Hard Delay Constraints:** In IoV, many applications have severe delays, even they do not require high-bitrate or bandwidth. Such as, during a serious incident, a message must be transmitted and displayed at a specific time to avoid a car accident. Current application, instead of the usual delay, undesirable delays will be important.
- **Big data in IoT:** One of the biggest challenges is the processing then storing of the large amount of data generated in the IoV due to a high number of connected autonomous vehicles, which has assumed to be 1Gb per second of data processing [15]. Therefore, it is a vital function for big data management in intelligent connected vehicles (ICV) using cloud computing and IoT analytics.
- **High-Reliability requirement:** Transportation-related applications are often very delicate, so there is a requirement for high-reliability. Since complex systems, network-level size, and poor power of connectivity, high-reliability in IoV is difficult to achieve.
- **Service Sustainability:** Ensuring the robustness of service delivery in IoV is still difficult as it is, it calls for high-level perceptual strategies, just as it is not easy to understand the network system. There is difficulty in adapting all vehicles to provide sensible types of assistance with amazing networks gradually, as it is necessary to have network bandwidth, remote access, low service levels and host status.

These security and trust-related issues of IoV are expected to be resolved in the future to make it more reliable and successful.

1.2. Applications. Most of the IoV applications are in the field of telecom. According to Cyber-physical systems (CPS) some important applications of the IoV model and their functionality are classified into different classes, given below:

- *Safety:* Cooperative collision avoidance, Lane changing warnings, automatic braking and speed control
- *Navigation:* Real-time traffic, Route navigation, Locating parked vehicle, Cooperative driving
- *Information and Infotainment:* Wi-Fi in vehicles for downloading of music, video streaming, content sharing
- *Remote Telemetric:* Vehicle remote locking, Car surveillance
- *Diagnostic:* Service spot detection, Self-repair, Fuel usage optimization
- *Car sharing:* Car pooling, Group parking booking
- *Others:* Electronic toll payments, Traffic flow monitoring, etc.

The Internet of things and Cyber-physical systems are not isolated technologies [16]. Cyber-physical systems (CPS) and IoT in the IoV network have some other important applications, such as Intelligent transportation systems (ITS), Connected autonomous vehicles (CAV), Smart grid, Smart city, and Smart manufacturing [17].

1.3. Motivation. In the upcoming time, IoV will become the future of the VANETs based on the expansion of the web services. An essential portion of the Internet of Vehicles having different exploration fields, including intelligent transportation system, wireless communication, cloud and fog computing, mobile computing, autopilot vehicles and era of CPS [18]. On the routing and packet data transmission point of view, both network security and robust data transmission are necessary for different applications. IoV networks mostly consist of vehicles, which operate in a very different way than wireless sensors. As a result, the IoV network has many features that can influence the formation of IoV technologies. Some features will present difficulties for the advancement of technology, and some may result in benefits.

1.4. Contributions. Motivated by existing problems in IoV, we present a new approach for improved traffic management and reliable communication in IoV using behavioral studies of IoT and VANETs. The significant contribution of our work lies in summary as follows:

- In this paper, we have proposed and designed a novel framework (CSI model) for IoV with studies based on the current VANET environment to prevent traffic congestion, maintain secure communication, and maximize data delivery.
- The purpose of this study is to analyze various mechanisms for better route selection based on vehicle range, residual energy, and vehicle condition for efficient communication using artificial intelligence.
- The future challenges and current issues for the design and development of IoV networks with the help of big data, IoT, and cloud services have discussed in detail.
- This paper sheds new light on the full layered architecture of IoV for a better analysis of existing models.
- This survey presented various communication strategies for IoV by reviewing existing routing protocols, topologies, and applications.
- In conclusion, comparisons have made for the performance of existing works in terms of Quality of Service (QoS) parameters to explore better possibilities towards faster and reliable data transmission in IoV networks.

1.5. Organization. This paper has divided into five sections. Section 1 gives a brief overview of the IoV network with current issues, challenges, applications, and motivation for contributions. The rest of this paper is structured as follows. Section 2 describes an in-depth background survey explaining their respective approaches, limitations, findings, and future research gaps. Section 3 presents the layered architecture of the IoV with its routing protocols and security requirements. The performance of existing works has evaluated and a proposed novel framework presented in Section 4. The paper concludes with future scope in Section 5.

2. Background Survey. The survey of existing work for the secure communication and traffic congestion minimization in IoV is discussed in this section of the paper.

In 2019, Muhammad Asim Saleem et al [1] has researched information transmission using IoT based VANETs with congestion control mechanisms and the key highlights are listed as:

- Researchers has used the concept of basic routing protocol standards for VANETs. The proposed routing mechanism requires a large amount of energy used to consider routed between vehicles or RSUs one after another.
- Here center of attention is to intend a multiple mediator system for completing the route discovery mechanism in VANETs. Designing routing mechanism is grounded upon four mediators (intermediate communication vehicles or RSUs), which are coordinated in finding ideal paths and in reducing the traffic congestion inside the network.
- Therefore, an efficient hybrid clustering based routing mechanism is a better option for data packet transfer with IoT based VANETs to reduce traffic interruption.

In 2018, Lucy Sumi and Virender Ranga [2] designed an IoT based VANET for the controlling the traffic to help emergency vehicles in a smart urban city and some of the observed conclusions are written as follows:

- The authors introduced a structure of merged concepts from IoT and VANET that target the easy passage for precedent emergency vehicles via the traffic path, and this traffic congestion is a big issue.
- The suggested system helps the emergency vehicles and ambulances in searching the adjacent feasible route to their destination based upon the real-time traffic data.
- The proposed system supports reducing the delay in transmission time for any medical help during mishaps and the timely delivery of medicines to patients.

In 2019, Amr Tolba [3] has presented a content accessibility preference approach for improving service-optimality in IoV; the key highlights are listed as:

- The author has given a technique of obtaining data at low-delay to enhance the service-optimality of smart vehicular applications.
- By taking optimal gateways, the vehicles for getting access to data use the advantages of epidemic spread routing.

- The author has designed a content accessibility preference (CAP) prototype that helps in accurate vehicle selection.
- The effectiveness of the suggested CAP design is validated using various QoS parameters such as throughput, packet transfer or service rate, vehicle service ratio, and delay.

In 2017, Zhenyu Zhou et al. [4] presented big data-based data propagation in VANET and the key highlights are listed as:

- The authors have examined the content propagation difficulty in IoV networks based on D2D and V2V communication.
- The physical, as well as social layer data in the context of link possibility as well as social association rigidity functioning to resolve the framed combined power control, peer discovery, and channel selection problem.
- The projected scheme was associated with two experiential processes, and its efficiency and power in the sum rate, as well as content approval, were authenticated using results.

In 2017, Zhaolong Ning et al. [5] projected a cooperative QoS access scheme for Social-IoV (S-IoV) and below is some relevant points relating to this research:

- Firstly, the authors studied an active access service assessment scheme to manage through the consequence carried by the dynamic network alteration. They construct a CQS access system, concentrating on consistency assurance and service excellence promotion in Social IoVs.
- Secondly, the paper offered a social association assessment technique for exploring internal as well as external similarities between vehicles. Additionally, they examined a forecast technique consistent with the vehicle movement path for communication time valuation.
- In the end, they presented a CQS technique, which initially constructed a node-centric formation of hierarchy and arrangement to calculate the contact excellence, and then selects a contact path based upon the existing state of the network. Also, the bi-direction buffering procedure is inspected to enhance the response competence as well as procedure accuracy.

In 2016, Jiafu Wanet al. [6] projected mobile crowd-sensing for traffic forecasts in IoV and the observed conclusions are written as follows:

- Originally from the perspective of service relationships between fog computing and IoV, the authors have talked about the classification of cloud or fog-based Internet of vehicles.
- Authors have projected mobile crowd-sensing for traffic forecasts on the Internet of Vehicles (IoV).

In 2017, Wenchao Xu et al. [19] plotted the Internet of Vehicles (IoV) in a Big Data era to analyze relationships between them and some of the observed conclusions are written as follows:

- The authors highlighted and extended the significant role of the IoV using big data for autonomous vehicles.
- In addition, they identified emerging problems in IoV to demonstrate several required guidelines for the future Internet of vehicles (IoV) in a big data period.

In 2018, Xiaojie Wang et al. [20] offered a reasonable solution that allows off-loading for real-time traffic organization in Edge-based Internet of vehicle (IoV), to reduce the system average execution time. The key points related to this study are as follows:

- Firstly, offered a reasonable solution, which allows for off-loading real-time traffic organization in Edge-based Internet of Vehicles (IoV) to reduce the average execution time of the system.
- The authors have not investigated how to use external vehicles from a range of roadside-units (RSUs) communication to offload the data from fog nodes for traffic management systems (TMS).

In 2016, Jiawen Kang et al. [21] purposed two computer-generated mechanisms for plotting outbreaks and safeguarding for IoV. And the key highlights are listed as:

- Firstly, present the IoV prototype and then designate location concealment and plotting by two representative virtual machines.
- In an active topology, the communication rules in a secrecy protection order among local or in-house clouds setup and vehicles should be secure and effective to reduce the usage of the framework.

In 2017, Jiawen Kang et al. [22] presented privacy-preserved pseudonym plan for Edge computing-based IoV network. And below are some relevant points relating to this research:

- They introduced path info secrecy issues on the Internet of Vehicles with a new model called Fog or Edge-based Internet of Vehicles (F-IOV) for successful pseudonym organization using devices on the fog layer.
- Pseudonyms are created and spread to automobiles with time for safe communication and privacy management in IoV.
- Their proposed system is not much valid for conditions related to light vehicles, also does not study social networks.

In 2016, Eun-Kyu Lee et al. [23] presented the IoV from a smart grid to self-directed fog or edge-based vehicles. And the key highlights are listed as:

- This editorial appealed that the vehicular cloud model which is equal to Internet-based cloud structure for vehicles, would be the essential framework surrounding that creates the progress probable and that the self-directed manipulating would be the main recipient in the cloud planning.
- The authors introduced a vehicle cloud model for future research, which emphasizes the use of unmanned aerial vehicles (UAVs) and also explained the future research approaches.

In 2017, Wenyu Zhang et al. [24] purposed a model of cooperative fog computing amid Big Data on the Internet of Vehicles. And below are some relevant points relating to this research:

- They presented difficulties with cloud-based IoV networks, and planned a local IoV design for low-delay communication services.
- They discussed resource management scheme for this improvement in IoV designed, including intra-edge power-efficient and inter-age QoS-efficient resource management.
- Further, they provided the necessary simulation outcomes to prove the efficiency of the model employed.

In 2015, Anand Paul et al. [25] purposed a helpful perceptive intellect for the internet of vehicles. And some of the observed conclusions are written as follows:

- There are conflicts among better vehicular communication and high wireless mobility with a deficiency of computational resources and low bandwidth.
- Therefore it requires an inventive Cognitive Radio (CR) and effective spectrum management as well.
- The projected scheme utilized to reduce both high vehicular mobility as well as the spectrum deficiency problems.

In 2018, Priyan Kumar et al. [26] offered an active traffic management plan with the help of the Internet of Vehicles for optimal route selection. And the key highlights are listed as:

- The road map presented here is further divided into small numbers of groups. In addition, the ant colony optimization (ACO) algorithm is used on these maps to find the best routes.
- Additionally, they have proposed a fuzzy logic function for estimating huge traffic capacity and designing.

2.1. Research Gaps. After careful analysis of the literature survey and findings discussed above, various solutions have been found for smart and secure communication, which provides better results in reducing traffic congestion. And the following main highlights are as follows:

- It can be deduced from the above existing works that, routing is used the concept of intermediate vehicles is a time-consuming process. A clustering-based routing mechanism would be a better solution to achieve the maximum data delivery ratio [1].
- Researchers have proposed various routing mechanisms such as proactive, reactive, hierarchical, and hybrid to prevent traffic congestion in VANETs. In such cases, both normal messages, as well as emergency messages, sent with the same delay in transmission [2].
- The existing mechanism for IoT-based VANET uses a general algorithmic framework that needs to be updated with the concept of Artificial intelligence and optimization to minimize delay in transmission [3, 4].
- IoT is one of the most emerging and innovative areas that need to be implemented in VANETs or existing ad hoc networks for secure communication, accident prevention, and traffic management [6, 5].
- A challenging area in the field of IoT based VANET is network protection from various attacks and secure communication, not found in existing works [21].
- The present VANET has a low data delivery rate due to the complex interaction between human

Table 2.1
Survey of Related Work

Year	Author/s	Proposed Work and Approach	Limitations
2019	Muhammad Asim Saleem et al. [1]	Data transmission using IoT based VANETs with congestion control mechanisms. Approach: MAC protocol with mediator selection method.	The expected routing mechanisms in practice comprise a large network field, where the mechanisms required a large amount of energy to visit one by one vehicle/RSU en-route. Hence, an effective hybrid clustering based routing mechanism for packets information transfer utilizing IoT driven VANETs to reduce vehicular congestion is a better option.
2018	Lucy Sumi and Virender Ranga [2]	Designed an IoT based VANET used for traffic controlling structure in a smart city for emergency vehicles. Approach: IoT based VANET.	Authors do not consider a secure reliable routing mechanism for controlling traffic influenced by attacks or intentional obstruction, which itself is exceptional and should be studied.
2019	Amr Tolba [3]	Content accessibility preference (CAP) approach for improving service optimality in internet of vehicles. Approach: CAP model with Epidemic spread routing (ESR).	The author has suggested a technique of obtaining data at low delay to enhance the service optimality of smart vehicular applications. By choosing optimal gateways, the vehicles for getting access to data exploit the advantages of ESR.
2016	Zhenyu Zhou et al. [4]	Presented dissemination of social Big Data information in IoV. Approach: Hybrid V2V-D2D driven Internet of Vehicles.	The experimental results of this designed network show that the system efficiency is increased by utilizing the hybrid approach, but the routing overhead problems of the network are not solved in case of a large number of vehicles.
2016	Z Ning et al. [5]	Proposed a co-operative quality-aware, S-IoV access service model. Approach: SIoV and co-operative Quality Aware Service Access Structure (CQS) model.	The work has been conducted assuming a network of fixed number of vehicles. In the case of new vehicles, which want to join the network during the simulation, the achieved system performance is degraded for high bandwidth requirements.
2017	Jiafu Wanet et al. [6]	Mobile crowdsensing for traffic prediction in IoV. Approach: IoT, data aggregation method, cloud computing.	The proposed model overcomes the deficiency of the existing model by using the data aggregation in case of less communicating devices, although for a large amount of communicating devices the rate of data transmission is reduced.

behavior and vehicles. As driving behavior shows the influence on human decisions, intelligent vehicles can not only act as human, but also make decisions according to requirements [22, 23].

The main issues not considered in the existing works are secure communication and network security, after which vehicle traffic congestion is another major problem for any developed or developing country. With the increment in the number of connected cars, a huge amount of data plus traffic also expected [27]. Due to traffic congestion a lot of inconveniences such as road accidents, the amount of fuel consumed by vehicles is large, air and sound pollution, damage to the vehicles occur. To analyzing precise traffic congestion control mechanisms are very much required to provide efficient and sophisticated functioning of a network with security. Web of things (IoT) helping the transformation of conventional VANETs into the intelligent vehicular ad-hoc networks (InVANETs) for Device-device (D2D) communication [28]. This helps improve road safety and efficiency of the VANETs by adopting a cluster-based routing scheme. The most important goal of any IoV model is to transmit the data packets within a network with maximum delivery rate and to achieve this task, IoT based communication is the best solution that can be observed by analysis of Table 2.1.

Table 2.1 represents an analysis of existing work based on the approaches used and their limitations, and the findings of existing research have described in Table 2.2 below with the work proposed by the authors.

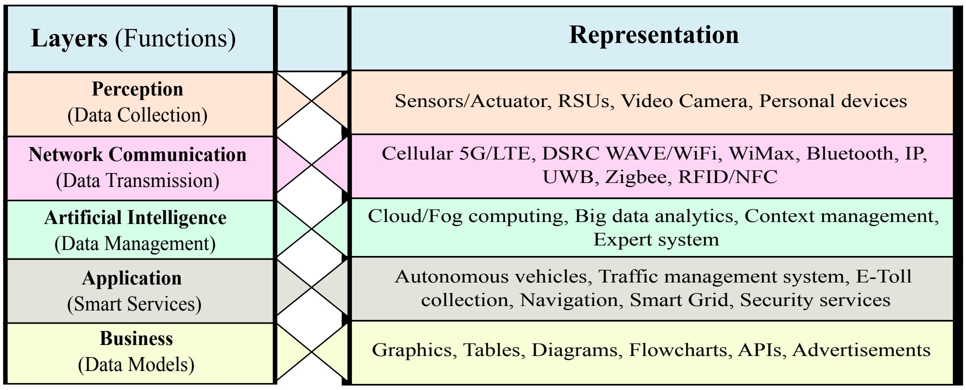
IoV depends on VANET, so it needs more analysis on VANETs based on its routing protocols, artificial intelligence, and optimization techniques. VANET is the resultant term originated from the general word “Ad-Hoc Networks”. VANET is a subcategory of Mobile Ad-Hoc Network (MANET), and VANETs also face many challenges that are common with MANETs [36]. Due to the development in vehicle industries, wireless communication and telecommunication results to the advancement in VANET. In this network, vehicles are taken as nodes that are movable and sharing the information between the nodes for creating a network. Vehicles that are under the field of another node, may communicate with that node plus turn out to be as a part of the network. There are mainly three kind of connection feasible in vehicular ad-hoc networks called Vehicle-vehicle (V2V), Vehicle-road side units (V2R) plus clustering based routing protocols. Several IT and vehicle industries are operating collectively for organizing VANET and offering protection to consumer, consistent data

Table 2.2
Findings from Related Work

Year	Author/s	Proposed Work	Findings
2019	Elgarej Mouhcine et al. [29]	A smart routing scheme in VANET using a distributed Ant Colony Optimization (ACO) was projected.	The projected scheme relies on the VANET architecture and is followed by a dispersed ant system (DAS) algorithm. DAS is an example of swarm intelligence, which provides good results for searching the shortest and finest routes. The results showed that the technology not only reduces traffic congestion by rerouting vehicles, but also helps reduce overall waiting time.
2019	Christy Jackson Joshua et al. [30]	A multi-objective firefly optimization approach (FOA) dependent weighted clustering routing protocol in VANET was designed.	The authors discussed the difficulties of clustering in a reliable weighted clustering routing protocol with the Firefly-optimization algorithm. The results differ from other related methods such as Comprehensive Learning Particle Self Adaptation (CLPSO) and Multi- objective PSO with a summary obtained from Chennai (Urban) in India. A similar form of the evolutionary algorithm used for Multi-objective with the help of multiple processors to increase performance with reduced computational time.
2019	Mrigali Gupta et al. [31]	Designed a Particle Swarm Angular Routing for Vehicular ad-hoc networks (PSARV).	The authors suggested an active particle swarm angular routing protocol for VANETs (PSARV) that applies the PSO technique as angular routing to search an appropriate route. The RREQ-swarms routing executes more proficiently than the DSR considering throughput and packet loss. The predicted process can be applied to a swarm leader who depends on swarm multicasting for the safety of passengers. Therefore, all the vehicles in this path are informed about hazardous situations onward, so that they are alert and aware in advance. The future goal is to estimate the performance of PSARV network scenarios in live traffic.
2019	Xi Hu et al. [32]	Proposed a model for VANETs using the concept of social co-operative driven vehicle-to-vehicle broadcasting optimization algorithm.	Proposed V2V broadcasting optimization algorithm based on social coalitions. Simulation results prove that the SCBO algorithm not only keeps the highest possible rate under varying node densities, but also increases the transmission saving rate to reduce the average end-to-end delay time.
2018	Jamal T. and Enrique A. [33]	Designed a vehicular network based on swarm intelligence algorithm for collaborative traffic.	Swarm FREDY advised a Swarm Intelligence inspired scheme for congestion control in VANET. FREDY has shown greater stability and more options for advancement than DIFRA. Since the beacon rates calculated through the first technique are often lower than those calculated by Swarm DIFRA.
2017	Ramesh C. Poonia [34]	Discussed the routing protocols for VANETs with swarm intelligence for performance evaluation of the network.	This article presented a VANET based on swarm intelligence and also examined the QoS of some routing protocols based on swarm intelligence like Time-ant, ACO-RA, Hy-BR, PRA, and Bee-Ad-hoc. They conclude Bee-Ad-hoc and Hy-BR is the most suitable routing algorithm in VANET but the network security is poor and will be improved by applying artificial intelligence techniques with swarm intelligence technology in the future.
2019	R. Yarinezhad and A. Sarabi [35]	Projected a new routing algorithm for VANETs driven Glowworm Swarm Optimization (GSO) algorithm.	Using GSO, the proposed algorithm detects optimal routes between 3-way congestion with intersections, and packets are delivered based on optimal routes, which are also fit for congested situations. Simulation results show that the designed algorithm is better than GSO and another two algorithms OSTD and SAMQ. Furthermore The GSO algorithm requires a large number of glowworms to build an optimal path and need artificial intelligence for better results.

interchange and providing the optimal path. One of the main tasks in VANET is to discover the optimal path or to find the target's node position. Varied researchers are trying their hand in this area for optimizing the route or searching the destination in VANET in a consistent manner. Numbers of optimization procedures are existing being encouraged from nature and also has own varied characteristics and calculates diverse method for solving problems and for optimization of the results.

After the survey, the next section discusses the layered architecture of IoV, which helps explore the characteristics and challenges of smart and secure communication for better understanding.



Layers (Functions)	**Representation**
Perception (Data Collection)	Sensors/Actuator, RSUs, Video Camera, Personal devices
Network Communication (Data Transmission)	Cellular 5G/LTE, DSRC WAVE/WiFi, WiMax, Bluetooth, IP, UWB, Zigbee, RFID/NFC
Artificial Intelligence (Data Management)	Cloud/Fog computing, Big data analytics, Context management, Expert system
Application (Smart Services)	Autonomous vehicles, Traffic management system, E-Toll collection, Navigation, Smart Grid, Security services
Business (Data Models)	Graphics, Tables, Diagrams, Flowcharts, APIs, Advertisements

Fig. 3.1. *IoV 5 layers architecture*

3. Layered Architecture and Protocols. Layered architecture design of complex networks such as IoV with various technologies is a challenging task that needs to identify and effectively group a set of elements with similar functionality and representation. Several researchers have proposed different layered IoV architectures.

3.1. Layered Architecture. A five-layered architecture is depicted in this section, which includes perception, network, artificial intelligence, application, and business layer. The functionality and representation of each layer have briefly described below, with the proposed Fig. 3.1 from studies on [37, 38] has the typical IoV architecture.

- **Perception Layer:** Perception layer includes each sensor inside the vehicle that collects natural data to make precise decisions for various driving patterns and traffic conditions use, such as satellite sightings, road traffic monitoring, vehicle position recognition, and vehicle monitoring.
- **Network Layer:** This layer establishes communication using heterogeneous networks like Cellular 5G/LTE, DSRC WAVE/Wi-Fi, WiMax, GSM, WLAN, Bluetooth, Radio frequency identification (RFID) signals and many other modes of wireless communication for Vehicle-everything (V2X) modes.
- **Artificial Intelligence Layer:** It is the heart of the IoV model which is used for computing, processing and storing of the information. This layer having sub-layers like sensing, data mining, analytics and intelligent control layer.
- **Application Layer:** This layer contains statistical tools, storage provision, and infrastructure operations that are responsible for analysis, processing, storage, and decision making for risk factors like traffic collisions, bad weather. Here smart real-time apps, traffic safety, efficiency, and multimedia data used.
- **Business Layer:** This layer contains operational and management logic related to the business aspect, mainly for the development of business models and statistical analysis of the vehicular data. For these tasks, various analysis tools use, such as flowcharts, graphs, tables, diagrams, and use cases.

3.2. Routing Protocols. As high flow in the vehicle network, the most difficult task is data routing or connection establishment process. There are five categories of routing protocols in the Internet of Vehicles (IoV) model, which are listed and discussed below with the help of diagram 3.2:

- **Position or Geographic Based (Unicast):** These routing protocols consist of a variety of algorithms for exchange details about the geographical positioning to pick the next forwarding nodes. The information is communicated to the one-node neighbour, which is nearest to the target, without any route information. So, this routing method is useful, as there is no requirement to build and maintain a universal route from source to target node like location-aware routing for Unmanned Aerial Vehicles in software-defined networks [41]. The rank based route is mainly divided into two classes: Location-based greedy Vehicle-vehicle networking and Delay-Tolerant Networking (DTN) Protocols.
- **Topology Based:** These protocols use the details links available on the channels for data packets delivery. These are mainly separated by Proactive, Reactive and Hybrid types. Proactive is table-

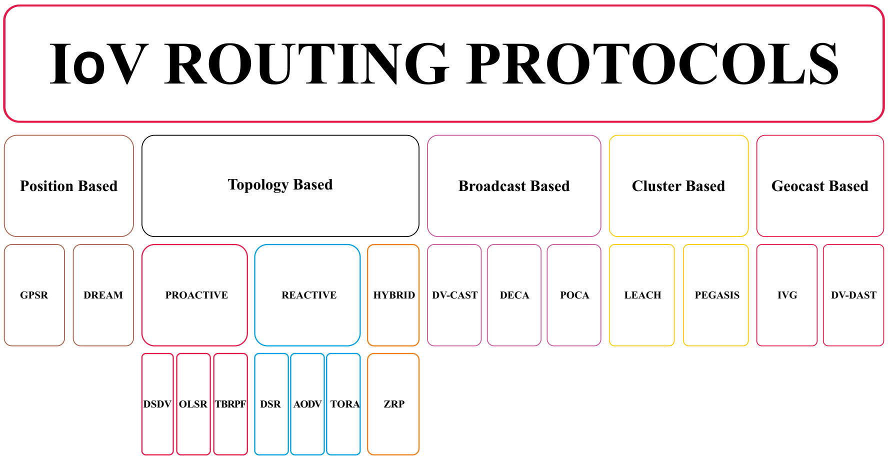


Fig. 3.2. *IoV Routing protocols taxonomy adapted from [39, 40]*

based routing, where every hop maintains a routing table. Reactive is on-demand routing, there is no need to maintain the routing table. Hybrid routing is a combination of both, where the network is divided into zones [42].

- **Broadcast Based:** Broadcast based routing is widely used in the vehicular network for data sharing regarding weather, emergency and traffic conditions between vehicles as well as delivery of advertisements and announcements.
- **Cluster Based (Multicast):** A collection of hops describes itself as part of a cluster from which a hop is selected as the cluster head (CH) to distribute the packets. Cluster head (CH) selected based on specific selection criteria. Scalability can be given in large networks, but delays will be found when grouping in most vehicle networks. To implement scalability, a virtual network structure needs to form by the clustering (grouping of hops).
- **Geocast Based (Multicast):** The Geocast route is a multicast-based geographic location [43]. Its purpose is to pass the packet from single-vehicle through multiple vehicles to reach the destination based on their local geographic region known as Zone. Unicast routing may be applied to deliver the packet inside the target zone. In Geocast, vehicles outside the simulation area are not warned to bypass the unwanted extra-fast response. One disadvantage of Geo Cast is network segmentation, as well as the next adverse nodes, which can interrupt the usual information delivery.

Nowadays, VANETs integrate with software-defined networking (SDN), which requires a more flexible content-centric networking approach as opposed to multicasting [44]. Whereas active assemblies of basic VANETs should be used for a unicast message communication at present; as there are still many applications that require seamless unicast routing.

3.3. Security requirements. The above survey concluded some basic security requirements to protect the vehicle network from attackers and maintain secure communications, listed in the following Table 3.1:
The layered architecture, routing protocols, and security requirements have discussed in this section, but the use of intelligence needs to be explored more with the help of a new proposed model. Therefore, to analyze the impact of artificial intelligence and machine learning, some important conclusions have proposed in the following section.

4. Analysis and Proposed Framework. In this part, the existing works have analyzed using QoS parameters like throughput and packet loss for comparing and observing values from the simulation section of [1, 2, 3, 4] in terms of throughput and packet loss. Throughput typically represents an average feasible or the transfer rate between two nodes. Throughput is usually inversely correlated with the possibility of packet loss, as packet loss leads to a slow throughput rate. Therefore, to compare and analyze existing works, calculated the average Throughput based on unit conversion (like Kbps to Mbps or Gbps to Mbps) and Packet Loss

Table 3.1
Security requirements in VANET based on[45]

Parameters	Description
Non-Repudiation	This avoids frauds from denying their offenses because in this case even if the attack happens, unrepentant will expedite the capability to detect attackers.
Availability	Vehicular networks will require real-time for many purposes and should therefore always be available. These apps require high speed from sensory networks or through the Ad-hoc network, the resultant elimination may be or the massage can be a little useless if there is a hold of seconds in various applications.
Confidentiality	This security requirement assured that information will be read by allowed users only. A need for confidentiality is required in group communications, where team members are not allowed to read such information.
Authentication	It assures that the data has been entered by an authentic user. The data accessing the physical stream must be precise and established by the authentic person because in IoV, the reaction nodes according to data established from the other end.
Integrity	It ensures that the data in the sender and the sender side are the same Message conversion is done by authorized users only. The recipient uses the same process as the one used on the sender's side to create the second call from the message comparing it to the first message. This process ensures the integrity of the data.

Table 4.1
QoS Parameters Comparison

S.no	Authors	Average Throughput (Mbps)	Packet Loss Rate (%)
1	**Muhammad Asim Saleem et al.** [1]	5.30	12.77
2	**Lucy Sumi and Virender Ranga** [2]	53.94	23.74
3	**Amr Tolba** [3]	95.17	9.83
4	**Zhenyu Zhou et al.** [4]	85.33	8.87

(Sometimes known as Total packet/Success transfer rate). In these research articles, simulation parameters have calculated with different scenarios, and comparisons between them are not possible depending on the scenarios considered. Analysis based on QoS parameters, such as throughput and packet loss, is shown in the Table 4.1 to compare existing work; their graphical representation is in Fig. 4.1.

Table 4.1 and Fig. 4.1 represent a comparative analysis of the IoV model in previous years proposed by various techniques and algorithms based on the QoS parameters. Fig. 4.1 (a) represents the comparative analysis of observed throughput in different scenario for IoV, but in case of work proposed by [3] and [4] is much better compare to the others because they use the concept of estimation of headway distance of vehicles as a Wiener process by exploiting Kolmogorov equation to maintain the connectivity of the vehicle. The transmission loss of existing work is represented in Fig. 4.1 (b) with comparative analysis in different scenarios for IoV and the transmission loss of work proposed by [3] and [4] is less due to stable connectivity between the vehicles. From the Fig. 4.1, some observations have shown that the models designed by different researchers have their advantages and disadvantages, but the use of artificial intelligence is a beneficial step for secure communication in IoV networks with existing works.
Additional studies are required to build a smart and reliable IoT based VANET model using artificial intelligence techniques.

4.1. Proposed Framework. Based on the above survey findings and gaps analysis, IoT-based VANET and intelligent services have used to design a proposed Cluster-based, Self-organized, and Intelligent (CSI) framework that helps in smart and secure communication thereby reducing traffic congestion, as shown in Fig. 4.2.

- **IoV Setup Phase:** Design and deploy IoT based VANET
  Step 1: IoT has helped develop the ad-hoc network of traditional vehicles into ad-hoc networks of intelligent vehicles.
- **Cluster Setup Phase:** Clusters formation and Data transmission
  Step 2, 3: Define the coverage area for each vehicle/RSU node, which helps to build the route from the source node to the destination node. To solve this problem the selection of Cluster Head (CH) node is

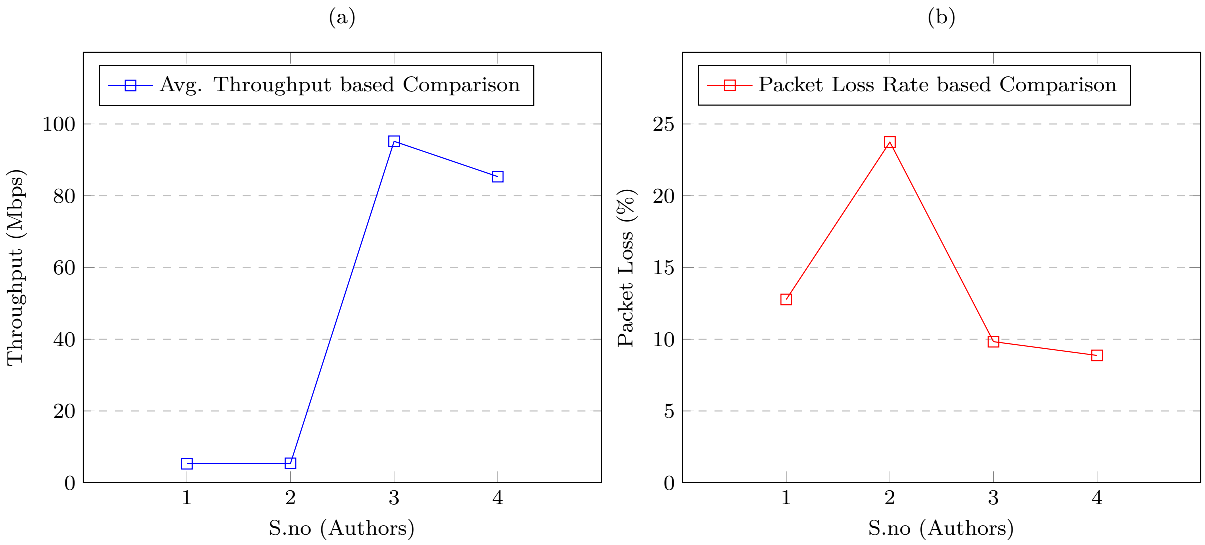


FIG. 4.1. *Comparison of QoS Parameters: (a) Average Throughput and (b) Packet Loss Rate*

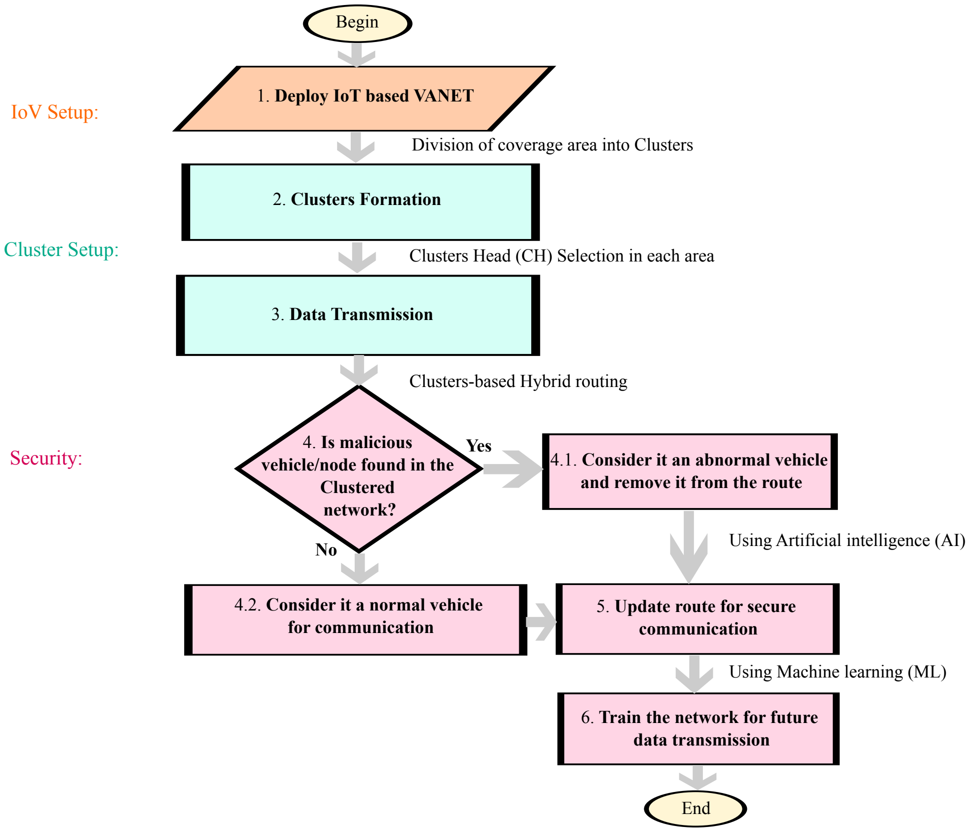

Begin
IoV Setup:
1. Deploy IoT based VANET
Division of coverage area into Clusters
2. Clusters Formation
Cluster Setup:
Clusters Head (CH) Selection in each area
3. Data Transmission
Clusters-based Hybrid routing
Security:
4. Is malicious vehicle/node found in the Clustered network?
Yes
4.1. Consider it an abnormal vehicle and remove it from the route
Using Artificial intelligence (AI)
No
4.2. Consider it a normal vehicle for communication
5. Update route for secure communication
Using Machine learning (ML)
6. Train the network for future data transmission
End

FIG. 4.2. *Proposed Cluster-based Self-organized Intelligent (CSI) Framework*

best suited for better data transmission.

- **Security Mechanism:** Malicious node, Update route, and Network training
  Step 4, 5, 6: When the performance of the network degraded, then the concept of Artificial intelligence (AI) is used to detect malicious or fail nodes in the network, which has not able to communicate with other vehicles/RSU thereby create traffic congestion. The machine learning (ML) algorithm acts as a processing unit that updates itself with changes in routes used for future data transmission.

Some important outcomes based on the proposed CSI model have listed below that will help solve real-life problems in existing IoV networks:

- The use of clustering-based routing approaches with trust management mechanisms will be beneficial for the future internet of vehicles.
- The integration of artificial intelligence with routing mechanisms helps to detect networks blocked by malicious or dead nodes.
- The concept of traffic congestion minimization will be introduced with the IoV network to provide a fast and robust communication.

Moreover, we have decided to use a robust cluster-based routing to validate the proposed work. Hence, future studies on the current topic are required to compare and verify the proposed and existing models based on parameters such as Throughput, Packet loss, Packet delivery ratio, and delay.

5. Conclusion and Future Scope. New communication technologies for vehicles grow mainly from the improvement of basic communication between Vehicle-vehicle and Vehicle-infrastructure, and Vehicle-network. IoV model is a reality at present which is acquired by interconnections of vehicles and traffic infrastructure including people. This paper presented a detailed overview of the IoV architecture along with its routing protocols, issues, and challenges, which helps to build secure communication. IoV network designing is still at an early stage of development, and requires many technical issues to be resolved before it is recognized globally and deployed in modern networks. With the fast growth of computing and wireless transmission techniques, the Internet of Vehicles network offers large business and research importance for security and fast communication. Therefore, in the future, the IoV network is a better option with the concept of Artificial intelligence and Machine learning technology acting as a classifier used to train IoV networks with a reliable routing mechanism based on hybridization and meta-heuristic optimization algorithms adopted for security and fast communication purposes.

REFERENCES

[1] M. A. Saleem, Z. Shijie, and A. Sharif, Data transmission using iot in vehicular ad-hoc networks in smart city congestion," *Mobile Networks and Applications*, vol. 24, no. 1, pp. 248–258, 2019.

[2] L. Sumi and V. Ranga, An iot-vanet-based traffic management system for emergency vehicles in a smart city," in *Recent Findings in Intelligent Computing Techniques.* Springer, 2018, pp. 23–31.

[3] A. Tolba, Content accessibility preference approach for improving service optimality in internet of vehicles," *Computer Networks*, vol. 152, pp. 78–86, 2019.

[4] Z. Zhou, C. Gao, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, Social big-data-based content dissemination in internet of vehicles," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 768–777, 2017.

[5] Z. Ning, X. Hu, Z. Chen, M. Zhou, B. Hu, J. Cheng, and M. S. Obaidat, A cooperative quality-aware service access system for social internet of vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2506–2517, 2017.

[6] J. Wan, J. Liu, Z. Shao, A. V. Vasilakos, M. Imran, and K. Zhou, Mobile crowd sensing for traffic prediction in internet of vehicles," *Sensors*, vol. 16, no. 1, p. 88, 2016.

[7] A. Azman, S. Yogarayan, S. L. W. Jian, S. F. A. Razak, K. J. Raman, M. F. A. Abdullah, S. Z. Ibrahim, A. H. M. Amin, and K. S. Muthu, Comprehensive study of wireless communication technologies for vehicular communication," in *2018 3rd International Conference on Computer and Communication Systems (ICCCS).* IEEE, 2018, pp. 314–317.

[8] M. A. Haque and M. D. Hossain, Technology survey of wireless communication for in-vehicle applications," in *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014).* IEEE, 2014, pp. 1–7.

[9] A. Nayyar, B.-L. Nguyen, and N. G. Nguyen, The internet of drone things (iodt): Future envision of smart drones," in *First International Conference on Sustainable Technologies for Computational Intelligence.* Springer, 2020, pp. 563–580.

[10] J. Chen, H. Zhou, N. Zhang, W. Xu, Q. Yu, L. Gui, and X. Shen, Service-oriented dynamic connection management for software-defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2826–2837, 2017.

[11] M. Sookhak, F. R. Yu, and H. Tang, Secure data sharing for vehicular ad-hoc networks using cloud computing," in *Ad Hoc Networks.* Springer, 2017, pp. 306–315.

[12] S. Sharma and A. Kaul, A survey on intrusion detection systems and honeypot based proactive security mechanisms in vanets and vanet cloud," *Vehicular Communications*, vol. 12, pp. 138–164, 2018.

[13] N. A. Khan, N. Jhanjhi, S. N. Brohi, R. S. A. Usmani, and A. Nayyar, Smart traffic monitoring system using unmanned aerial vehicles (uavs)," *Computer Communications*, 2020.

[14] M. Priyan and G. U. Devi, A survey on internet of vehicles: applications, technologies, challenges and opportunities," *International Journal of Advanced Intelligence Paradigms*, vol. 12, no. 1-2, pp. 98–119, 2019.

[15] S. Kumar, H. Sharma, G. Singh, Neetu, and H. Chugh, Internet of vehicles (iov): A 5g connected car," *Advances and Applications in Mathematical Sciences*, Dec. 2019.

[16] A. Kumar, K. Rajalakshmi, S. Jain, A. Nayyar, and M. Abouhawwash, A novel heuristic simulation-optimization method for critical infrastructure in smart transportation systems," *International Journal of Communication Systems*, p. e4397, 2020.

[17] K. J. Borah, J. Borah, and M. Kantipudi, Optimal control of cyber physical vehicle systems," *International Journal of Intelligent Systems Design and Computing*, vol. 1, no. 3-4, pp. 205–213, 2017.

[18] S. Chaba, R. Kumar, R. Pant, and M. Dave, Secure and efficient key delivery in vanet using cloud and fog computing," in *2017 International Conference on Computer, Communications and Electronics (Comptelix).* IEEE, 2017, pp. 27–31.

[19] W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, Internet of vehicles in big data era," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 19–35, 2017.

[20] X. Wang, Z. Ning, and L. Wang, Offloading in internet of vehicles: A fog-enabled real-time traffic management system," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4568–4578, 2018.

[21] J. Kang, R. Yu, X. Huang, M. Jonsson, H. Bogucka, S. Gjessing, and Y. Zhang, Location privacy attacks and defenses in cloud-enabled internet of vehicles," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 52–59, 2016.

[22] J. Kang, R. Yu, X. Huang, and Y. Zhang, Privacy-preserved pseudonym scheme for fog computing supported internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2627–2637, 2017.

[23] E.-K. Lee, M. Gerla, G. Pau, U. Lee, and J.-H. Lim, Internet of vehicles: From intelligent grid to autonomous cars and vehicular fogs," *International Journal of Distributed Sensor Networks*, vol. 12, no. 9, p. 1550147716665500, 2016.

[24] W. Zhang, Z. Zhang, and H.-C. Chao, Cooperative fog computing for dealing with big data in the internet of vehicles: Architecture and hierarchical resource management," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 60–67, 2017.

[25] A. Paul, A. Daniel, A. Ahmad, and S. Rho, Cooperative cognitive intelligence for internet of vehicles," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1249–1258, 2015.

[26] P. M. Kumar, G. Manogaran, R. Sundarasekar, N. Chilamkurti, R. Varatharajan *et al.*, Ant colony optimization algorithm with internet of vehicles for intelligent traffic control system," *Computer Networks*, vol. 144, pp. 154–162, 2018.

[27] Y. Agarwal, K. Jain, and O. Karabasoglu, Smart vehicle monitoring and assistance using cloud computing in vehicular ad hoc networks," *International Journal of Transportation Science and Technology*, vol. 7, no. 1, pp. 60–73, 2018.

[28] S. Debroy, P. Samanta, A. Bashir, and M. Chatterjee, Speed-iot: spectrum aware energy efficient routing for device-to-device iot communication," *Future Generation Computer Systems*, vol. 93, pp. 833–848, 2019.

[29] E. Mouhcine, K. Mansouri, and Y. Mohamed, Intelligent vehicle routing system using vanet strategy combined with a distributed ant colony optimization," in *International Conference on Advanced Information Technology, Services and Systems.* Springer, 2018, pp. 230–237.

[30] C. J. Joshua, R. Duraisamy, and V. Varadarajan, A reputation based weighted clustering protocol in vanet: a multi-objective firefly approach," *Mobile Networks and Applications*, vol. 24, no. 4, pp. 1199–1209, 2019.

[31] M. Gupta, N. Sabharwal, P. Singla, J. Singh, and J. J. Rodrigues, Psarv: Particle swarm angular routing in vehicular ad hoc networks," in *International Conference on Wireless Intelligent and Distributed Environment for Communication.* Springer, 2018, pp. 115–127.

[32] X. Hu, T. Wu, and Y. Wang, Social coalition-based v2v broadcasting optimization algorithm in vanets," in *International Conference on Swarm Intelligence.* Springer, 2019, pp. 318–325.

[33] J. Toutouh and E. Alba, A swarm algorithm for collaborative traffic in vehicular networks," *Vehicular Communications*, vol. 12, pp. 127–137, 2018.

[34] R. C. Poonia, A performance evaluation of routing protocols for vehicular ad hoc networks with swarm intelligence," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 4, pp. 830–835, 2018.

[35] R. Yarinezhad and A. Sarabi, A new routing algorithm for vehicular ad-hoc networks based on glowworm swarm optimization algorithm," *Journal of AI and Data Mining*, vol. 7, no. 1, pp. 69–76, 2019.

[36] S. Glass, I. Mahgoub, and M. Rathod, Leveraging manet-based cooperative cache discovery techniques in vanets: A survey and analysis," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2640–2661, 2017.

[37] L. Tuyisenge, M. Ayaida, S. Tohme, and L.-E. Afilal, Network architectures in internet of vehicles (iov): Review, protocols analysis, challenges and issues," in *International Conference on Internet of Vehicles.* Springer, 2018, pp. 3–13.

[38] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero Ibáñez, A seven-layered model architecture for internet of vehicles," *Journal of Information and Telecommunication*, vol. 1, no. 1, pp. 4–22, 2017.

[39] O. Senouci, Z. Aliouat, and S. Harous, A review of routing protocols in internet of vehicles and their challenges," *Sensor Review*, 2019.

[40] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibañez, Internet of vehicles: architecture, protocols, and security," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3701–3709, 2017.

[41] S. Vashist and S. Jain, Location-aware network of drones for consumer applications: Supporting efficient management between multiple drones," *IEEE Consumer Electronics Magazine*, vol. 8, no. 3, pp. 68–73, 2019.

[42] S. Kumar, D. Kaur, and M. Talwar, Comparative performance analysis of aodv, dsdv and zrp under various network density and traffic in manets," *Journal of Computer Technology & Applications*, vol. 7, no. 2, pp. 9–16, 2019.

[43] M.-S. Pan and S.-W. Yang, A lightweight and distributed geographic multicast routing protocol for iot applications," *Computer Networks*, vol. 112, pp. 95–107, 2017.

[44] A. Gulati, G. S. Aujla, R. Chaudhary, N. Kumar, and M. S. Obaidat, Deep learning-based content centric data dissemination scheme for internet of vehicles," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[45] M. S. Sheikh, J. Liang, and W. Wang, A survey of security services, attacks, and applications for vehicular ad hoc networks (vanets)," *Sensors*, vol. 19, no. 16, p. 3589, 2019.

Edited by: Anand Nayyar
Received: Apr 21, 2020
Accepted: Jul 1, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 441–450, DOI 10.12694:/scpe.v21i3.1742

A NEW APPROACH FOR NAVIGATION AND TRAFFIC SIGNS INDICATION USING MAP INTEGRATED AUGMENTED REALITY FOR SELF-DRIVING CARS

HARSHAL DEORE[*] AKSHAT AGRAWAL[*] [†] VIVEK JAGLAN[‡] POOJA NAGPAL[§] AND MAYANK MOHAN SHARMA[¶]

Abstract. Self-driving vehicles are one of the emerging technologies. This technology has potential to save lives and make lives comfortable. However, the technology used in self driving cars has to perform series of task for building perceptions. This has some certain prerequisites related to road infrastructure and is affected by daylight and weather conditions of the place. If these prerequisites are not satisfied then it could affect the performance of the vehicle and can be considered as compromise with safety of the users. This research work is focused on trying to find a new approach using which the underdeveloped countries will also be able to implement self driving cars in their county. The objective of this paper is to propose a new approach to supplement the technology used in the self-driving cars for perception. Using this approach the countries who don't satisfy the prerequisites would be eligible to implement them without compromising the safety.

The proposed approach uses the technology Augmented Reality to create and augment artificial objects of navigational signs and traffic signals based on vehicles location to reality. Later the augmented scene is fed into the conventional Deep learning object detection algorithm to detect the navigational artificial objected along with other real objects. This approach help navigate the vehicle even if the road infrastructure does not have very good sign indications and marking.

The approach was tested locally by creating a local navigational system and a smartphone based augmented reality app. The approach performed better than the conventional method as the objects were clearer in the frame which made it each for the object detection to detect them.

Key words: Self Driving Cars, Augment Reality,Machine Vision

AMS subject classifications. 68T40, 68T45

1. Introduction.

Every year huge amount of money is spent on research for finding cures and vaccines, but people often fail to recognize magnitude and cost of road traffic accidents. Each year 2.2% of deaths are caused by road traffic accidents this makes it 9th leading cause of death in the world. Nearly 12.5 lakh people each year which mean on an average 3287 people die daily on the road due to road traffic accidents. Addition to that every year 2-5 crores of people are left injured. In these deaths the people aging between 15-44 account for more than 50%. Surprisingly, among 15-29 years old it is leading cause of death, and among 5-14 years old it is second leading cause of death globally. The number of people below age 25 dying in road traffic accidents is around 4 lakhs, which makes it 1000 young people each day. More than 90% of these deaths occur in countries which are underdeveloped or have low income. These accidents don't only cause fatalities but harm the economy also, It has been found that these accidents costs $518 worldwide, on an average 1 to 2 percent of an individual countries GDP. This expense is more than the amount few underdeveloped countries receive for development assistance [1].

Innovation has improved technology to a great extent over past few decades. The technologies have disrupted various industries for betterment of the society. One such innovation in the field of Artificial Intelligence was Deep Learning. Deep learning has enabled us to bring human like or even better perception in machines. The technology is set to disrupt various industries one such industry is transport. Innovation of deep learning has led people to build self-driving cars. Self-driving cars concept is to replace need of person to drive them from one place to another. These vehicles have cameras mounted over them which continuously capture the

[*] Amity University Haryana (harshaldeore7@gmail.com).

[†] Amity University Haryana(akshatag20@gmail.com).

[‡] Graphic Era Hill University(jaglanvivek@gmail.com).

[§] Amity University Haryana(pbnagpal@ggn.amity.edu).

[¶] Zillow Inc., San Francisco, U.S.A.(mayank.mohan.sharma@gmail.com).

surrounding and is fed into the deep learning algorithm which detects if there is any object which could cause an accident. The vehicle's control system totally depends on the detection algorithm and responds based on its findings. Self-driving vehicles are currently need human driver presence just work as an assistant as it is not yet fully developed. These vehicles are improving day by day and will be there in each and every corner of the world in the coming years.

Considering the statistics of deaths and injuries caused by road accidents it is evident that self- driving vehicle (self driving vehicles) are the future of transport. Around 47 cities in the world are piloting self driving vehicles projects. A report prepared by KPMG shows Autonomous Vehicles Readiness Index (AVRI) for 25 countries. The AVRI is a tool to measure how well a country is prepared to implement self driving vehicles, this tool considers 25 different factors and all these factors are scored in a single digit. It is quite clear that to successfully deploy self driving vehicles in a country, the country must have good road infrastructure. The good infrastructure for self driving vehicles means the lane marking should be proper, the traffic signals and sign boards should be present at all the essential locations. The lack of infrastructure has been the main reason behind the lag in implementation of this life saving technology [2].

The self driving vehicles have multiple sensor and cameras mounted on and around it. These cameras capture the surrounding and those images are fed into Deep learning DL algorithms to detect and recognize various objects. The results get affected with quality of cameras, surrounding condition like sunlight falling at low angle or it is too bright causes difficulty in reading traffic lights. Also fog, smog, snowfall make it difficult to work. Deep learning algorithm depends on data, providing it with more data with different scenarios its results improve. Hence it is very important it gets implemented in more countries and different places. This will help in collecting more data and gradually lead to better and safer self driving vehicles [3].

Countries like India are in the most need of self driving vehicles. In India, there were 464,674 road accidents which caused 148,707 deaths. In the population of 1.31 billion there are 182.45 million vehicles. In 2015 the collision rate was 0.8 per 1000 vehicles and per 1000,000 people there were 11.35 deaths [4].

Poor road infrastructure and poor surrounding conditions are the two problems which create a question mark over the performance and safety of self driving vehicles. The approach proposed in this paper has aim to help these self-driving vehicles reach the countries where they are most needed. The approach uses augmented reality technique to supplement the conventional method.

Objectives:

- Developing an algorithm to combine deep learning and augmented reality technologies
- Creating test setup to test the proposed approach
- Comparing results of the proposed approach

The paper is spread in four sections. Section 2 discusses related work followed by methodology in section 3. Section 4 presents results and comparison and finally section 5 discusses conclusion of current study with future directions.

2. Related Works.

2.1. Augmented Reality. Virtual Reality VR is used to create a virtual world with the use of computer graphics which can experienced through wearable devices. Telepresence is also same kind of technology used to facilitate someone's problem solving skills from a remote place [5].

However, Augmented Reality AR is another technology which is often referred as Mixed Reality MR as how it works is combination of both the VR and Telepresence technology.

In AR, a layer of artificial objects created by us to the real-world perception is added, these artificial can be considered as extra information to make the real world more informative. This extra information or the layer of the artificial objects does not cover up any important or meaning detail from the real-world but enriches the experience. In simple words AR is VR with added real world. The technology is being used widely to help students learn topics practically which are otherwise difficult to image. AR make learning interesting and joyful as students get to see virtual objects in real-life environment [6].

AR was invented with a desire to make build airplanes easier. In 1990, Caudell and Mizell developed a system for workers to help them during assembly for guiding them they could wear the device on their heads. The device would overlay real world with graphical instructions [7].

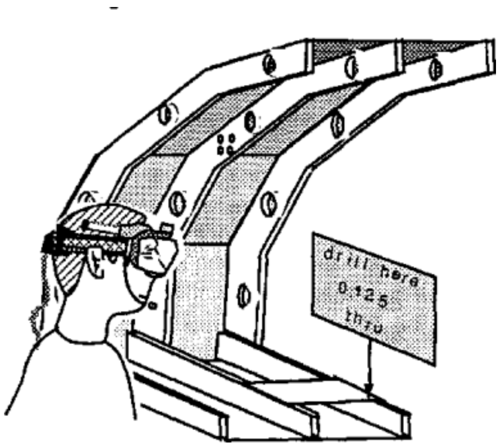


FIG. 2.1. *AR guiding worker for drilling [7]*

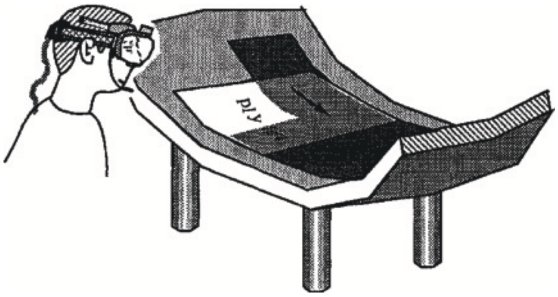


FIG. 2.2. *AR guiding worker with ply [7]*

However, the initial attempts were not successful as the system could not track head positions efficiently and was not responsive enough when people moved around. The main reason behind that was the computer inside the wearable was not powerful enough back then.

Furthermore, the idea was not limited to industrial applications. In 2000, Bruce Thomas along with other researchers developed an extended version of a desktop game called Quake. They converted that first-person desktop game to outdoor/indoor mobile AR game [8].

Similarly, it can be used for navigation purpose and people have been putting efforts in creating mobile based applications to help foreign people navigate at new place like Airport, Hotels or offices [9].

2.2. Deep Learning. Deep learning's birth came from people's aspiration to create machines which could act, think and behave like humans. In 1943, McCulloch and Pitts studied functioning of neurons inside the brain and tried to understand how they could produce complex patterns with a system of tiny cell connected to each other. They then created a small simplified model of neuron. This model had limitations but made an important contribution to development of artificial neural networks. Later they could add learning feature to the model [10].

Furthermore, in 1958 Frank Rosenblatt proposed the concept of perceptron this is considered as major development in neural networks. In his model he first passed the input through some pre-processors also called association units. Pre-processors work was to detect if there are any specific features present in the inputs [11].

However, in 1969 Minsky along with Seymour Papert attacked the work of Frank Rosenblatt where they introduced the XOR problem. They proved that a single perceptron a grandparent to the computational units which compose modern neural networks was incapable of learning the exclusive-or (aka XOR) function [12].

As a result, in 1986 David E. Rumelhart along with two other researchers introduced a new learning procedure for neural networks. The aim of the procedure was to minimize the difference between actual output

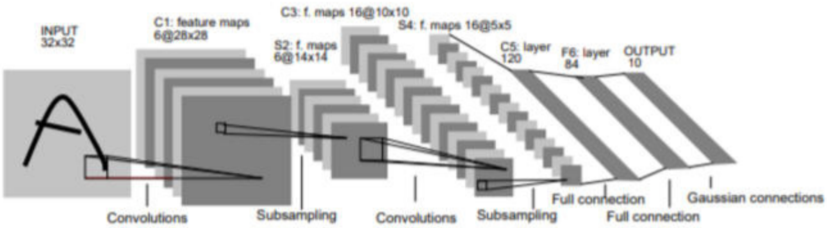


Fig. 2.3. *Convolutional Neural Network [14]*

and neural networks output, to achieve this it would repeatedly adjust weights of neurons. The procedure introduced was very different from the perceptron convergence method proposed earlier. They also introduced the concept of hidden units [13].

Subsequently, in 1998, Yann LeCun along with three other researchers introduced Convolutional Neural Networks for object recognition. They combined different operation like convolution max-pooling to extract features from images with Neural networks. Their work laid foundation to create a real time object recognition algorithm which is being used in self driving vehicles [14].

Current self-driving cars use Deep Learning technique to recognize objects. In this multi-layered neural network is trained on millions of images of different objects, the trained AI model is used to make prediction on the live camera feed being captured by the cameras mounted on the self driving vehicles. The algorithm draws bounding box around the object which helps in calculating the distance of the object from the self driving vehicles. Lane detection is also done on the live feed using image processing technique [15].

2.3. GPS. GPS's (Global Positioning System) birth came in 1960s during the cold war between the US and the Soviet Union. There was competition between these two countries on who could develop the best technology. American military understood importance of the technology when the Soviet Union the world's first satellite "Sputnik". Back then they used this technology to track submarines [16].

However, later in 1983 when the Korean airplane veered off into restricted Soviet airspace, they shot it down mistaking it for spy plane. To prevent such incidents, then President of the Soviet Reagen decided to release GPS technology.

Furthermore, After the release the first consumer GPS would cost 3000 USD. With time the technology kept improving day by day the initial device ran on short lasting batteries wasn't accurate enough. Other technologies also improved and supported GPS, with the developed of mobile phones GPS become a worldwide popular technology [17].

Today the technology has improved to a great extent and is very reliable due to this the same technology is being used in self driving vehicles for localization, tracking and route deciding [18].

3. Methodology. Current state of the art self driving vehicles include cameras, LiDAR, radar, GPS, wheel odometry, and IMUs [19]. An algorithm finds best possible route to reach the selected destination and the GPS constantly tracks current location of the vehicle. The vehicle's cameras records surrounding continuously, and the recorded video is processed in real time to find if any object is present or coming towards the vehicle. The cameras are also responsible for detecting lanes on the road this is how the vehicle the maintains lane discipline. However, night-time or different weather conditions increases risk of misjudging the situation which can result in accident. This can also lead to people loose faith in the technology which is improving day by day. Furthermore, some countries could not implement self driving vehicles because of their bad road infrastructure [20].

Our approach on other hand takes advantage of the deep learning technology to recognize objects and indication signs and overcome the disadvantages using GPS and AR technologies. The proposed approach uses GPS to get location and moving direction of the vehicle. From its location information virtual object is added to the original video. This new rendered video is fed into deep learning algorithm for detection various objects like traffic signal, lane lines, etc. based on this detection vehicle makes decision to which direction to steer [21].

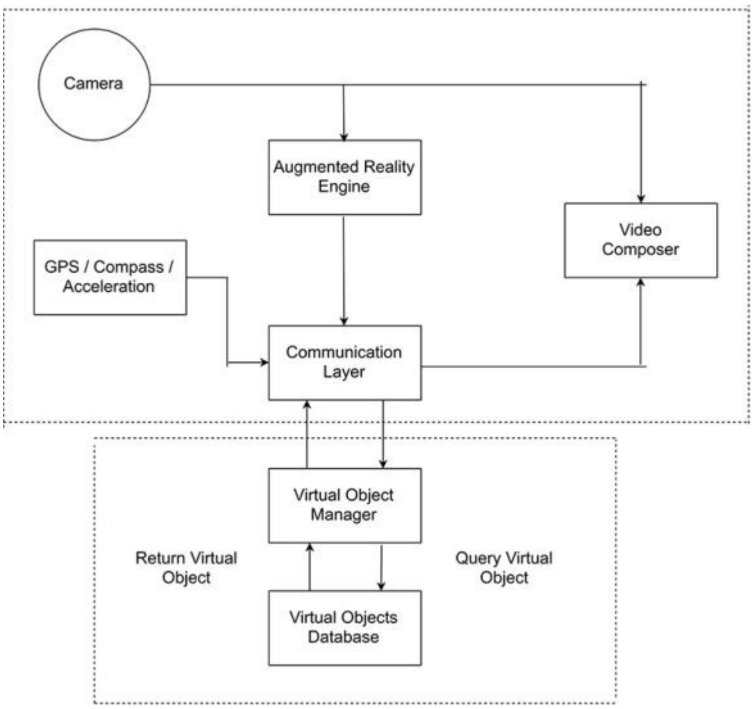


Fig. 3.1. *AR Server Platform Architecture*

Proposed ARIM algorithm process flow:

1. Getting current location of the vehicle using GPS.
2. Send the location of the vehicle to AR engine.
3. Based on location of the vehicle the virtual object manager makes request to Virtual object database.
4. Virtual object manager responds with the requested artificial object.
5. Video composer combines the graphical object and the original video captured by camera.
6. The new composed video is fed into the deep learning algorithm.

AR has two main components in it: Scene Generator and Tracking System. Scene generator is about showing the three-dimensional virtual object in the real-life environment. Rendering those virtual objects can be done to various levels depending on the requirement but simple graphics can also serve the purpose. Tracking system is about tracking the location of the user, the scene generator also depends on this as it has to modify the scene visuals depending on the perception of the user. Indoor navigation takes help of Bluetooth or local wi-fi devices to track current location of the person and based on that it shows indication on the phone to the destination.

The Virtual Object Manager and Virtual Object Database are two components of the system which work remotely. Communication layers manages the communication between the remote place and the actual place from where the request is made. GPS, compass and Accelerometer are one the most important components as the provide the information about location orientation of the device. Video composer composes the actual video with augmented virtual object [22].

The database of virtual objects is kept at a remote location and the Virtual Object Manager manages them [23]. The signs and lane lines are created in virtual navigational environment where will be able to see various signs while moving around which will guide us towards the destination.

The AR environment is created using FME AR software is explained in [24] the preparation process is described below:

1. First step is to take measurements of the place where the artificial objects are to be augmented.
2. Create an 3D model of the measurements on a 1:1 scale.

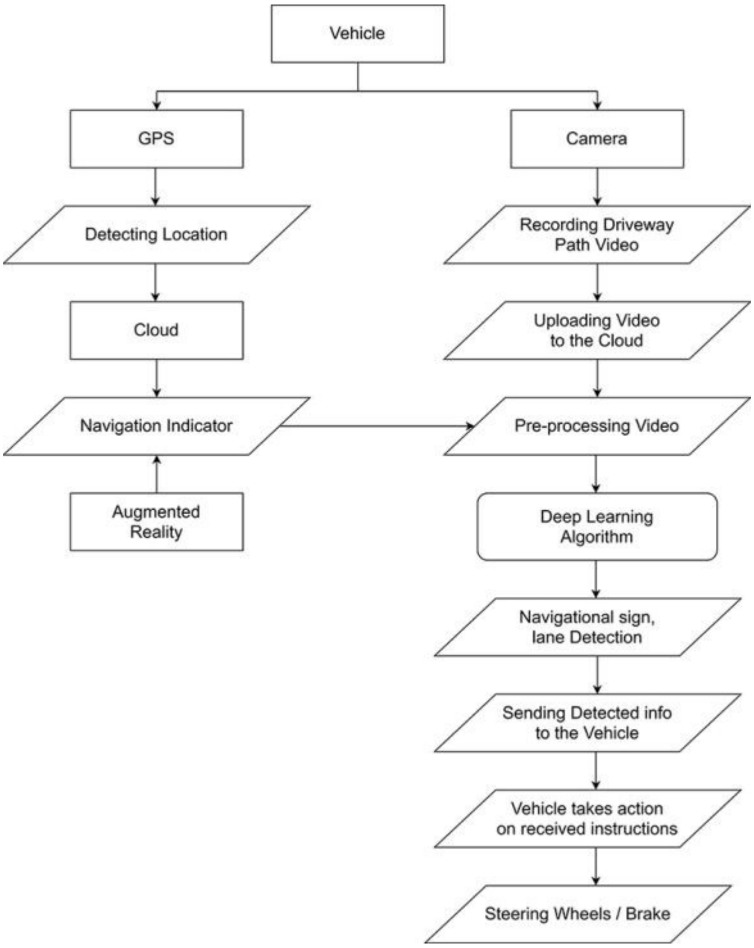

Vehicle
GPS
Camera
Detecting Location
Recording Driveway Path Video
Cloud
Uploading Video to the Cloud
Navigation Indicator
Pre-processing Video
Augmented Reality
Deep Learning Algorithm
Navigational sign, lane Detection
Sending Detected info to the Vehicle
Vehicle takes action on received instructions
Steering Wheels / Brake

Fig. 3.2. *ARIM Process Flow*

3. Draw a rectangle of the A4 sheet size somewhere in the corner. The sheet will act as an anchor.
4. Now convert the model to .fmear using the FME software. Bring the file to the phone and open it in the FME AR app.
5. Place a real sheet of A4 size paper on the exact same location where you drew the virtual sheet of paper in the model.
6. Match the two sheets of paper by scaling, offsetting and rotating the model on the screen.

Now the FME AR application installed in the phone shows the signs and screen is shared to the cloud source where it is analysed for detection [25].

The approach is tested on a small self driving vehicles model created using Arduino microprocessor [26]. The model has a smartphone fixed on it [27]. The smartphone is fixed in a way that its back camera faced the travelling path. The smartphone is connected to a laptop over wi-fi network and the laptop is connected to the Arduino over Bluetooth network [28].

The smartphone records video of the travelling path and sends it to the laptop over wi-fi network. Here the laptop plays the part of the cloud server [29]. Laptop processes the video and feeds it into the deep learning algorithm to detect lanes, signs and objects [30]. Based on the detection later on the algorithm calculates the steering angle required to keep vehicle model in the lane and follow indications [31]. The calculated angle is then sent to the Arduino over Bluetooth, Arduino then takes action based on the received information [32].

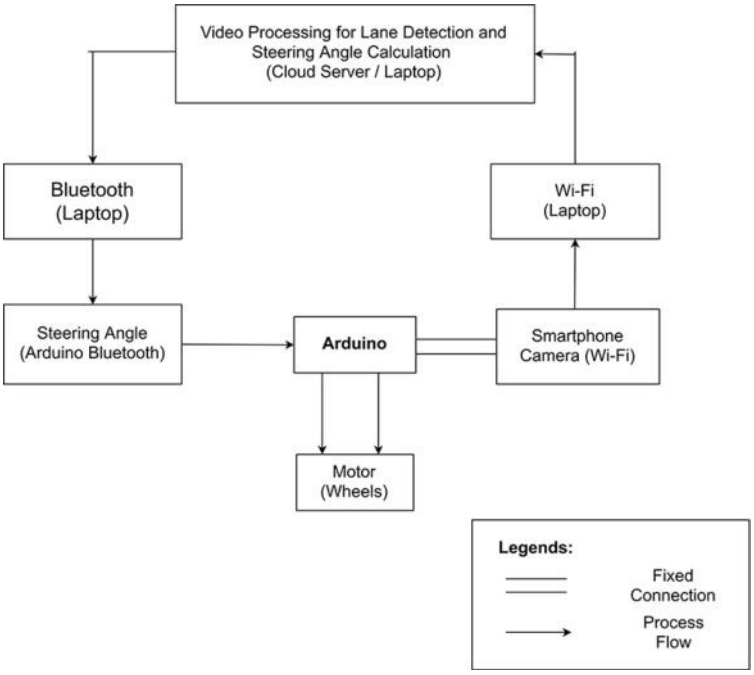

Video Processing for Lane Detection and Steering Angle Calculation (Cloud Server / Laptop)
Bluetooth (Laptop)
Wi-Fi (Laptop)
Steering Angle (Arduino Bluetooth)
Arduino
Smartphone Camera (Wi-Fi)
Motor (Wheels)
Legends:
Fixed Connection
Process Flow

Fig. 3.3. *Arduino Autonomous Vehicle Model Architecture*

4. Results and Discussion. The most important work of detecting a sign or object is performed by the deep learning algorithm [33]. To test the approach, a deep neural network with 7 layers (1 Input layer, 5 hidden layers and an output layer) was trained. The hidden layers consist of 3 pairs of convolutional and max pooling layers and 2 fully connected layer [34]. The output layer makes prediction using SoftMax function [35].

The dataset used for the training consists of 51839 colour images of size 32x32. The dataset was split into 3 parts train, test and validation with 34799 images in train, 4410 images in validation and 12630 images for testing. The images are of 43 different classes. The classes included different road traffic signs like stop, right turn, one way etc. After training, the model achieved 93% accuracy on the test data [36].

To test the efficiency of the approach [38], the trained model was later used for prediction on traditional method and the AR method and results are compared [37]. Comparing the average precision on precision vs recall plot showed improved results with the proposed AR approach. To compare results of traditional and

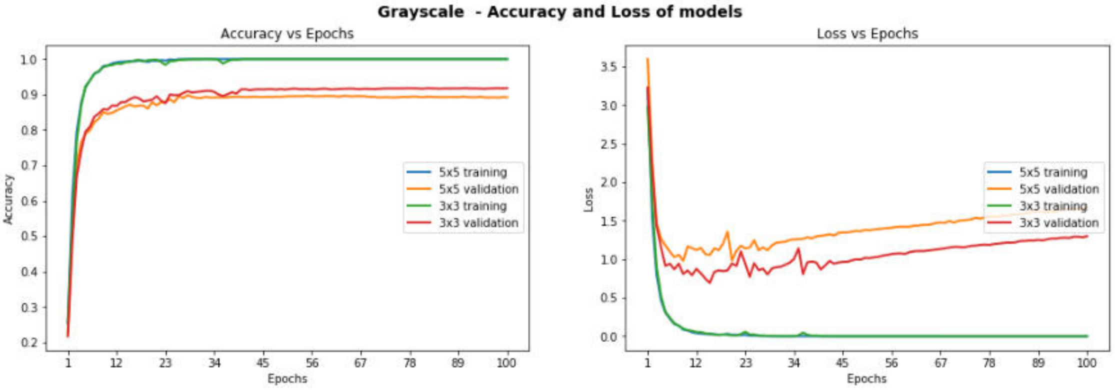


Fig. 4.1. *Model Training Performance*

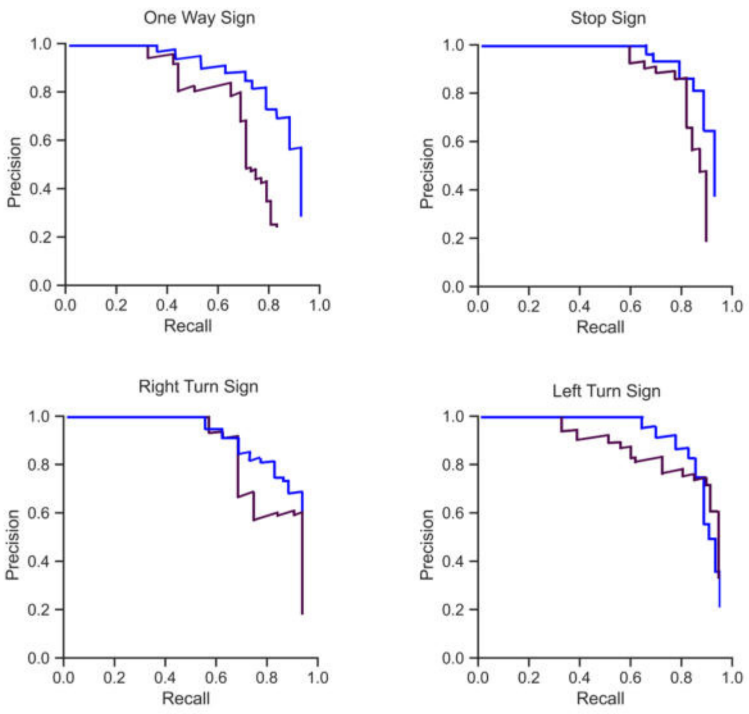


Fig. 4.2. *Average precision comparison between traditional approach and AR approach*

proposed approach, testing was performed on four different classes i.e. Stop Sign, One Way Sign, Right Turn Sign and Left Turn Sign [39].

Self driving vehicles don't just classify objects in a frame but also find their exact location. The algorithm draws bounding box around them for visualization. Performance of classification can be evaluated using confusing matrix but to calculate accuracy of an object detection algorithm mean average precision score is used. For better comprehension the scores are plot on Precision vs Recall graph. Below graphs show the performance of proposed approach comparing with traditional approach on precision vs recall plot [40].

In the graph the Magenta line indicates the traditional method and Blue line denotes proposed AR approach. The AR approach outperformed the traditional method with clear results. The ability of the traditional algorithm to detect a sign in a picture depends on how well it is visible in the picture. However, in the proposed approach signs are always present in a well accessible position making it easy to detect this was proved and validated by the results obtained.

5. Conclusion and Future Scope. The proposed algorithm is supposed to be used in self driving vehicles where safety of the passengers is of huge concern, hence it is important that the approach is properly tested. Testing the proposed algorithm on the real self driving vehicles was difficult, the test results presented in this paper was obtained by testing the approach in a local area map but the real self driving vehicles work on global map hence it is important to test the approach on a real self driving vehicle. However, it is expected that testing it locally with Bluetooth tracking was less accurate as compared to GPS hence the tracking will be better in the real self driving vehicles due to GPS and it can be integrated in a universal map like Google Maps.

The Arduino robot imitated role of self driving vehicles, the trained deep learning test model performed well on detecting the AR artificial navigational signs. As the objects were clear compared to real signs it was easier for the algorithm to detect. Using raspberry pi instead of Arduino could improve test results significantly and make the testing process seamless. Arduino's computational power is less compared to Raspberry pi, because of less computaional power the Arduino used in the testing has caused lag in responding to the instructions. In future work, Raspberry pi can be used to improve results [41].

The proposed work aims to lay a brick propose combining to technologies to promote the use of self-driving by improving its safety and reliability. However, there is huge future scope to improve the system. One problem that was encountered during the testing was the artificial objects created using AR were hiding important surrounding details. The detail in real self driving vehicles can be any human or moving object which should be detected to avoid an accident. This problem can be resolved in future improvements of the algorithm. The AR system used for testing was not up to the mark as there was latency in between Bluetooth and Wi-Fi networks. By tunnelling the AR research approach for use in the self-driving vehicle can produce much better results.

REFERENCES

[1] World Health Organization, *Global status report on road safety*, Inj. Prev., 2015.

[2] KPMG, *Autonomous Vehicles Readiness Index*, Auton. Veh. Readiness Index, 2018.

[3] A. Teichman and S. Thrun, *Practical object recognition in autonomous driving and beyond* Proceedings of IEEE Workshop on Advanced Robotics and its Social Impacts, ARSO, 2011.

[4] G. National Crime Records Bureau, *National Crime Records Bureau, Crime In India Reports from the Year 2002 to 2016* Crime In India Report, National Crime Records Bureau. 2016.

[5] M. Billinghurst, A. Clark, and G. Lee, *A survey of augmented reality* Foundations and Trends in Human-Computer Interaction. 2014.

[6] H. Kato, *Introduction to Augmented Reality* Kyokai Joho Imeji Zasshi, Journal Inst. Image Inf. Telev. Eng., 2012.

[7] T. P. Caudell and D. W. Mizell, *Augmented reality: an application of heads-up display technology to manual manufacturing processes* 2003.

[8] B. Thomas et al., *ARQuake: an outdoor/indoor augmented reality first person application* Int. Symp. Wearable Comput. Dig. Pap., 2000.

[9] W. Narzt et al., *Augmented reality navigation systems* Univers. Access Inf. Soc., 2006.

[10] W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity* Bull. Math. Biophys., 1943.

[11] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain* Psychol. Rev., 1958.

[12] M. Minsky and S. Papert, *Perceptron: an introduction to computational geometry* MIT Press. Cambridge, Expand. Ed., 1969.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors* Nature, 1986.

[14] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, *Object recognition with gradient-based learning* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1999.

[15] A. Azim and O. Aycard, *Detection, classification and tracking of moving objects in a 3D environment*IEEE Intelligent Vehicles Symposium, Proceedings, 2012.

[16] M. J. Rycroft, *Understanding GPS. Principles and applications* J. Atmos. Solar-Terrestrial Phys., 1997.

[17] H. Sugimoto, *Introduction to GPS* Seimitsu Kogaku Kaishi/Journal of the Japan Society for Precision Engineering. 2006.

[18] W. Rahiman and Z. Zainal, *An overview of development GPS navigation for autonomous car* Proceedings of the 2013 IEEE 8th Conference on Industrial Electronics and Applications, ICIEA 2013,

[19] A. Kumar, M. Mukherjee, and P. Mukhopadhyay, *Self Driving Car* Advances in Intelligent Systems and Computing, 2020.

[20] G. H. Lee, F. Faundorfer, and M. Pollefeys, *Motion estimation for self-driving cars with a generalized camera* in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2013.

[21] A. Y. C. Nee, S. K. Ong, G. Chryssolouris, and D. Mourtzis, *Augmented reality applications in design and manufacturing* CIRP Ann. - Manuf. Technol., 2012.

[22] O. Bimber and R. Raskar, *Spatial augmented reality: Merging real and virtual worlds* 2005.

[23] Q. Augmented and R. Lecture, *ORB-SLAM: a Real-Time Accurate Monocular SLAM System* IEEE Trans. Robot., 2015.

[24] Dmitri Bagh. (2020, March 01). Augmented Reality Indoor Mapping,

[25] Anand Nayyar, Bandana Mahapatra, D Le, G Suseendran, *Virtual Reality (VR) and Augmented Reality (AR) technologies for tourism and hospitality industry* International Journal of Engineering and Technology, 2018.

[26] M. Matijevic and V. Cvjetkovic, *Overview of architectures with Arduino boards as building blocks for data acquisition and control systems* in Proceedings of 2016 13th International Conference on Remote Engineering and Virtual Instrumentation, REV 2016.

[27] Vikram Puri, Anand Nayyar, *Real time smart home automation based on PIC microcontroller, Bluetooth and Android technology* 3rd International Conference on Computing for Sustainable Global Development, 2016.

[28] A. Cotta, N. T. Devidas, and V. K. N. Ekoskar, *Hc-05 Bluetooth Module Interfaced With Arduino*Int. J. Sci. Eng. Technol. Res., 2016.

[29] Anand Nayyar, Vikram Puri, *Smart farming: IoT based smart sensors agriculture stick for live temperature and moisture monitoring using Arduino, cloud computing and solar technology* The International Conference on Communication and Computing Systems, 2016.

[30] Vikram Puri, Anand Nayyar, Dac-Nhuong Le, *Handbook of Ardunio: Technical and Practice* Scholars Press, 2017.

[31] RS Batth, A Nayyar, A Nagpal, *Internet of robotic things: Driving intelligent robotics of future-concept, architecture, applications and technologies* 4th International Conference on Computing Sciences (ICCS), 2018.

[32] A. K. Jain, *Working model of Self-driving car using Convolutional Neural Network, Raspberry Pi and Arduino* Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, 2018.

[33] Faisal Saeed, Anand Paul, P Karthigaikumar, Anand Nayyar, *Convolutional neural network based early fire detection* Multimedia Tools and Applications, 2019.

[34] C. Y. Lee, S. Xie, P. W. Gallagher, Z. Zhang, and Z. Tu, *Deeply-supervised nets* Journal of Machine Learning Research, 2015.

[35] N. Ketkar, *Deep Learning with Python. 2017*

[36] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, *Detection of traffic signs in real-world images: The German traffic sign detection benchmark* Proceedings of the International Joint Conference on Neural Networks, 2013.

[37] Jafar Alzubi, Anand Nayyar, Akshi Kumar, *Machine learning from theory to algorithms: an overview* Journal of physics: conference series, 2018.

[38] Jonathan Hui, *mAP (mean Average Precision) for Object Detection* Jonathan Hui – Medium, Mar 7 2018.

[39] C. Caragea et al., *Mean Average Precision*in Encyclopedia of Database Systems, 2009.

[40] T. Zeugmann et al., *Precision and Recall* Encyclopedia of Machine Learning, 2011.

[41] Anand Nayyar, Vikram Puri, *Raspberry Pi A Small, Powerful, Cost Effective and Efficient Form Factor Computer: A Review* International Journal of Advanced Research in Computer Science and Software Engineering.

Edited by: Anand Nayyar
Received: Apr 24, 2020
Accepted: May 30, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 451–462, DOI 10.12694:/scpe.v21i3.1743

A FRAMEWORK TO SYSTEMATICALLY ANALYSE THE TRUSTWORTHINESS OF NODES FOR SECURING IOV INTERACTIONS

INDU BHARDWAJ [*], SIBARAM KHARA [†] AND PRIESTLY SHAN [‡]

Abstract. Trust plays essential role in any securing communications between Vehicles in IOV. This motivated us to design a trust model for IoV communication. In this paper, we initially review literature on IoV and Trust and present a hybrid trust model that separates the malicious and trusted nodes to secure the interaction of vehicle in IOV. Node segregation is done using value of statistics (S_t). If S_t of each node lies in the range of mean (m) plus/minus 2 standard deviation (SD) of PDR then nodes behaviour is considered as normal otherwise malicious. The simulation is conducted for different threshold values. Result depicts that PDR of trusted node is 0.63 that is much higher than the PDR of malicious node that is 0.15. Similarly, the Average no. of hops and trust dynamics of trusted nodes are higher than that of malicious node. So, on the basis of values of PDR, number of available hops and Trust dynamics, the malicious nodes can be clearly identified and discarded.

Key words: Internet of vehicles, Security, Trust Model, Challenges, Malicious Node

AMS subject classifications. 68M11

1. Introduction. The Internet of vehicle (IoV) [1] is a network in which vehicles, on-board sensors, roadside infrastructure and vehicular cloud are connected wirelessly to exchange traffic safely related information. It allows vehicles and infrastructure to be connected using internet connectivity [2]. In IoV, the concept of Internet of things (IoT) [3] is applied to the vehicles. It can be said that IoV is the joint version of VANET and IoT that enhances the road safety and security. The prime objective behind its implementation is to allow communication between different entities involved in it. In [4], IoV is defined as the dynamic mobile communication systems that enables communication between vehicles and public networks using V2V (vehicle-to-vehicle), V2R (vehicle-to-road), V2H (vehicle-to-human) and V2S (vehicle-to-sensor) interaction. Authors in [5], considers IoV as large-scale, distributed, wireless communication network for exchanging data between vehicle, road, human and internet, as per the agreed data interaction standards and communication protocols. To ensure the security of information exchanges in among the entities of IoV, the trust is established.

IoT is a swiftly developing system in which all entities of a network directly or indirectly connect to Internet. The evolution of IoT have revolutionized the vehicles to the great extent. For e.g. cars are equipped with navigation systems that gives information about traffic jams/ weather condition via internet and they also update the routing maps automatically. Some car models are equipped with vehicular communication network modules to communicate with other cars and alert the driver from invisible risk. They can find the car parking lots by themselves and react to any forthcoming accident.

Since IoV enables communication among various vehicles (which may also belong to malicious drivers/ pranksters), a serious interrogation arises on whether to trust vehicle or not. The idea of IoV network has brought various security, privacy and reliability issues that are imitated in a common term “trust” [6]. For example, IoV network is open and more vulnerable to attack by malicious users. Modelling trust is quite challenging in IoV network [7]. It is difficult to recognize which node is trusted and which is malicious.

Everybody recognizes what trust is, but nobody really knows how to define it to everyone’s satisfaction. Trust is a feature that exist in every communication but it is hard to formulate trust. After reviewing the

[*] Faculty of Electronics & Communication Engineering, Galgotias University, Greater Noida, India (indubhardwaj2011@gmail.com)

[†] Sharda University, Greater Noida, India (sianba@rediffmail.com)

[‡] Galgotias University, Greater Noida, India (priestlyshan@gmail.com).

current trust models for vehicular network, we present an effective trust model for IOV to separate the trusted and non-trusted nodes.

The main contributions of this paper are the followings:

1. We propose a fuzzy logic-based approach to evaluate the trust of one-hop neighbors. The proposed approach takes into account three different factors, namely, cooperativeness, honestness, and responsibility factors. Since the fuzzy logic-based approach is able to handle the complex and uncertain behavior of vehicles, it is suitable for dynamic and lossy vehicular networks.
2. We propose a Q-learning approach to evaluate indirect trust of nodes that are not directly connected to a trustor node. An evaluation about a non-neighbor-node is conducted by averaging the evaluation reports.
3. We propose a fuzzy logic-based approach to evaluate the trust of one-hop neighbors. The proposed approach takes into account three different factors, namely, coop- erativeness, honestness, and responsibility factors. Since the fuzzy logic-based approach is able to handle the complex and uncertain behavior of vehicles, it is suitable for dynamic and lossy vehicular networks.
4. A threshold-based trust approach is proposed to evaluate the trustworthiness of the nodes. This approach authenticates the nodes by comparing their trust values with a pre-set trust threshold. Since threshold-based approach is able to validate nodes without involving complex computation. So, secure node interaction can be established in timely manner that suits to the dynamic and decentralized nature of IoV network.
5. A trust initialization and storage mechanism is provided by the proposed model. to handle the cold start problem and scalability issues faced by existing models.
6. A joint probability-based approach is presented to update the trust at online centres. The trust is calculated by evaluating the trust worthiness of data using various statistics collected during interaction.
7. Computer simulations used to evaluate the effectiveness of the proposed trust model in separating malicious nodes from the trusted nodes and discarding them. By using this model, the malicious nodes will no longer be able to harm the network.

The rest of the paper is organized as follows. Section 2 includes the review and categorization of existing trust models on basis of their types, methodologies used and the network type. It also presents various challenges in modelling trust in IOV network. Section 3 includes the proposed trust model along with a descriptive discussion. Section 4 includes the simulation scenario and the results based on analytic study and simulation. At last, Section 5 concludes the paper and presents future scope.

2. Related Work.

2.1. Definition of Trust. Trust has different definitions in different contexts and subjects. But the importance of trust is same in every aspect. There is no particular definition of trust in vehicular network. Most of trust definitions are taken from the social sciences. However the trust has direct relation with the network security and basic concept of trust can be used to enhance the security of the network [8,9]. The trust is generally considered as a belief that one entity has about other entities depending upon the past experiences, data about the nature of entity, and on recommendations from other trusted entities. Authors in [10] stated that trust is a prime component in forming a trusted environment for VANET which endorses security in the network. In study [11], trust is an expectation and the belief about upcoming behaviour, depending upon past experiences. Authors in [12], defines trust as a relation among different entities established depending on the observations of past interactions.

2.2. Existing Trust Models. Research on development of trust model have been previously explored by various researchers in the field of MANET, VANET. But there are limited models which are proposed till now for IoV environment.

Authors in [13] presents a Multi-faceted approach to model the trustworthiness of data. This model is decentralized, task specific, scalable but it has not addressed the robustness. Gomez et al. model called TRIP [14] to differentiate malicious nodes from trusted nodes. It is a scalable model but it did not consider overhead introduced. Fangyu Gai [15] presented Ratee-Based Trust Management scheme model where each node maintains its own reputation rated by other during previous interactions. In extension to the work in [15],

Table 2.1
Types of trust models and their references

Types of Model	**Study**
Entity Based Model	[13], [14], [15], [16]
Data-based models	[17], [18], [19], [20] , [22], [23]
Hybrid Models	[10], [12], [21], [24],

Table 2.2
Various methodologies used in trust models and their references

Methodology	**Study**
Weighting	[10], [12], [21]
Ratings	[15], [16] , [19]
Probability	[18]
Bayesian network	[17], [22], [23]
Fuzzy logic	[14], [20]
Observations/Opinion gathering	[13], [24]

authors proposed a trust model for Social IoV [16]. This model is also ratee based This model also includes the Certification authority server and public key cryptography to avoid any alteration in the trust information by the ratee.

Study [17] includes a data based trust model for VANET to evaluate the trustworthiness of messages related to road safety. It used data trust instead of entity trust and utilizes Bayesian Inference approach in voting algorithm to enhance the robustness of network. Work in [18], presented a scheme to compute the reputations based on Hidden Markov Model (HMM). The proposed scheme evaluates the message reliability and predicts the legitimacy for broadcast messages. Authors in [19] proposed an announcement scheme for VANET based on reputations. Reputation value is evaluated by using a aggregation algorithm that is based on binary feedback ratings.

Study [20] includes an experience -based fuzzy trust model for securing the vehicular network. The proposed model executes various security checks to confirm the accuracy of received information. Yao et al. [21] proposed hybrid model including entity-centric trust evaluation based on weight and data-centric trust evaluation on the basis of experiences and the utility theory. L. Cong [22] et al. proposed data-based trust model to evaluate the correctness of vehicle to vehicle incident reports. This model computes the trust score by using behavior history of the incident report accuracy for a vehicle. Shu Yang et. al [23] proposed a trust model to elect anomaly nodes in IOV environment by forming cluster heads. Authors also provided mutual supervision model to handle tempering behaviors. Chen & Wei provided RSU and beacon-based trust management model[24] that prevents sending of false messages. Author in [12] proposed a Beacon-based trust management (BTM) model which computes entity trust from beacon messages. Merrihan Badr Monir et al. [10] combined experience and Role based trust to give Categorized trust based message reporting scheme for VANET.

According to the literature review, trust models are divided in three categories:

1. Entity trust model - evaluates the trustworthiness of the entity.
2. Data trust model - calculates the trustworthiness of data sent by entity.
3. Hybrid trust models – performs trustworthiness of data as well as entity.

Table 2.1 Summarizes trust models existing in each category. These trust models used different methodologies to model trust in network. Table 2.2 summarises various methodologies used in existing trust models. The methodologies used in for modelling trust are weights, ratings, probability, Bayesian network, fuzzy logic, and opinion gathering. From table 2.2, it can be clearly seen that Weights, ratings and Bayesian network are commonly used for trust modelling whereas probability approach is least used methods. Only one out of 14 trust models studied in literature used probabilistic approach in VANET. So, in our work we will focus on probabilistic approach to model evaluate the trust value. Table 2.3 shows the types of network for which the existing trust models are proposed. Out of 14 trust models studied in the literature, most of the trust models are designed for VANET and few are proposed for IoV environment. In our work we will focus on modelling trust for IoV network.

Table 2.3
Types of network and their references

Network	Study
VANET	[10], [12], [13], [14], [17], [18], [19], [20], [21], [22], [24]
IoV	[15], [16], [23]

2.3. Challenges in Trust management and Properties of Trust Model. Managing trust in IoV environment is quite important so that as to prevent malicious node to spread traffic-related false or tempered information. False information circulated by malicious nodes may create traffic jams and collision on roads. Dissemination of false information sometimes may result in dire consequences like loss of life. It is very challenging in IoV environment to manage trust in IoV network is various characteristics.

Trust verification in real-time: Vehicles randomly enter and leave IoV environment and move at very high speed so it is challenging to build up the trust in timely fashion. As vehicles interact for small time, it is difficult to judge which node is untrusted and up-to which extent.

Dynamicity: vehicular nodes are continuously moving so it is not necessary that the behaviour of trusted node will remain same always. Besides that, conditions of road are highly unpredictable [25]. Trust model developed for IoV should be able to handle these varying situations and characteristics of Network.

Large scale network: Number of vehicles in IoV are very large. Also, this situation become worse in the peak rush hours. This situation may arise problem like network congestion as vehicles are interacting through shared channel, and data overload – as vehicles may receive lot of data at one time from other vehicles stuck in a congested area.

Decentralization: There is no centralized infrastructure in IoV environment. Nodes can come and leave the network at any time. If a node interacts with a vehicle now, it is not guaranteed to interact with the same vehicle in the future.

According to the above characteristics of IoV and challenges in modelling trust, trust model should have following characteristics:

Fast computation: In order to evaluate trustworthiness of entity and data in real time for making quick decisions in IoV, trust model should have less complex so that trust computation can be fast. low complexity with also result in low computation overhead.

Distributed trust computation: Computation of trust in distributed manner is more suited for IoV due to its open, dynamic and self-organizing characteristics. When every node will calculate trust, there will be no need of central server to calculate the trustworthiness of nodes. Moreover, the system will have less chances of complete failure.

Scalable: Since the traffic is unpredictable so, the trust model should be scalable enough to handle the large number of nodes avoiding network congestion.

Literature review concludes that most of the existing trust model has been proposed for VANET. Trust modelling in IoV is still in infant stage. In our literature review, only three models out 14 models reviewed are proposed for IOV out of which two are entity trust model [15, 16] and one is data trust model [23]. There is no hybrid trust model proposed so far for IoV. Besides that, the existing IoV trust models suffers two main problem a) Scalability - when the number of nodes in network increases then it becomes difficult for each node to maintains the trust values of all. b) Cold start problem: this problem arises when a new joined node wants to communicate other nodes. Other nodes do not find the trust value to authenticate new node. In addition to this, we have identified some character that a trust model should have to overcome the above-mentioned challenges in IoV. To address these issues, we have proposed a framework for probability distribution-based hybrid trust model for IoV that initially computes the trustworthiness of nodes and then that of data exchanged between them by calculating trust. The proposed model is designed to solve the scalability issues and cold start problem and achieve the desired characteristics of trust model.

3. Proposed Model.

The proposed trust model for IoV is event driven. Trust values are stored online at trusted centres and updated after every event. System works as distributed protocol in which nodes computes the trust value of other node with which it interacted after every interaction. In this model, each node will be assigned a initial trust value of 0.5 whenever it joins the network. This will solve the cold start problem of

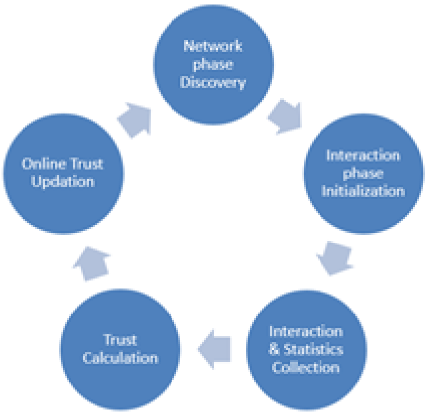

Network phase Discovery
Interaction phase Initialization
Interaction & Statistics Collection
Trust Calculation
Online Trust Updation

Fig. 3.1. *Trust Modelling Process*

the existing trust models. The nodes store the trust value of only limited set of nodes with whom it interacted recently. Besides this all other trust values will be stored at the online trusted centre. This will make the network more scalable in the sense that when the number of nodes increase during peak hours then nodes need not maintain the trust value of all other nodes as the trust values will be maintained at online centres. This will solve the problems associated with limited storage at node. This model Suits well to be decentralized architecture of IOV as there is no centralized authority for trust computation. Each entity in network has the capability to compute the trust value itself after interaction and update it online.

3.1. Trust Modelling. IOV has several advantages like internet connection, fast computations etc. over VANET that make it more useful for securing vehicular communication. Nonetheless, IOV is openly accessible and has huge data set involved in computation. Moreover, IOV is quite dynamic network where vehicles are joining and leaving the network continuously to cope up with these properties of IOV, we propose a trust model to secure communication in IOV. Trust modelling process used in proposed model is shown in Figure 3.1.

In an IOV trust model sender vehicle has to locate another vehicle with whom it wants to interact to get or provide the service as per the requirement of situation. If there are multiple requests then receiver node initialize the network phase discovery with the node having good reputation (high trust value). Once the network phase is over the interaction between node takes place. During this interaction, statistics is collected by RSU which is further used to calculate the new trust value of sender and receiver node. The trust is finally updated at online centers.

3.2. The Proposed Trust Model. The trust model proposed to secure communication in IOV is depicted as flow chart in Figure 3.2. The Proposed model is event driven. The process starts when A tries to interact with any node B to provide any service or get service from it. Node A initiates RSU to find location of Node B. To achieve this, RSU initially looks up to the past trust value of Node A saved at online centre to judge to trustworthiness of A. If trust value of A is available, RSU checks whether A's trust is greater than past threshold (T_0). If A fulfils the condition for minimum level of trust threshold, it is considered as legitimate node. If the A's trust value is less than T_0 then A can't interact with B.

Once RSU finds A as trusted Node, it initiates the procedure to find location of B. After locating B, RSU repeats the same procedure to judge trustworthiness of Node B. If B also meets the minimum trust requirement set for a node to be a legitimate node, the interaction between A & B starts. If B's Trust value is less than T_0 then node A cannot interact with node B. During interaction between both trusted nodes A & B, RSU collects the trust statistics like Packet Delivery Ratio (PDR). PDR is given by the following equation

$$PDR_t = \frac{\text{Total packet received}}{\text{Total packet transmitted}} \tag{3.1}$$

If the value of statistics (S_t) of each node lies in the range of mean (m) plus/minus 2 standard deviation

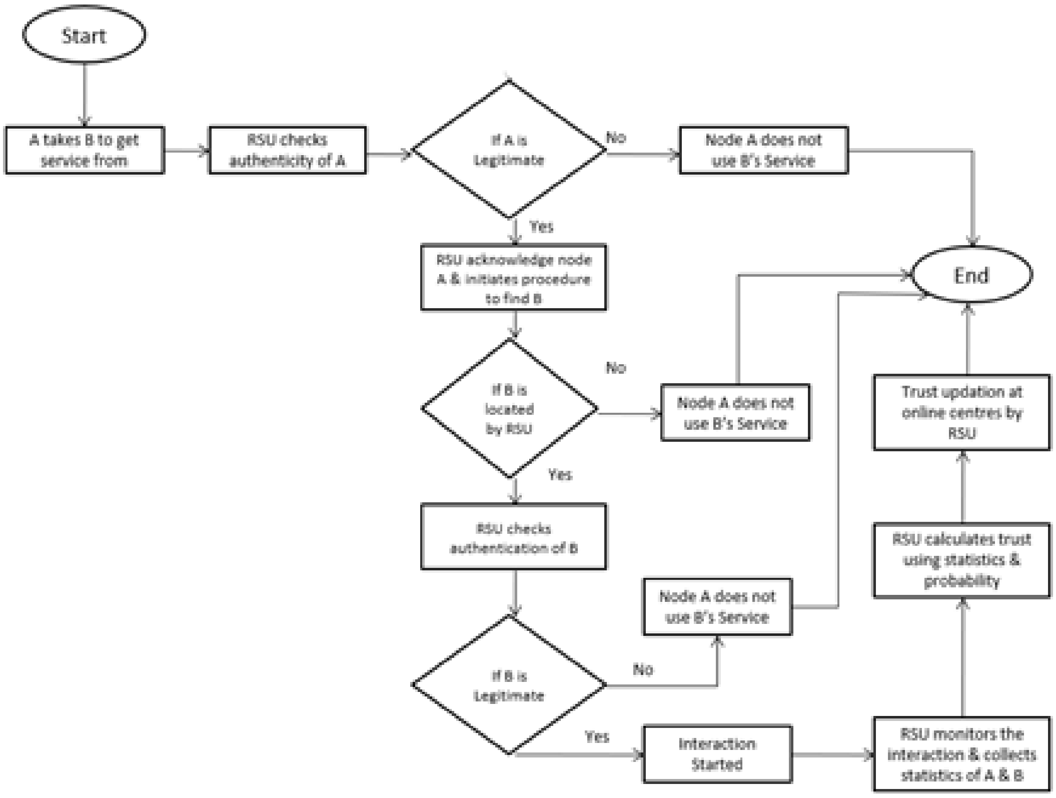

Start
A takes B to get service from
RSU checks authenticity of A
If A is Legitimate
No
Node A does not use B's Service
Yes
RSU acknowledge node A & initiates procedure to find B
End
No
If B is located by RSU
Node A does not use B's Service
Trust updation at online centres by RSU
Yes
RSU checks authentication of B
RSU calculates trust using statistics & probability
Node A does not use B's Service
No
If B is Legitimate
Yes
Interaction Started
RSU monitors the interaction & collects statistics of A & B

Fig. 3.2. *Flow Diagram of the Proposed Trust model for IoV*

(SD) of PDR then nodes behaviour is considered as normal otherwise malicious. Once the interaction between A & B is over, RSU calculates the new trust value for both node A & B by using conditional probability. RSU then updates the new calculated trust value at online centres.

$$\textit{if}\left\{\begin{array}{c} S_t > \ m+2SD \\ S_t < \ m-2SD \end{array}\right\} \quad \textit{Malicious Behaviour} \tag{3.2}$$

and

$$\textit{if}\left\{\begin{array}{c} S_t \leq \ m+2SD \\ S_t \geq \ m-2SD \end{array}\right\} \quad \textit{Trusted Behaviour} \tag{3.3}$$

Each time the successful / failed communication will take place between nodes the trust will be recalculated using conditional probability and updated (increased in case of successful interaction & decreased in case of failed interaction). It is to be noted that the above algorithm is inspired by interaction of humans in real life. We want to take services from trusted service provider and after taking service we update the feedback (trust value in this algorithm).

4. Results and Discussion.

4.1. Simulation Scenario. This section presents the simulation scenario our proposed trust model for IOV. The main aim of conducting this simulation is to study how efficiently the proposed trust model works in presence of non-trusted nodes in IoV environment. To achieve this the Simulation is conducted on SUMO (1.4.0 version) and MATLAB (2016a version). SUMO is used as traffic simulator for generating the traffic patterns and MATLAB is used as event simulator. Figure 4.1 shows the traffic scenario of real-world map for

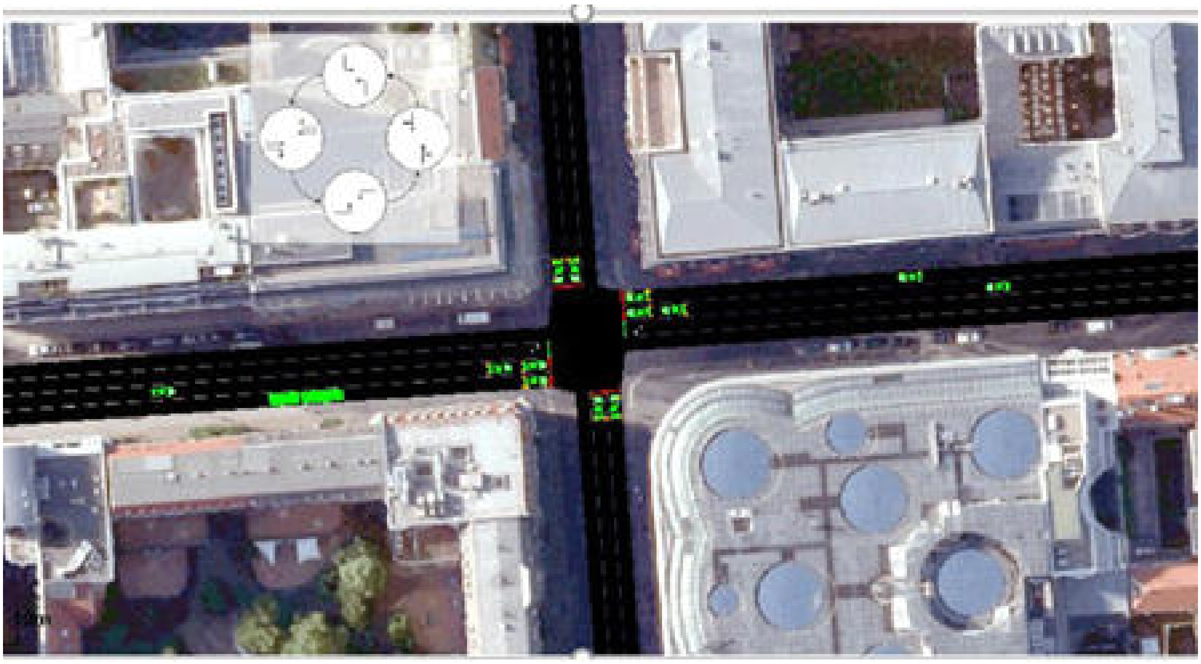


Fig. 4.1. *Traffic Scenario - Open Street Map for a Manhattan City*

Manhattan city generated using SUMO. The figure 4.1 includes top-view of traffic scenario near intersection of two roads having buildings. The objects in green color are vehicles moving on roads.

In experimental setup, the status of each node is changing dynamically. The input parameters provided to the simulation are listed in Table 4.1. Simulation is conducted by randomly setting some of the nodes as abnormal nodes. Initially, all nodes are assigned with equal trust value of 0.5. But with passage of time trust value of normal nodes increases with every successful interaction and that of abnormal node will decrease due to their malicious behavior.

To simulate this model, an IoV environment consisting of 30-100 nodes is considered in which 10% are malicious nodes. These nodes randomly move in 1000*1000 meters square area and has range of 250m for communication. The total time of simulation is taken 180 mins (3 hrs). The performance of proposed trust model is evaluated using three metrics i.e. number of available hops, PDR and trust value as metrics.

Simulation is conducted for three different value of threshold to study the impact of threshold value on evaluation metrics like PDR, average number of available hops. This study of different threshold will show how the value of evaluation metrics (PDR, number of available hops) vary for trusted and untrusted node under normal threshold policy (θ= 0.65), slightly strict (θ= 0.70) and highly strict threshold policy ($\theta = 0.75$).

4.2. Analytical Evaluation.

Fast computation: Instead of cryptography, the proposed model makes use of trust values for validating the trustworthiness nodes. This reduces the computation complexity and overhead involved in key management. This makes the computations fast.

Distributed trust computation: The proposed model does not involve any central authority to calculate the trustworthiness of nodes. Every entity in the network is connected to internet and able to calculate and update the trust value of nodes after every interaction. The distributed trust computation reduces the chances of complete system failure and is more suited for IoV due to its open, dynamic and self-organizing characteristics.

Scalable: Any node in the network need not maintain the trust of all the nodes in the network rather only for small set of nodes with which node plans to have interaction. So, the proposed trust model is scalable enough to handle the large number of nodes avoiding network congestion. It also solves cold start

Table 4.1
Parameters for simulation

Simulation Parameters	**Values**
Monitoring Area	1000 × 1000 meters
Number of nodes (n)	30-100
Communication Range	250 meters
Packet Interval	2 ms
Length of Data Packet	923 bits
Symbol rate	256KB/S
Bit rate	512KB/S
Simulation time	180 (s)
No Malicious Nodes	10%
Routing Protocol	A-STAR
Mac Layer Protocol	802.11p
Trust Range	[0,1]
Initial trust value of each node	0.5

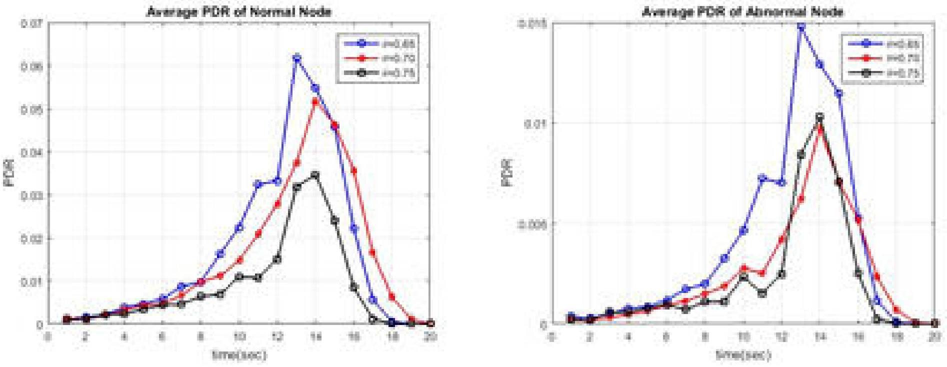


Fig. 4.2. *Average PDR for normal node (left) and abnormal (right) nodes*

problem by assigning minimum trust to each node initially.

4.3. Simulation-based Evaluation. The simulation shows easily how the proposed model can filter trusted and malicious nodes se depending upon the Packet delivery ratio, number of available hops curves. The network simulation is conducted for three different values of threshold values.

4.3.1. Packet delivery ratio. It is the ratio of number of packets received successful to the total number of packets sent. As we know that malicious nodes will not forward all the received packet so estimation of its PDR will be less as compared to that of trusted node. The curves show how the packet delivery ratio of normal and malicious nodes varies with time. The results show that average PDR value for trusted as well as malicious nodes, varies continuously with increase in time. Ideally the value of PDR should be as high as possible for better performance of network. The graphs presented in figure 4.2 depicts that average PDR of normal nodes is high approx. i.e. 0.063 in threshold $T_0 = 0.65$ as compared with the average PDR of abnormal nodes i.e. 0.015 at threshold $T_0 = 0.65$.

4.3.2. Effect of threshold policies on PDR. Table 4.2 shows the values of Average PDR of trusted nodes for different thresholds (θ = 0.75,0.70,0.65) at different instants of time starting from t= 0 sec to 20 sec. Initially at t=0, the Average PDR of trusted nodes is zero for each value of θ. As the times increases from t=0 to t=14 seconds, the average PDR value of trusted nodes is increasing for each threshold and After t=14, Average PDR values are decreasing for each threshold value till t=20. It means the maximum PDR achieved at t=14 sec for all the thresholds. For θ=0.65 is 96% which is very high as compare to the maximum PDR achieved for θ=0.75 i.e. 64%.

The PDR reading at almost every instant of time is less for higher threshold values, for e.g. at t=10, PDR

Table 4.2
PDR value of trusted nodes at different threshold

θ \ t	0	2	4	6	8	10	12	14	16	18	20	Max
0.75	0	0.31	0.32	0.35	0.37	0.41	0.45	0.64	0.38	0.30	0.30	0.64
0.70	0	0.30	0.33	0.36	0.40	0.44	0.58	0.82	0.52	0.30	0.31	0.82
0.65	0	0.31	0.35	0.38	0.63	0.84	0.80	0.96	0.81	0.38	0.33	0.96

Table 4.3
PDR value of non-trusted nodes at different threshold

θ\t	0	2	4	6	8	10	12	14	16	18	20	Max
0.75	0	0.00	0.00	0.00	0.00	0.00	0.0025	0.0100	0.0025	0	0	0.0100
0.70	0	0.00	0.00	0.00	0.00	0.00	0.0030	0.0103	0.0052	0.0001	0	0.0103
0.65	0	0.00	0.00	0.00	0.00	0.00	0.0042	0.0135	0.0053	0.0007	0	0.0135

for θ=0.65 is 0.84, which decreases for θ=0.70 i.e. 0.44 and further decreases for θ=0.75 i.e. 0.41. Similarly, at t=20 seconds, PDR for θ=0.65 is 0.33, which decreases for θ=0.70 i.e. 0.31 and further decreases for θ=0.75 i.e. 0.30. This discussion on PDR values concludes that the PDR for trusted nodes decreases significantly (i.e. from 96% to 64%) with increase in trust threshold (i.e. from 0.65 to o.75). This is due to the reason that if threshold policy is strict then sometimes trusted nodes may be considered as untrusted.

Table 4.3 shows the values of PDR of non-trusted nodes for different thresholds (θ = 0.75,0.70,0.65) at different instants of time starting from t= 0 sec to 20 sec. Initially at t=0, the Average PDR of non- trusted nodes is zero for each value of θ. It remains zero from t=0 to t=10 seconds irrespective of threshold value. After t=10 seconds, the average PDR value of non-trusted nodes increases for each threshold till t=14 seconds and After t=14 seconds, Average PDR values are decreasing for each threshold value till t=20 seconds. It means that the maximum PDR achieved at t=14 seconds for θ=0.65 is 1.35% which is comparable to the maximum PDR achieved for θ=0.75 i.e 1.0%. For each non zero values of PDR (from t=12 to t=18), it is observed that, the PDR reading at every instant of time is less for higher threshold values. For e.g. at t=12, PDR for θ=0.65 is 0.42, which decreases for θ=0.70 i.e. 0.003 and further decreases for θ=0.75. This discussion on PDR values concludes that with increase in trust threshold, the PDR for non- trusted nodes decreases but not significantly. Reason behind this is that malicious nodes has nothing much to do with threshold policies as their main motive is to affect PDR.

4.3.3. Average number of available hops to the non-trusted node. We estimated the average number of available hops to the trusted nodes with the progression of time for different trust threshold. Fig 4.3 depicts that as the time progresses, the available number of hops to the trusted nodes increases because of their good behaviour. More number of hops helps them in getting shortest path. So, behaving good is rewarding.

It is evident from the figure 5 that as the time progresses the number of average hops to the non-trusted nodes approaches to zero. Within first 10 seconds the average number of hops drops significantly. This drop is more prevalent in stricter threshold policy (θ=0.75).

4.3.4. Effect of threshold policies on Available number of hops. Table 4.4 shows the values of number of hops available of trusted nodes for different thresholds (θ = 0.75,0.70,0.65) at different instants of time starting from t= 0 sec to 20 sec. Initially at t=0, the no. of available hops for trusted nodes is zero for each value of θ. As the times increases from t=0 to t=20 seconds, the available no. of hops for trusted nodes increases continuously for each threshold value due to their good behaviour in the interactions. Hops readings at every instant for different values of threshold shows that available number of hops is comparatively less for higher threshold values. for e.g. at t=10, hops available for θ=0.65 is 27, which decreases for θ=0.70 i.e. 18 and further decreases for θ=0.75 i.e. 15. Moreover, Average no. of hops available for θ=0.65 is 26 which is very high as compare to the that for θ=0.75 i.e. 17. This discussion concludes that the growth of average available hops is higher when θ is less (θ=0.65) and smallest during the strict policy. This is due to the reason that under strict threshold policies the trusted node sometimes may be misunderstood as non-trusted.

Table 4.5 shows the values of number of hops available of non-trusted nodes for different thresholds (θ =

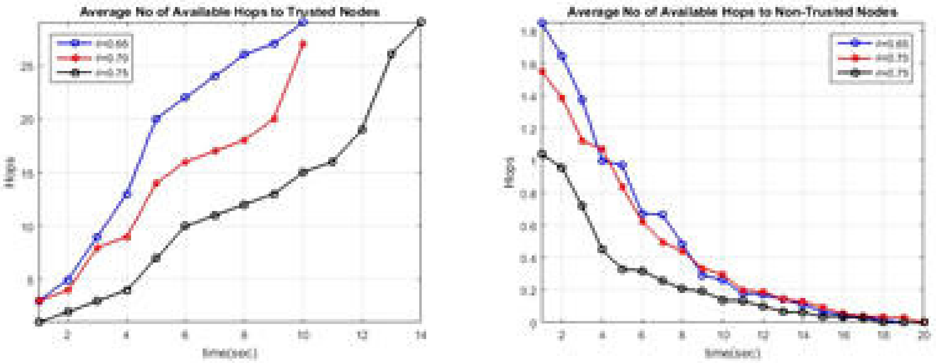


Fig. 4.3. *Average number of Available hops to trusted (left) and non-trusted (right) nodes*

Table 4.4
Number of hops available for trusted nodes at different threshold

θ\Time	0	2	4	6	8	10	12	14	16	18	20	Avg.
0.75	0	2	4	10	12	15	19	29	28	32	36	17
0.70	0	2	4	9	16	18	27	30	33	38	42	20
0.65	0	3	9	20	24	27	30	33	41	45	50	26

0.75,0.70,0.65) at different instants of time starting from t= 0 sec to 20 sec. Initially the malicious node shows some available no. of hops to mislead the other nodes. But, as the times increases from t=0 to t=20 seconds, the available no. of hops for non-trusted nodes decreases continuously for each threshold value and becomes Zero at t=20. This is due to their misbehaviour in the interactions. Hops reading at each instant of time for different values of threshold shows that available number of hops is comparatively less for higher threshold values. for e.g. at t=14, hops available for θ=0.65 is 0.11, which decreases for θ=0.70 i.e. 0.13 and further decreases for θ=0.75 i.e. 0.06. The growth of average available hops is higher i.e. 0.438 when θ is less (θ=0.65) and smallest according to the equation during the strict policy θ=0.75. This discussion concludes that the growth of average available hops is higher when θ is less (θ=0.65) and smallest during the strict policy but the value is almost negligible in both cases.

4.3.5. Trust Dynamics. Trust dynamic of a node shows the trust worthiness of nodes. The trust dynamics changes dynamically after completion of each interaction. The trust value varies between 0 to 1. Figure 4.3.5 shows the combined graph of trust dynamics for trusted as well as abnormal nodes. The result depicts that the trust values of trusted nodes are continually increasing with time and that of abnormal node is gracefully decreasing with passage of time.

Every successful interaction contributes further increase in the trust value of trusted node. The reduction in trust value of malicious node is due to its misbehaviour. Initially there is not much difference in the trust dynamics of normal and abnormal node but as the time increases and more events are encountered the difference increases to great extent that helps in clearly separating the abnormal nodes from normal nodes and discarding them.

5. Conclusion and Future Scope. Modelling trust in IoV network is quite challenging. This paper presents various challenges faced by researchers in modelling trust for IOV network and the characteristics of the trust model. Additionally, we have proposed probability-based hybrid trust model that is combination of entity based and data-based trust models. The entity-based trustworthiness is evaluated by using pre-set threshold policy and data- based trustworthiness is evaluated by collecting the statistics during communication. The model used joint probability distribution to separate the malicious and trust nodes. If the measure statistics lies in the range of mean plus/minus twice of standard deviation then it is considered as trusted otherwise untrusted. The model also resolves the cold start problem by providing initial trust value to each node. The analytic evaluation of proposed model shows the model is scalable and involves fast and distributed trust

Table 4.5
Number of hops available for non-trusted nodes at different threshold

θ\Time	0	2	4	6	8	10	12	14	16	18	20	Avg.
0.75	1.06	0.95	0.45	0.32	0.21	0.14	0.10	0.06	0.03	0.00	0.00	0.226
0.70	1.55	1.39	1.07	0.62	0.44	0.29	0.19	0.13	0.06	0.04	0.00	0.423
0.65	1.8	1.64	1.00	0.67	0.49	0.23	0.17	0.11	0.05	0.02	0.00	0.438

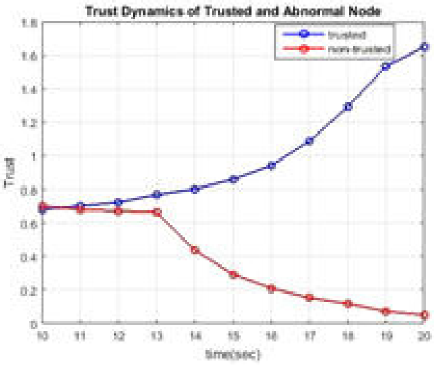


Fig. 4.4. *Trust dynamics trusted and non-trusted nodes wrt time*

computations, so is well suited for IoV. The experimental results for proposed model depict that PDR of trusted node is 0.63 that is much higher than the PDR of malicious node that is 0.15. Additionally, the average number of available hops, and trust value of trusted nodes are also significantly higher than that of non-trusted node. Thus, the malicious nodes can be clearly identified and discarded on the basic of value of PDR, available hops and Trust dynamics. The effects of threshold on evaluation metrics shows PDR and available number of nodes for both trusted and non-trusted nodes decrease with increase in threshold (θ). But this decrease is less significant in non-trusted nodes.

In future, the work might be extended to investigate the following aspects:

1. The current model secures the traffic information exchanged between vehicles. This model might be extended to secure the data transactions in other application scenarios of IoV network.
2. In proposed model, a vehicle and its driver are considered as a single node. Our model might be extended to identify the malicious behaviours of drivers and vehicles separately and discard it.
3. The proposed system might be extended by using better techniques to improve the robustness of the model.
4. In this paper, we present a separate approach to evaluate the trustworthiness of entity and data. A single approach might be used to compute the trustworthiness of both data as well as entity to make the computation much faster than that in this model.

REFERENCES

[1] O. Kaiwartya et al., Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects, IEEE Access, vol. 4, pp. 5356–5373, 2016.

[2] L. Angeles and L. Angeles, Internet of Vehicles?: From Intelligent Grid to Autonomous Cars and Vehicular Clouds, IEEE World Forum Internet Things, pp. 241–246, 2014.

[3] A. El Bekkali, M. Boulmalf, M. Essaaidi, and G. Mezzour, Securing the Internet of Things (IoT), Proc. - 2018 Int. Conf. Wirel. Networks Mob. Commun. WINCOM 2018, vol. 44, pp. 51–58, 2019.

[4] L. M. Ang, K. P. Seng, G. K. Ijemaru, and A. M. Zungeru, Deployment of IoV for Smart Cities: Applications, Architecture, and Challenges, IEEE Access, vol. 7, pp. 6473–6492, 2019.

[5] R. Silva and R. Iqbal, Ethical Implications of Social Internet of Vehicles Systems, IEEE Internet Things J., vol. 6, no. 1, pp. 517–531, 2019.

[6] K. Zaidi and M. Rajarajan, Vehicular internet: Security & privacy challenges and opportunities, Futur. Internet, vol. 7, no. 3, pp. 257–275, 2015.

[7] R. Iqbal, T. A. Butt, M. Afzaal, and K. Salah, Trust management in social Internet of vehicles?: Factors , challenges , blockchain , and fog solutions, Int. J. Distrib. Sens. Networks, vol. 15, no. 1, 2019.

[8] D. D. K. Nigahat, A review of blackhole attack in mobile adhoc network, Int. J. Eng. Sci. Res. Technol., vol. 6, no. 4, pp. 314–319, 2017.

[9] S. Mandala, K. Jenni, M. A. Ngadi, M. Kamat, and Y. Coulibaly, Quantifying the severity of blackhole attack in wireless mobile Adhoc networks, Commun. Comput. Inf. Sci., vol. 467, pp. 57–67, 2014.

[10] M. Monir, A. Abdel-Hamid, and M. A. El Aziz, A Categorized Trust-Based Message Reporting Scheme for VANETs, Commun. Comput. Inf. Sci., vol. 381 CCIS, no. 5, pp. 65–83, 2013.

[11] N. Bismeyer, S. Mauthofer, K. M. Bayarou, and F. Kargl, Assessment of node trustworthiness in VANETs using data plausibility checks with particle filters, IEEE Veh. Netw. Conf. VNC, pp. 78–85, 2012.

[12] Y. M. Chen and Y. C. Wei, A beacon-based trust management system for enhancing user centric location privacy in VANETs, J. Commun. Networks, vol. 15, no. 2, pp. 153–163, 2013.

[13] J. Finnson, J. Zhang, T. Tran, U. M. Farooq, and R. Cohen, A framework for modeling trustworthiness of users in mobile vehicular ad-hoc networks and its validation through simulated traffic flow, in springer book series Lecture Notes on Computer Science, 2012.

[14] F. Gómez Mármol and G. Martínez Pérez, TRIP, a trust and reputation infrastructure-based proposal for vehicular ad hoc networks, J. Netw. Comput. Appl., vol. 35, no. 3, pp. 934–941, 2012.

[15] F. Gai, J. Zhang, Z. Peidong, and X. Jiang, Ratee-Based Trust Management System for Internet of Vehicles, in Part of the Lecture Notes in Computer Science Springer book series, springer, 2017.

[16] F. Gai, J. Zhang, P. Zhu, and X. Jiang, Trust on the Ratee?: A Trust Management System for Social Internet of Vehicles, Wirel. Commun. Mob. Comput., vol. 2017, 2017.

[17] G. Wang and Y. Wu, BIBRM: A Bayesian Inference Based Road Message Trust Model in Vehicular Ad Hoc Networks, in IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, 2014.

[18] A. Shrivastava, K. Sharma, and B. K. Chaurasia, HMM for Reaputation Computation in VANET, in IEEE International Conference on Computing, Communication and Automation (ICCCA2016), 2016, pp. 667–670.

[19] Q. Li, A. Malip, K. M. Martin, S. Ng, and J. Zhang, A Reputation-Based Announcement Scheme for VANETs, in IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, 2012, vol. 61, no. 9, pp. 4095–4108.

[20] S. A. Soleymani et al., A Secure Trust Model based on Fuzzy Logic in Vehicular Ad Hoc Networks with Fog Computing, in IEEE., 2017, vol. 3536, no. c, pp. 1–10.

[21] X. Yao, X. Zhang, H. Ning, and P. Li, Using trust model to ensure reliable data acquisition in VANETs, Ad Hoc Networks, vol. 55, pp. 107–118, 2016.

[22] C. Liao, J. Chang, I. Lee, and K. K. Venkatasubramanian, A trust model for vehicular network-based incident reports, in IEEE 5th International Symposium on Wireless Vehicular Communications, WiVeC 2013, 2013.

[23] S. Yang, Z. Liu, J. Li, S. Wang, and F. Yang, Anomaly detection for internet of vehicles: A trust management scheme with affinity propagation, Mob. Inf. Syst., vol. 2016, 2016.

[24] Y.-M. C. Yu-Chih Wei, Reliability and Efficiency Improvement for Trust Management Model in VANETs, in Human Centric Technology and Service in Smart Space, springer, 2012, pp. 105–112.

[25] J. Zhang, Trust Management for VANETs: Challenges, Desired Properties and Future Directions, Int. J. Distrib. Syst. Technol., pp. 48–62, 2012.

Edited by: Anand Nayyar
Received: Apr 26, 2020
Accepted: May 13, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 463–477, DOI 10.12694:/scpe.v21i3.1746

A PARALLELIZATION BASED DATA MANAGEMENT FRAMEWORK FOR PERVASIVE IOT APPLICATIONS

SANIYA ZAHOOR, ROOHIE NAAZ MIR*

Abstract. Pervasive Internet of Things (IoT) is a research paradigm that has attracted considerable attention nowadays. The main aim of pervasive IoT is that in the future, the everyday objects (devices) would be accessible, sensed, and interconnected inside the global structure of the Internet. But in most of the pervasive IoT applications, the resources of an IoT device such as storage, processing, and energy are limited; as such there is a need for management of resources in such applications. Multiple aspects related to the data such as the type of data, size of data, number of transmission and reception of data packets, the structure of data, etc are taken into consideration while managing the resources of pervasive IoT applications. Therefore data management is essential for the management of limited resources in such applications. This paper presents the recent studies and related information in data management for pervasive IoT applications having limited resources. This paper also proposes a parallelization based data management framework for resource-constrained pervasive applications of IoT. The comparison of the proposed framework is done with the sequential approach through simulations and empirical data analysis. The results show an improvement in energy, processing, and storage requirements for the processing of data on the IoT device in the proposed framework as compared to the sequential approach.

Key words: Internet of Things, Resource-constraints, Security, Tiny encryption, Data aggregation, Parallelization

AMS subject classifications. 68M11

1. Introduction.

Pervasive Internet of Things is the evolution of the Internet designed to sense, collect, analyze, and distribute the data via smart, programmable, light-weight, miniaturized IoT devices (nodes). The main goal of pervasive IoT is to form the network of day-to-day life objects and make them programmable using wireless and sensor technologies [1]. The pervasiveness of IoT eases everyday activities such as data exchange by devices while sensing and reacting to events. There are numerous applications of pervasive IoT that come from different sectors such as healthcare, agriculture, homes, offices, waste management, transport, weather monitoring, water supply, etc. But in most of these applications, resources such as storage, processing, and energy are constrained [4]. These resource constraints impact the software design at various levels e.g., the lack of sufficient hardware features facilities the design of small memory footprints. The hardware constraints also impact the design of many protocols and algorithms e.g., data aggregation algorithms executed in resource-constrained IoT networks. It also prohibits the integration of many desirable components such as GPS receiver on the IoT devices.

Due to the resource-constrained nature of pervasive applications, there are restrictions on the type of data processing algorithms that run on an IoT device. Various lightweight data processing algorithms are used on the IoT node as the feasible solution in such scenarios [3]. But as the size of data increases, even implementing a lightweight algorithm on IoT node adds resource overheads. Therefore, the data is an essential factor that imposes restrictions on the use of resources in IoT nodes of pervasive applications [4]. Multiple aspects related data are taken into consideration while managing resources of an IoT device and these include the size of data, type of data, the structure of data, number of transmission and reception of data packets, aggregation of data, duration of data storage, etc. Therefore, data management becomes necessary to manage resources in these environments.

There have been recent studies in data management solutions for the management of resources in such applications through data aggregation mechanisms [5], data storage solutions [6], virtualization techniques [7], architecture-based solutions [8], lightweight data security approaches [9], and data parallelization approaches

*Department of Computer Science Engineering, National Institute of Technology Srinagar, India (saniyazahoor@nitsri.net).

[10]. However, most of the work has been reported in Wireless Sensor Networks (WSNs) and lacks comprehensive experimental evaluations. There has been little work in Internet of Things, as such efforts are required to design a data management solution for pervasive IoT. This paper proposes a parallelization based data management framework to minimize resource overheads on the individual IoT nodes in pervasive IoT applications. The comparative analysis of resource consumption is drawn at the node level for the proposed framework with respect to the sequential approach [11] through simulations and empirical data analysis.

The rest of the paper is organized as: Section 2 presents the literature survey, Section 3 discusses proposed data management framework, Section 4 gives experimental and evaluation details, Section 5 presents the results and discussions, and section 6 gives conclusions.

2. Literature Survey.

Due to the resource-constrained nature of IoT devices, pervasive IoT faces great challenges of data management in terms of storing, processing, and communicating the data [4]. Traditionally, data management is related to the data lifecycle requirement of a system. In the context of IoT, data management serves as a layer between the devices generating the data and the applications utilizing the data for analysis [12]. In our context, data management in pervasive IoT is visualized as the deployed algorithm, protocol, architecture, or framework that primarily focuses on managing limited resources of IoT devices through the proper management of data. There has been significant research in data management for pervasive applications through data aggregation mechanisms, data storage solutions, data management models, virtualization techniques, architecture-based solutions, lightweight data security approaches, and data parallelization approaches as discussed below.

2.1. Data Aggregation Mechanisms.

One of the widely used and recognized techniques for data management in most of the pervasive IoT applications is data aggregation. Few comparative studies have been carried out in data aggregation for optimization of resources such as network lifetime, processing, storage, data redundancy, reliability, latency, etc [13]. The possible mechanisms for data aggregation include centralized, cluster-based, and tree-based. In centralized data aggregation, a central node aggregates the data from each IoT node [14]. The work in [15] presents a centralized data aggregation approach that addresses scalability and heterogeneity issues but the limitation of the work is the single point of failure. In cluster-based data aggregation, a cluster head is selected in each cluster to reduce the data traffic in the network. Considerable work in this includes Chinese remainder based theorem [16], cross-layer data aggregation [17], low resource cost data aggregation [18], etc. In tree-based data aggregation, the intermediate nodes perform the aggregation of data. Considerable work in this includes energy-efficient tree approach [19][20], tree-based data aggregation approach to balance the energy and network load [21], etc. However, most of the work in IoT data management through data aggregation has been reported in WSNs and lacks a comprehensive performance evaluation framework.

2.2. Data Storage Solutions.

In pervasive IoT applications, IoT devices generate data rapidly, as such, it becomes necessary to store the data efficiently. Further, the heterogeneity of IoT data collected from different sources necessitates the need for efficient data storage solutions ready to deal with heterogeneous IoT data. To address such problems, various data storage solutions have been proposed that allow efficient storage ad integration of heterogeneous data viz., structured, and unstructured data [6]. Several aspects are taken into consideration for data storage solutions and these include the type of data, location of data, duration of storage, etc. Several data storage schemes that help in management of resources include data-centric storage [22], provenance-aware storage [23], real-time databases [24], centralized storage [25][26], etc.

2.3. Architecture based solutions.

Depending on the specific requirements and design constraints of applications, the architecture of IoT varies from one application to another. Based on the available study, IoT architectures have been classified into centralized, distributed, and service-oriented architectures. The centralized architectures are one of the most widespread models for data management in IoT applications having limited resources. In this, the sensed data from the IoT devices are transferred to a single central device (or location), which combines, processes and presents the information to the end-users. The work in [7] presents a centralized scheduling method, 6TiScH in which the amount of resource consumption is modeled using historical information about resource consumption. The work in [27] also presents a centralized architecture, based on web resources, to decouple the domain of heterogeneous devices from the application development.

But these approaches fail to guarantee scalability and interoperability among devices of different application domains.

To overcome the disadvantages of centralized architectures, the distributive IoT architectures have been introduced. The distributive IoT architectures provide services at the node level and at the network level by the collaboration of nodes and users to achieve a common goal. One of the disadvantages encountered in this approach is a security breach that occurs due to device mobility and network heterogeneity [28]. To overcome this in pervasive and distributive IoT, few studies introduce master key for entity authentication [29] [30] [31]. Another disadvantage of distributive IoT architecture is excessive energy consumption and to overcome this, the work in [32] presents a decentralized scheduling approach based on Proportional, Integral, and Derivation algorithm that operates dynamically and controls the data traffic in the network to reduce the unnecessary consumption of energy.

The Service-Oriented architecture (SOA) is a distributive architecture of autonomous services executing on nodes with different service providers. SOA enables the decomposition of large networks into well-defined IoT networks in which nodes execute an arbitrary number of services and exchange information among them without human intervention [33]. Considerable work has been carried out in Service-Oriented Architectures for WSN and IoT applications. These include OASIS Reference Model for Service Oriented Architecture (SOA-RM) [34], a Web Service Middleware for Ambient Intelligence (aWESoME) [35], Knowledge-Aware and Service-Oriented architecture (KASO) [36], micro-subscription management system (mSMS) [37] , etc. SOA is considered applicable for IoT environments but there are many challenges such as security, limited resources, etc that need to be taken care of while adopting SOA into the IoT environment [38].

2.4. Virtualization Techniques. IoT environment consists of a huge number of IoT devices that produce data of variety in type, size, and formation, and this imposes great challenge of data management in pervasive applications. Virtualization techniques can efficiently handle this complexity in data. In pervasive IoT, the trend is to introduce virtualized environments at the node level or the network level. At the node level, multiple applications run their tasks concurrently on an individual IoT node Considerable work in this includes the container-based virtualization used on IoT devices for creation and initiation of virtualized instances [39], docker containerization used on Edge platforms to improve the manageability of resources and services [40], light-weight virtualization for IoT gateways that provides better IoT services [41], container Edge-cloud PaaS architecture that minimizes the consumption of energy [42], etc. However, it considers only a limited number of Raspberry Pi boards for performance evaluation and lacks comprehensive energy and power evaluations. On the other hand, the network-level virtualization creates a Virtual Sensor Network (VSN) containing a subset of sensor nodes that performs a given task, while other sensor nodes remain reserved for other tasks e.g. the Radio Access Network (RAN) virtualization that dynamically provides an isolated network in IoT [43], Long-Term Evolution (LTE) network virtualization using hypervisor software [44], etc.

2.5. Lightweight Data Security Approaches. Enabling data security in pervasive IoT is a challenging task due to the resource limitations [45]. Several data management strategies are used for optimizing the performance of security algorithms at hardware and software levels [46][47], but only a few are implemented in resource-constrained IoT networks. Recently, there has been a huge demand for lightweight authentication and encryption for securing resource-constrained IoT devices. These lightweight algorithms aim to balance security and resource costs to achieve privacy and performance advantages in IoT [48][49][50]. Due to their low processing power, limited battery life, small size, and small memory, the lightweight data security algorithms are considered efficient for securing resource-constrained devices in pervasive IoT applications [51][52][53].

Several lightweight two-factor user authentication schemes have been proposed for WSNs, but such schemes seem vulnerable to several attacks such as replay, denial of service, etc. Few lightweight security schemes, based on the computation of hashing function, have also been proposed for WSNs [54][55] but the work is vulnerable to several attacks (e.g., login identity attack). Various schemes such as password-based user authentication, mutual authentication, etc have also been proposed but these fail to satisfy mutual authentication between the base station and the sensor node [56]. To address such issues, the work in [57] proposes a lightweight protocol to secure IoT devices via portal controllers; the proposed protocol preserves the privacy of communications; however, it generates an overhead regarding the number of messages exchanged among IoT devices.

Predominant research is done on lightweight key management protocols for IoT devices that guarantee data confidentiality and constrained node authentication during data transmission along the channel; the limitation of the work is that the security protocol does not specify resource overheads in IoT environments [58]. The work in [59] proposes a similar lightweight security framework for resource-constrained smart objects but the proposed framework was not integrated into the resource-constrained IoT environments to evaluate its suitability. The work in [60] presents a lightweight authentication protocol to enable security on computationally constrained RFID tags; the proposed protocol guarantees minimum computation overhead with better authentication among RFID tags. To address the security and privacy concerns in constrained IoT environments, SecKit, a security toolkit has been proposed [61]; the drawbacks of this approach is it does not provide information on how to deploy security and privacy solutions for devices operating in a dynamic IoT environment.

There has been also tremendous research on lightweight data security schemes based on attribute encryption [62][63][64]. The work in [62] proposes a lightweight Attribute-Based Encryption scheme that decreases the resource overhead in terms of computation and communication of IoT environments. The work in [63] presents a similar attribute-based encryption scheme that ensures a trade-off between computation and storage capacity of constrained devices. The work in [64] discusses a similar lightweight attribute-based encryption scheme for heterogeneous IoT applications with a disadvantage of high bandwidth consumption. Therefore, security in such environments is a serious issue because of resource constraints in IoT devices and networks; also the number of attacks is a bit higher [65]. The existing security techniques do provide a basis for privacy and security, but these techniques cannot be used in resource-constrained IoT without modifications [66].

2.6. Data Parallelization Approaches. There has been considerable research in data parallelization for efficient management of limited resources. Several serial approaches are being used for analyzing small data sets but it results in increased resource overheads especially storage and processing [10]. To address this, data parallelization approaches are being used that harness multiple processing units to solve this problem wherein multiple processing units execute the computation simultaneously to reduce computing time [67].

In IoT applications, which follow in-network processing IoT architectures, the individual IoT nodes are overloaded due to the size of data [68]. To address this, parallel approaches are being widely used in many IoT applications such as medical imaging, bio-informatics, graph mining, etc. The work in [69] presents a parallel computing framework to integrate diverse computing resources for manufacturing IoT applications. The work in [70] proposes a similar framework that uses the cloud to optimize process planning. The work in [71] develops a parallel algorithm to achieve highly efficient service decomposition and optimal selection. However, most of these studies focus on strategical problems in manufacturing e.g., job scheduling and service optimization. In pervasive IoT, parallel algorithms are mapped on FPGA to obtain minimal power consumption [72][73], while others use data parallelization to reduce the device overload [74]. However, most of the work on improving the resources of IoT devices is done at the hardware level only.

Table 2.1 presents the classification of recent studies and related information in data management for pervasive IoT.

3. Proposed Data Management Framework. In pervasive applications of IoT, the resource utilization should be optimized because of resource constraints, as such, the algorithms and protocols used in the data management framework should be designed accordingly. In resource-constrained nodes, there are restrictions on the use of data processing algorithms due to resource overheads. To reduce this overloading on IoT nodes, we propose a data management framework that implements any lightweight algorithm with low resource overheads. The framework uses the concept of data parallelization on multiple nodes for minimizing the overloading on the individual IoT node in IoT networks. For the implementation of a lightweight algorithm in such a scenario, the Tiny Encryption Algorithm is used. The TEA is a lightweight cryptographic algorithm and due to its simplicity of description and implementation, the algorithm is more suited for the resource-constrained IoT scenario.

Since, for the implementation and evaluation purposes, we have used a data security algorithm to test the proposed framework, so the prime objectives of the proposed framework shall be security along with optimal resource utilization. Table 3.1 highlights the various resource and security concerns of various pervasive IoT applications.

3.1. Formulation of Proposed Framework. The proposed data management framework applies to homogeneous in-network processing architectures of IoT which includes distributed data-centric IoT architec-

Table 2.1
Classification of Recent Studies and Related Information in Data Management for Pervasive IoT

IoT Data Management	**Related Work**	**Performance Metrics**	**Issues**	**Research Possibilities**
Data Aggregation Mechanisms	Centralized data aggregation [19]	Heterogeneity, Scalability	Single point of failure	With modifications manage resources in IoT
	Cluster based data aggregation [16]	Energy efficiency,Security	High Hardware costs	
	Tree based data aggregation [14]	Network lifetime, Throughput, Delay	High computational cost	
Data Storage Solutions	Data-centric storage [22], Provenance-aware storage [23], Real time databases [24], etc	Storage utilization	Energy Consumption Overhead	Limited work in IoT
Architecture based Solutions	Centralized Solutions [27]	Resource Management	Congestion and single point of failure, Scalability issues	Standardization of IoT resource management architecture
	Distributed Solutions [28]	Bandwidth utilization, Reduced number of transmissions	Security Breaches, Excessive energy consumption	
	Service Oriented Solutions [38]	Reduced number of transmissions	Security Breaches	
Virtualization Techniques	Node Level Virtualization [41, 42]	Energy, manageability of resources and services	Limited number of devices for performance evaluation	Exploration at the system level
	Network Level Virtualization [43, 44]	Dynamic network provisioning	Limited work in IoT	Exploration at the system level
Lightweight Data Security	Two-Factor User Authentication Schemes [56, 57]	Data Security	Message Exchange Overhead	Limited Work in IoT
	Key Management Protocols [58, 59]	Data Confidentiality	Limited work in resource-constrained IoT	
	Attribute-based Encyption Schemes [64, 63]	Less Computation and Communication Costs	Bandwidth Utilization	
Data Parallelization Approaches	Parallel computing frameworks [69, 70], Parallel algorithms on FPGA [72]	Less Power Consumption	Limited work at software level	More attention needed on data management at hardware level

tures, where data processing such as encryption, data fusion, etc are done at the node level. In the proposed framework, we consider IoT applications wherein the IoT nodes are resource-constrained and non-replaceable after the deployment for a particular application.

3.1.1. Assumptions. Following assumptions are taken in our IoT scenario:

- An IoT scenario consists of super-master, master and slave nodes,
- Each slave node, S_{in} must have a unique ID, I_{Miid},
- Each master node, M_i must have a unique ID, M_{iid},
- Resources of slave nodes $\{E_i,\ P_i,\ M_i\}$ is equivalent to that of master nodes $\{E_{Mi},\ P_{Mi},\ M_{Mi}\}$,
- Resources of super master node $\{E_{1i},\ P_{1i},\ M_{1i}\}$ are higher than that of slave nodes $\{E_i,\ P_i,\ M_i\}$ and master nodes $\{E_{Mi},\ P_{Mi},\ M_{Mi}\}$,

Table 3.1
Pervasive IoT Applications: Resource and Security Perspectives

Pervasive IoT Application	**Main Resource Concern**	**Main Security Concern**
Healthcare [75]	Limited Storage, Processing, Energy and Bandwidth	Identity of Patients, Health records of Patients, Patient-Doctor Communication, etc
Transportation and Parking [76, 77]	Limited Storage, Processing and Bandwidth	Information of driver/Traveler, Inter-vehicular communications, Breaching Traffic Lights for hijacking, etc.
Monitoring of Weather, Environments, Waste [78, 79]	Limited Storage, Processing, Energy and Bandwidth	Confidentiality of critical information, Integrity of sensed data, etc.
Agriculture [80, 81]	Limited Storage, Processing, Energy and Bandwidth	Personal details of farmers, Integrity of sensed data, False negative and False positives lead to disastrous results for crop, etc.
Offices and Homes [82]	Limited Storage, Processing and Bandwidth	Owner Authentication, Household data, Monetary loss due to privacy breaches in online purchases from home, etc.

- Distance between the slave node and the master node , $d(M_i$ to $I_{Mi})$ is negligible as compared to the distance between the master node and super master $d(E_1$ to $I_{Mi})$,
- TEA encryption is performed by slave nodes S_{in},
- Data aggregation and distribution is done by the master nodes M_i,
- Data aggregation and decryption is done by super-master node (Edge node) E_1.

3.1.2. Proposed Algorithm. In the proposed algorithm, the data is divided into the data chunks on the master node and is sent to the slave nodes that perform the encryption of data. After encryption of data chunks is performed, the encrypted data chunks are then pushed to the super-master node that performs the decryption of data.

As shown in Figure 3.1, the proposed framework aims to parallelize TEA implementation across the virtualized OS containers of neighboring slave nodes that help in mass data parallelism. The sensed and collected data from the master node is split among N slave nodes e.g., consider a scenario with 104 blocks of 32-bit plaintext on the master node and four slave nodes are volunteering to work for the master, then each slave node gets 26 blocks of 32-bit plain text.

A slave node is selected based on availability quotient, A which is a function of resources such as energy (e) and memory (m):

$$A = f(e, m) \tag{3.1}$$

For the encryption process, the slave nodes in the cluster run TEA in their OS containers in parallel and simultaneously. After the encryption process on slave nodes is completed, the encrypted data is gathered from all slave nodes to the super master. The super master node then sends an acknowledgment to the master node on the reception of the data. As the master node receives the acknowledgment of data chunks, the cluster vanishes, and the data is deleted from the memory of the master node and slave nodes.

For the decryption process, the super master node performs the decryption of the encrypted data chunks; the resulted decrypted data chunks will be stored on separate OS containers of the super master node pertaining to the master node. The pseudocode for the proposed algorithm is given in Algorithm 1.

4. Experimental Setup. At the time of deployment, all the nodes start functioning and acquire sensing data from the physical environment. It is only at the time of data dissemination to the edge, the node requires the help of nearby nodes for the implementation of a lightweight algorithm in parallel mode. The proposed data management framework has been tested through simulations and empirical data analysis. The simulations are carried out in MATLAB and simulation parameters are set as shown in Table 4.1.

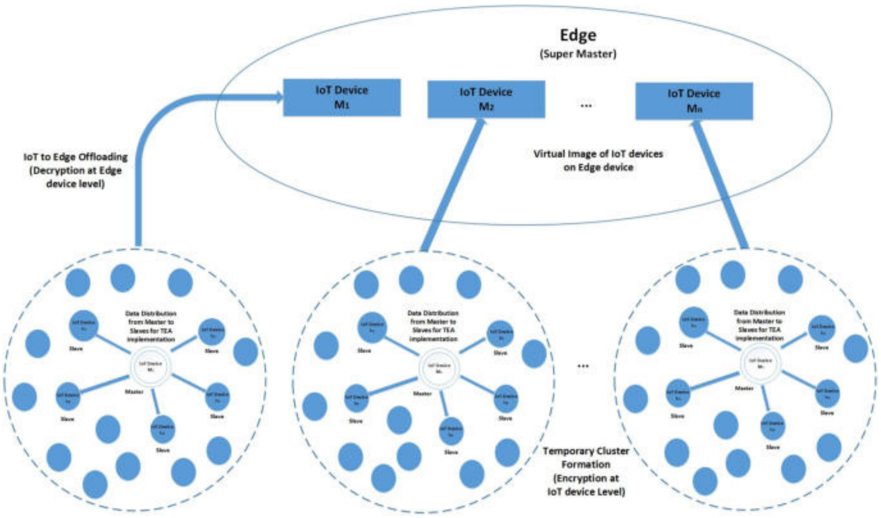

Edge
(Super Master)
IoT Device M1
IoT Device M2
...
IoT Device Mn
IoT to Edge Offloading (Decryption at Edge device level)
Virtual Image of IoT devices on Edge device
Data Distribution from Master to Slaves for TEA implementation
Slave
Master
...
Temporary Cluster Formation (Encryption at IoT device Level)

FIG. 3.1. *Proposed Data Management Framework for Resource Constrained IoT Environment*

Algorithm 1 Proposed Algorithm

```
A node, i collect data over a period of time, t
Calculate the data size periodically
As the data size of the node, i reaches a maximum size of a node D
Node i acts as a master node, $M_i$ with ID $M_{id}$
(Temporary Cluster Formation)
For each master node, Mi there is a cluster of available nearby nodes, i to act as slaves, $S_{in}$
$M_i$ broadcasts a request for cluster formation to the nearby nodes, i
N nodes accepts the request of Mi to act as slave nodes,$S_{i1}$,$S_{i2}$,$S_{i3}$,..., $S_{in}$
$S_{in}$ are selected by Mi based on Availability quotient, A where:
                    $A=f(E_i,M_i)$
Divide the data of data size, D into data chunks $D_{i1}$,$D_{i2}$,$D_{i3}$, ..., $D_{in}$
$M_i$ sends $D_{i1}$,$D_{i2}$,$D_{i3}$, ..., $D_{in}$ to the selected slave nodes, $S_{i1}$,$S_{i2}$,$S_{i3}$, ..., $S_{in}$
Each of N selected $S_{in}$ implements TEA in parallel
Each of N selected $S_{in}$ sends the encrypted data to the super master node, $E_1$
Encrypted data is collected by $E_1$ and acknowledgement is send to $M_i$
Cluster vanishes
Data is deleted on $M_i$ and $S_i$
Decryption process is performed on the $E_1$
Encrypted data Chunks from $S_{in}$ are decrypted and aggregated on the $E_1$
Data is stored on the separate containers for each master nodes ($M_i$) maintained on the super master node, $E_1$
```

TABLE 4.1
Simulation Parameters

Parameter	**Values**
Number of Slave IoT devices	100
Number of Master IoT devices	20
Number of super master device	1
Initial Energy of slave IoT device	300 mAh
Initial Energy of Master IoT device	300 mAh
Transmission range of a slave IoT device	40 m
Transmission range of master IoT device	40 m
Block Size	256
Key Size	128 bits

For the empirical data analysis, an Arduino based IoT device with sensors (air humidity, temperature, and pressure)is used to capture data from the physical environment. The proposed framework is tested on this data to check its suitability to map with the real-life IoT scenarios.

4.1. Performance Metrics. The comparative resource analysis is performed in terms of following metrics:

- *Round*: It is defined as the complete process which starts when the IoT node aggregates the data, implements the lightweight algorithm, and pushes the data to the edge.
- *Energy*: The energy of a node is the difference between total energy and consumed energy in an IoT system. To extend the network lifetime, it may be desirable to avoid routing through nodes with low residual energy. It is calculated as:

  $$E_i = E_t - E_c \tag{4.1}$$

  where E_i is the residual energy, E_t is the total energy of a node, and E_c is the energy consumed in one round.

  The average energy of the IoT network, E_n is calculated as:

  $$E_n = \sum E_i / i \tag{4.2}$$

  where i denotes the number of active IoT nodes.

  Let E_{TEA} and E_{P-TEA} is the average energy for TEA implementation in sequential and proposed framework respectively.
- *Storage*: Let S_r represents the residual storage on the node, then we have:

  $$S_r = S_t - S_o \tag{4.3}$$

  where S_t is the total storage of a node and S_o is the occupied space on the node. Let M_{TEA} and M_{P-TEA} is the measure of average storage requirements for TEA implementation on a single node in sequential and proposed framework respectively.
- *Processing Time*: It is the time required to process the data in a network. Processing of data involves activities such as sensing, storing, aggregation, offloading, etc. Let T_{TEA} and T_{P-TEA} is the average processing time for TEA implementation in sequential and proposed framework respectively.
- *Number of Alive Nodes versus the number of rounds*: Let A_{TEA} and A_{P-TEA} is the number of alive nodes in sequential and proposed framework respectively.
- *Degree of improvement*: Degree of improvement is referred to the percentage increase in performance of proposed framework with respect to sequential TEA implementation.

  The degree of improvement in processing time is calculated as:

  $$D_{P-TEA} = (T_{TEA} T_{P-TEA}) / T_{TEA} \tag{4.4}$$

  and the degree of improvement in storage is calculated as:

  $$Dm_{P-TEA} = (M_{TEA} M_{P-TEA}) / M_{TEA} \tag{4.5}$$

5. Results and Discussions. The proposed data management framework has been tested through simulations and empirical data analysis. The performance parameters used to evaluate the proposed framework with respect to the sequential approach include energy, number of alive nodes, storage, and processing time. The results are presented in the form of graphs and explanation.

Figure 5.1 represents the energy of a node during simulation (as shown by eq. 4.1 and 4.2). It is evident from the graph that the energy consumption of a node in the proposed framework is lesser as compared to the single node implementation of a lightweight algorithm. In the case of the sequential implementation of a lightweight algorithm, the energy of a node gets exhausted earlier as compared to the proposed approach. There is a 12.5% improvement in the energy of a node in the proposed framework as compared to a sequential approach.

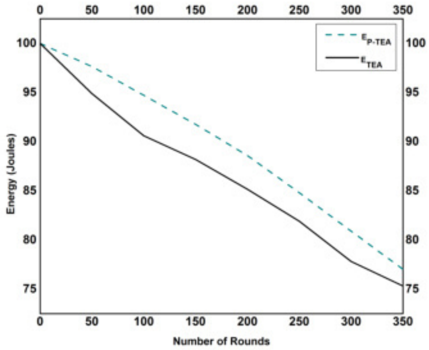


FIG. 5.1. *Energy versus Number of Rounds*

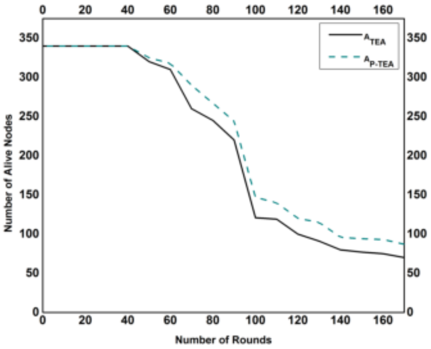


FIG. 5.2. *Number of Alive Nodes versus Number of Rounds*

Because of the parallelization of data, significant energy is saved on an IoT node. Since each node contributes to the overall network lifetime, therefore, the lifetime of the network is increased. Also, in a sequential approach, data overloading on a single node leads to the early death of the node, which can create voids in the network, and the lifetime is further reduced. Further, there is a direct relationship between energy and network lifetime. More is the energy of a node in the network, the longer is the network lifetime.

Figure 5.2 shows the number of alive nodes versus the number of rounds. It is evident from the graph that the number of alive nodes in the proposed framework is more at any particular instant as compared to the single node implementation of lightweight algorithm i.e., nodes in the proposed framework last for a longer time. Since the proposed approach balances the energy consumption and node deaths gracefully, it results in network stability. There is a 25% improvement in the number of alive nodes in the proposed framework as compared to the sequential approach.

In the proposed approach, the processing in each node is divided among different nodes and it results in fair utilization of resources in an IoT network. While in the case of a sequential approach, a particular node is overloaded while other neighboring nodes are underutilized. This results in the formation of holes in the network.

Figure 5.3 represents the comparison of storage used for the TEA implementation on a single node versus on the multiple salve nodes in the proposed framework (calculated by eq. 4.3). We observed that the storage required in sequential implementation is higher as compared to the proposed approach at a particular instant of time. In the proposed approach, as the number of slave nodes in the cluster increases, the memory requirements decrease by a factor of N i.e., higher the number of slave nodes in the cluster for the implementation of TEA

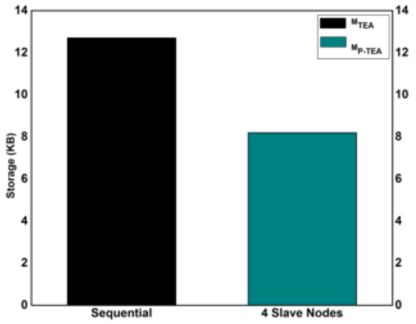


FIG. 5.3. *Storage for 256 Data Blocks*

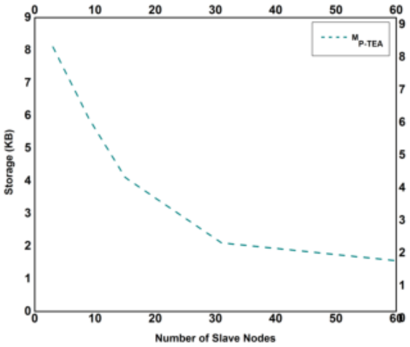


FIG. 5.4. *Storage versus Number of Slave Nodes*

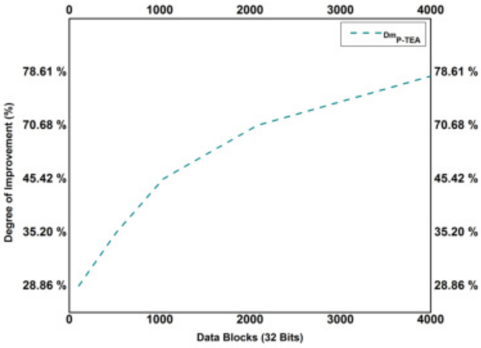


FIG. 5.5. *Degree of Improvement in Storage versus Data Blocks*

in parallel mode, the lesser the memory requirement.

Figure 5.4 shows the impact of the number of slaves on the storage in the proposed framework for 4160 blocks of 32-bit data. It is observed that the storage used is further reduced if the number of slave nodes for data processing is increased.

Figure 5.5 illustrates the degree of improvement in the storage of TEA implementation in a sequential approach versus the proposed data management framework. It is observed that the value of the degree of improvement in the proposed framework increases as the size of input data increases i.e. if there is huge data, the proposed framework proves to be more advantageous in saving resources in IoT networks. There is 28.86% to 78.61% improvement in the storage requirements of a node in the proposed framework as compared to the sequential approach (calculated by eq. 4.4).

Figure 5.6 shows the impact of cluster size on the processing time in the proposed data management framework for 4160 blocks of 32-bit data. It is observed that the processing time is reduced if the number of slave nodes for data processing is increased because the data size is reduced and time to execute on the same processor is divided by a factor N number of nodes.

Figure 5.7 shows the execution time of TEA on multiple nodes is less compared to the single node TEA implementation due to data parallelization for processing.

Figure 5.8 illustrates the degree of improvement in the processing time of TEA implementation in sequential versus the proposed approach. It is observed that the value of the degree of improvement in the proposed framework increases as the size of input data increases i.e. if there is huge data, the proposed framework proves to be more advantageous in saving the resources in IoT networks. There is 41.86% to 68.61% improvement in

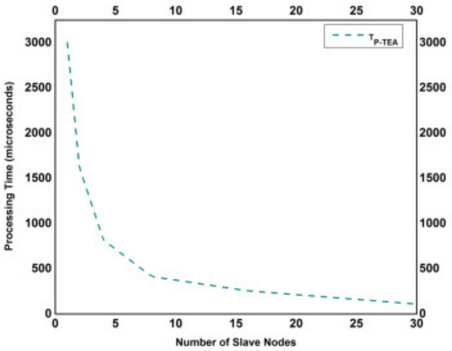


FIG. 5.6. *Processing Time versus Number of Slave nodes*

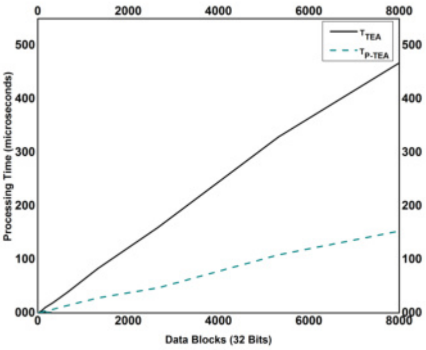


FIG. 5.7. *Processing Time versus Data Blocks*

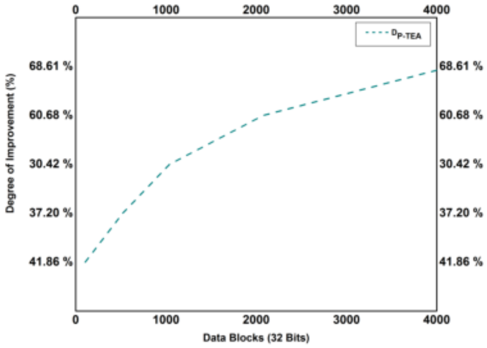


FIG. 5.8. *Degree of Improvement in Processing Time versus Data Blocks*

processing time in the proposed framework as compared to the sequential approach (calculated by eq. 4.5).

The proposed framework is evaluated using empirical data as well. Figure 5.9, 5.10 and 5.11 shows the energy versus time, storage versus the number of slave nodes, and processing time versus the number of slave nodes in the proposed framework as compared to the sequential approach.

It is evident from the Figures 5.9, 5.10 and 5.11 that the energy, storage and processing time requirements on the IoT device in the proposed framework are less as compared to a sequential approach.

6. Conclusions. In most of the pervasive applications of IoT, the resources of an IoT device are constrained and due to these resource limitations, there are restrictions on the use of data processing algorithms on an IoT node e.g. encryption and decryption of data causes an overhead of storage, processing, and energy to the existing data on an IoT node. To reduce these resource overheads in such applications, data management is important as data is the common factor that imposes restrictions on the use of resources. This paper proposed a data management framework that uses the concept of data parallelization on multiple nodes for minimizing the amount of data for processing on an individual IoT node and communication of data over longer distances.

From the comparative analysis of resources used in implementing a lightweight algorithm on the proposed framework and in sequential mode, the simulation results showed the proposed framework is better than the sequential one in terms of energy, storage, and processing time. Because of parallelization, the proposed framework results in fair utilization of resources and significant saving of resources. While in the case of a sequential approach, a particular node is overloaded while neighboring nodes are underutilized, resulting in fast depletion of resources especially energy which creates holes in the network. There is an improvement of 12.5% in

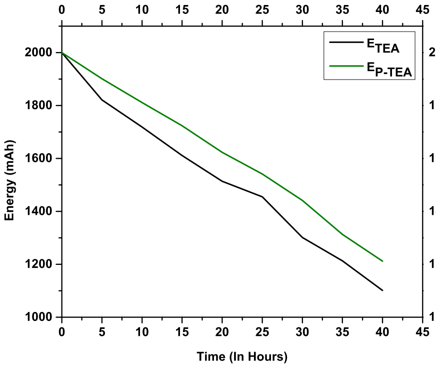


FIG. 5.9. *Energy versus Time*

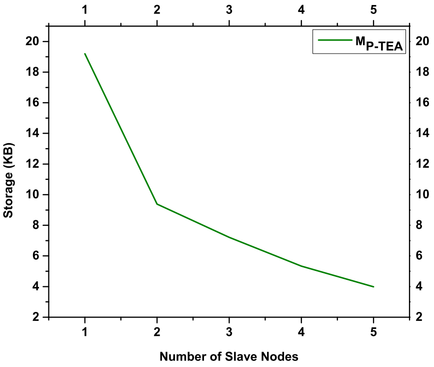


FIG. 5.10. *Storage versus Number of Slave nodes*

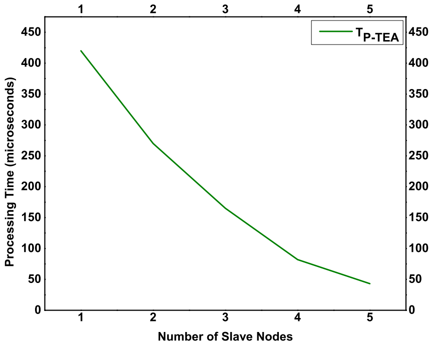


FIG. 5.11. *Processing Time versus Number of Slave nodes*

energy and 25% in the number of alive nodes in the proposed framework as compared to the sequential approach. We have observed that the storage and processing requirements of a node in the proposed framework is lesser and as the number of slave nodes for a master node increases to N, the storage and processing requirements of a node is decreased by a factor of N. There is 28.86% to 78.61% improvement in the storage of a node as the size of data on the master node increases to 4000 data blocks. Also, better performance has been observed in terms of processing time with 41.86% to 68.61% improvement in the network.

The proposed framework has been tested on the empirical data as well and it also showed improved resources in the proposed framework as compared sequential approach, making the proposed solution suitable to map with the real-life scenarios of pervasive IoT applications. We also conclude from the implementations that larger the number of slave nodes for the TEA implementation, lesser the resource consumption of a node for a particular interval of time but there has to be a limit for the division of work among the slave nodes in the cluster, otherwise it will increase the operational cost both at the node and network level.

In this paper, we have not compared the proposed data management framework with any other framework because the literature available for parallelization based data management frameworks for resource-constrained IoT applications is very scarce and the ones described in data parallelization approaches are mostly designed and implemented at hardware level only. One of the interesting future research developments of this paper lies in comparing the proposed parallelization based data management framework at the hardware level with the existing solutions. Also, the proposed framework can be extended to include more resource parameters and constraints to map with the real-world IoT scenarios.

REFERENCES

[1] Di Martino Et.al *Internet of things reference architectures, security and interoperability: A survey*, *Internet of Things* , 1, pp.99-112 (2018).

[2] S. Zahoor and R.N. Mir, *Resource management in pervasive Internet of Things: A survey*, *Journal of King Saud University-Computer and Information Sciences*, (2018).

[3] Z. Sheng Et.al, Lightweight management of resource-constrained sensor devices in internet of things, *IEEE internet of things journal*, 2(5), pp.402-411 (2015).

[4] S. Zahoor and R.N. Mir, *Resource management in pervasive Internet of Things: A survey*, *Journal of King Saud University-Computer and Information Sciences*, (2018).

[5] B. Guidi and L. Ricci, *Aggregation Techniques for the Internet of Things: An Overview*,*In The Internet of Things for Smart Urban Ecosystems*,pp. 151-176, Springer, Cham (2019).

[6] L. Jiang Et.al,*An IoT-oriented data storage framework in cloud computing platform*, *IEEE Transactions on Industrial Informatics*, 10(2), pp.1443-1451 (2014).

[7] P. Thubert Et.al, *6TiSCH centralized scheduling: When SDN meet IoT*, *In Proc. of IEEE Conf. on Standards for Communications & Networking (CSCN'15)*, (2015).

[8] A. Celesti Et.al, *Exploring container virtualization in IoT clouds*, *In 2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 1-6, IEEE (2016).

[9] S. Li and L. Da Xu, *Securing the internet of things*, *Syngress*,(2017).

[10] J.T. Townsend, *Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished*, *Psychological Science*, 1(1), pp.46-54 (1990).

[11] M.J. Quinn,*Parallel programming*, TMH CSE, 526 (2003).

[12] B. Diene,*Data management techniques for Internet of Things*, *Mechanical Systems and Signal Processing*, 138, p.106564 (2020).

[13] H. Rahman Et.al, *Comparison of data aggregation techniques in Internet of Things (IoT)*, *In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1296-1300, IEEE (2016).

[14] B. Pourghebleh. and N.J. Navimipour,*Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research*, *Journal of Network and Computer Applications*, 97, pp.23-34 (2017).

[15] T. Zhu Et.al, *An architecture for aggregating information from distributed data nodes for industrial internet of things*, *Computers & Electrical Engineering*, 58, pp.337-349 (2017).

[16] F. Xie,*CaCa: chinese remainder theorem based algorithm for data aggregation in internet of things on Ships*, *In Applied Mechanics and Materials*, 701, Trans Tech Publications Ltd, pp. 1098-1101, (2015).

[17] A. Alkhamisi Et.al,*A cross-layer framework for sensor data aggregation for IoT applications in smart cities*, *In 2016 IEEE International Smart Cities Conference (ISC2)*, IEEE, pp. 1-6 (2016).

[18] Z. Li, *Lifetime balanced data aggregation for the internet of things*, *Computers & Electrical Engineering*, 58, pp.244-264 (2017).

[19] A. Koike Et.al, *Iot network architecture using packet aggregation and disaggregation*, *In 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, IEEE, pp. 1140-1145,(2016).

[20] Y. Liu Et.al *A novel trust-based secure data aggregation for internet of things*, *In 2014 9th International Conference on Computer Science & Education*, pp. 435-439, IEEE.

[21] S.G. Shilpa and S.M. Sundaram, *Data Aggregation Techniques Over Wireless Sensor Network-A Review*, (2013).

[22] Y. Qin Et.al, *When things matter: A survey on data-centric internet of things*, *Journal of Network and Computer Applications*, 64, pp.137-153 (2016).

[23] J. Ledlie Et.al, *Provenance-aware sensor data storage*, *In 21st International Conference on Data Engineering Workshops (ICDEW'05)*, pp. 1189-1189, IEEE (2005).

[24] W.J. Li Et.al, *Just IoT Internet of Things based on the Firebase real-time database*,*In 2018 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE)*, pp. 43-47, IEEE (2018).

[25] T. Fujita Et.al,*Centralized storage management method*, *U.S. Patent* 7,152,144 (2006).

[26] M.E. Rottsolk and S.P. Nolan, *Microsoft Corp*, *Centralized healthcare data management*, U.S. Patent Application 12/345,334 (2010).

[27] M. Kovatsch Et.al,*Moving application logic from the firmware to the cloud: Towards the thin server architecture for the internet of things*, *In Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, IEEE, pp. 751-756 (2012).

[28] R. Roman Et.al, *On the features and challenges of security and privacy in distributed Internet of Things*, *Computer Networks*, 57, 10, pp. 2266–2279 (2013).

[29] F. Zhu Et.al, *Private entity authentication for pervasive computing environments*, *International Journal of Network Security*, 14, 2, pp. 86–100 (2012).

[30] S. Shin Et.al, *An effective authentication mechanism for ubiquitous collaboration in heterogeneous computing environment*, *Peer-to-Peer Networking and Applications*, (2013).

[31] Y. Liu Et.al, *PKC based broadcast authentication using signature amortization for WSNs*, *IEEE Transactions on Wireless Communications*,11,6, pp. 2106–2115 (2012).

[32] M. Domingo-Prieto Et.al,*Distributed pid-based scheduling for 6tisch networks*, IEEE Communications Letters, 20(5), pp. 1006-1009 (2016).

[33] H. Gomaa, *Software modeling and design: UML, use cases, patterns, and software architectures*, *Cambridge University Press*, (2011).

[34] M. Kushwaha, *Oasis: A programming framework for service-oriented sensor network*, *In 2nd International Conference on Communication Systems Software and Middleware*, IEEE, pp. 1-8 (2007).

[35] T. G. Stavropoulos Et.al, *aWESoME: A web service middleware for ambient intelligence*, *Expert Systems with Applications*, 40(11), pp. 4380-4392 (2013).

[36] I. Corredor Et.al, *Knowledge-aware and service-oriented middleware for deploying pervasive services*, *Journal of Network and Computer Applications*, 35(2), pp. 562-576 (2012).

[37] M. S. Familiar Et.al, *Building service-oriented smart infrastructures over wireless ad hoc sensor networks: A middleware perspective*, *Computer Networks*, 56(4), pp. 1303-1328 (2012).

[38] L. Atzori, A. Iera and C. Giacomo Morabito, “The internet of things: a survey ”, Comput. Netw., 2010.

[39] M.J. Scheepers, *Virtualization and containerization of application infrastructure: A comparison*, *In 21st Twente Student Conference on IT*, 21 (2014).

[40] R. Hussain Et.al, *Federated Edge Computing for Disaster Management in Remote Smart Oil Fields*, *In 2019 IEEE 21st International Conference on High Performance Computing and Communications, IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 929-936, IEEE (2019).

[41] R. Petrolo Et.al, *The design of the gateway for the cloud of things*, *Annals of Telecommunications*, 72(1-2), pp.31-40, 2017.

[42] C. Pahl Et.al, *A container-based edge cloud paas architecture based on raspberry pi clusters*,*In 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pp. 117-124, IEEE (2016).

[43] L.E. Li Et.al, *Toward software-defined cellular networks*, *In 2012 European Workshop on Software Defined Networking*, pp. 7-12, IEEE (2012).

[44] L.E. Li Et.al *Toward software-defined cellular networks*, *In 2012 European Workshop on Software Defined Networking*, pp. 7-12, IEEE (2012).

[45] C.Y. Chen Et.al, *Securing real-time internet-of-things*, *Sensors, 18(12)*, p.4356 (2018).

[46] M. Nagendra and M.C. Sekhar, *Performance improvement of Advanced Encryption Algorithm using parallel computation*, International Journal of Software Engineering and Its Applications, 8(2), pp.287-296 (2014).

[47] T. Balamurugan and T. Hemalatha, *Parallelization of Symmetric and Asymmetric Security Algorithms for Multi-Core Architectures*, *International Journal of Science and Research (IJSR)*, 3(12),(2014).

[48] W. Sun Et.al, *Security and privacy in the medical Internet of Things: A review*, *Security and Communication Networks*, (2018).

[49] A. Shah and M. Engineer, *A survey of lightweight cryptographic algorithms for iot-based applications*, *In Smart Innovations in Communication and Computational Sciences*, Springer, Singapore, pp. 283-293 (2019).

[50] J. Srinivas Et.al, *Secure and efficient user authentication scheme for multi-gateway wireless sensor networks*, *Ad Hoc Networks*, 54, pp.147-169 (2017).

[51] F. Wu Et.al, *A lightweight and privacy-preserving mutual authentication scheme for wearable devices assisted by cloud server*, *Computers & Electrical Engineering*, 63, pp.168-181, 2017.

[52] Y. Qiu and M. Ma, *A mutual authentication and key establishment scheme for m2m communication in 6lowpan networks*, *IEEE transactions on industrial informatics*, 12(6), pp.2074-2085 (2016).

[53] R. Giuliano Et.al, *Security access protocols in IoT capillary networks*, *IEEE Internet of Things Journal*, 4(3), pp.645-657 (2016).

[54] I. Butun Et.al, *A survey of intrusion detection systems in wireless sensor networks*, *IEEE communications surveys & tutorials*, 16(1), pp.266-282 (2013).

[55] W. Chen, *An ibe-based security scheme on internet of things*, *in: Cloud Computing and Intelligent Systems (CCIS), 2012 IEEE 2nd International Conference*, 3, IEEE, pp. 1046–1049 (2012).

[56] C. Perera Et.al, *Context aware computing for the Internet of Things: a survey*, *IEEE Commun. Surv. Tutor.*, 16 (1), pp. 414–454(2014).

[57] T. Choi Et.al, *The best keying protocol for sensor networks*, *Pervasive and Mobile Computing*, 9(4), pp.564-571 (2013).

[58] Z. Wang Et.al, *Secure and efficient control transfer for IoT devices*, *International Journal of Distributed Sensor Networks*, 9(11), pp.503404 (2013).

[59] M.R. Abdmeziem and D. Tandjaoui, *An end-to-end secure key management protocol for e-health applications*, *Comput. Electr. Eng.*, 44, 184–197 (2015).

[60] F. Al-turjman and M. Gunay, *CAR Approach for the Internet of Things, approche de la CAR pour l internet des objects*, *Can. J. Electr. Comput. Eng.*, 39 (1), 11–18 (2016).

[61] R. Neisse Et.al, *SecKit: a Model-based Security Toolkit for the Internet of Things*, *Comput. Secur.*, 54, 60–76 (2015).

[62] X. Yao Et.al, *A lightweight attribute-based encryption scheme for the Internet of Things*, *Future Gener. Comput. Syst.*, 49, 104–112 (2014).

[63] N. Oualha and K.T. Nguyen, *Lightweight attribute-based encryption for the internet of things*, *in: 2016 25th International Conference on Computer Communication and Networks (ICCCN)*, IEEE, pp. 1–6 (2016).

[64] L. Touati Et.al, *C-cp-abe: Cooperative ciphertext policy attribute-based encryption for the internet of things*, *In 2014 International Conference on Advanced Networking Distributed Systems and Applications*, pp. 64-69, IEEE (2014).

[65] F. Restuccia Et.al, *Securing the internet of things: New perspectives and research challenges*, (2018).

[66] L. Da Xu Et.al, *Internet of things in industries: A survey*, *IEEE Transactions on industrial informatics*, 10(4), pp.2233-2243 (2014).

[67] C. Kan Et.al, *Parallel computing and network analytics for fast Industrial Internet-of-Things (IIoT) machine information processing and condition monitoring*, *Journal of manufacturing systems*, 46, pp.282-293 (2018).

[68] P.P. Ray *A survey on Internet of Things architectures*, *Journal of King Saud University-Computer and Information Sciences*,

30(3), pp.291-319 (2018).

[69] N.S. Raghavan and T. Waghmare, *DPAC: an object-oriented distributed and parallel computing framework for manufacturing applications*, *IEEE transactions on robotics and automation*, 18(4), pp.431-443 (2002).

[70] D. Mourtzis Et.al, *A cloud-based approach for maintenance of machine tools and equipment based on shop-floor monitoring*, *Procedia Cirp*, 41, pp.655-660 (2016).

[71] F. Tao eT.AL, *FC-PACO-RM: a parallel method for service composition optimal-selection in cloud manufacturing system*, *IEEE Transactions on Industrial Informatics*, 9(4), pp.2023-2033 (2012).

[72] V. Venugopal and D. Manikantan Shila, *Hardware acceleration of TEA and XTEA algorithms on FPGA, GPU and multi-core processors*, *In Proceedings of the ACM/SIGDA international symposium on Field programmable gate arrays*, ACM, pp. 270-270 (2013).

[73] P. Yalla and J.P. Kaps, *Lightweight cryptography for FPGAs*, *In 2009 International Conference on Reconfigurable Computing and FPGAs*, pp. 225-230, IEEE (2009).

[74] S. Kerckhof Et.al, *Towards green cryptography: a comparison of lightweight ciphers from the energy viewpoint*, *In International Workshop on Cryptographic Hardware and Embedded Systems*, Springer, Berlin, Heidelberg, pp. 390-407 (2012).

[75] S. Kim and S. Kim, *User preference for an IoT healthcare application for lifestyle disease management*,*Telecommunications Policy*, 42(4), pp.304-314 (2018).

[76] J. Sherly and D. Somasundareswari, *Internet of things based smart transportation systems*, *International Research Journal of Engineering and Technology*, 2(7), pp.1207-1210 (2015).

[77] S. Liniger and B. Stiller, *Parking prediction techniques in an iot environment*, Master's thesis (2015).

[78] F. Montori Et.al, *A collaborative internet of things architecture for smart cities and environmental monitoring*, *IEEE Internet of Things Journal*, 5(2), pp.592-605 (2017).

[79] V. Vishwarupe Et.al, *Zone specific weather monitoring system using crowdsourcing and telecom infrastructure*, *In 2015 International Conference on Information Processing (ICIP)*, pp. 823-827, IEEE (2015).

[80] J.M Talavera Et.al, *Review of IoT applications in agro-industrial and environmental fields*, *Computers and Electronics in Agriculture*, 142, pp.283-297 (2017).

[81] T. Qiu Et.al, *Framework and case studies of intelligence monitoring platform in facility agriculture ecosystem*, *In 2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pp. 522-525, IEEE (2013).

[82] S. Nagarkar, *IOT concept, technologies and applications for smart homes-A Review*, *Conference proceedings CTIcon* (2017).

Edited by: Anand Nayyar
Received: Apr 27, 2020
Accepted: Jul 17, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 479–496, DOI 10.12694:/scpe.v21i3.1764

PERFORMANCE ANALYSIS OF VIDEO ON-DEMAND AND LIVE VIDEO STREAMING USING CLOUD BASED SERVICES

UJASH PATEL[*], SUDEEP TANWAR[†] AND ANUJA NAIR[‡]

Abstract. The advent of Cyber-Physical Systems (CPS) has brought a revolutionary change coined as a mixture of information, communication, computation, and control. With applications in smart grid, health monitoring, automatic avionics, distributed robotics, etc., CPS is currently an area of attention among the academia and industry. The advancement of mobile communications and embedded technology has made it possible to build large scale CPS consisting of the interconnection of mobile phones. These devices collect information about the surrounding environment at any time anywhere basis through real-time video capture. Video streaming has proven to be a massive industry that is growing rapidly playing an important role in everyday life. Customer-driven approach wanting best experience with quality has to be the core offering of contemporary scenario. Video streaming is categorized into Video-On-Demand Streaming (VoDS) and Live Video Streaming (LVS) showing the current state-of-art opportunities. Many diverse applications of video streaming are military video surveillance using drones, live sports match player face recognition, on-demand video characters recognition, movie summarization like identifying parts of the movie which are viewed many times by different users, movie and series recognition, motion detection, gesture recognition, image segmentation, etc. This paper introduces an approach to develop video analysis on VoD and LVS using cloud-based services and analyzes the impact of Quality of Experience (QoE), cost, and bandwidth on the cloud. To achieve the best user experience for video streaming and video analysis, Content Delivery Network (CDN) offers the best QoE at various analyzed locations using various cloud providers like Amazon Web Services (AWS) CloudFront, Google Cloud CDN, Azure CDN, Akamai CDN, etc.

Key words: Cyber-Physical Systems, Live video streaming, Video on-demand streaming, Kafka, RabbitMQ, OpenCV, Spark, AWS Services

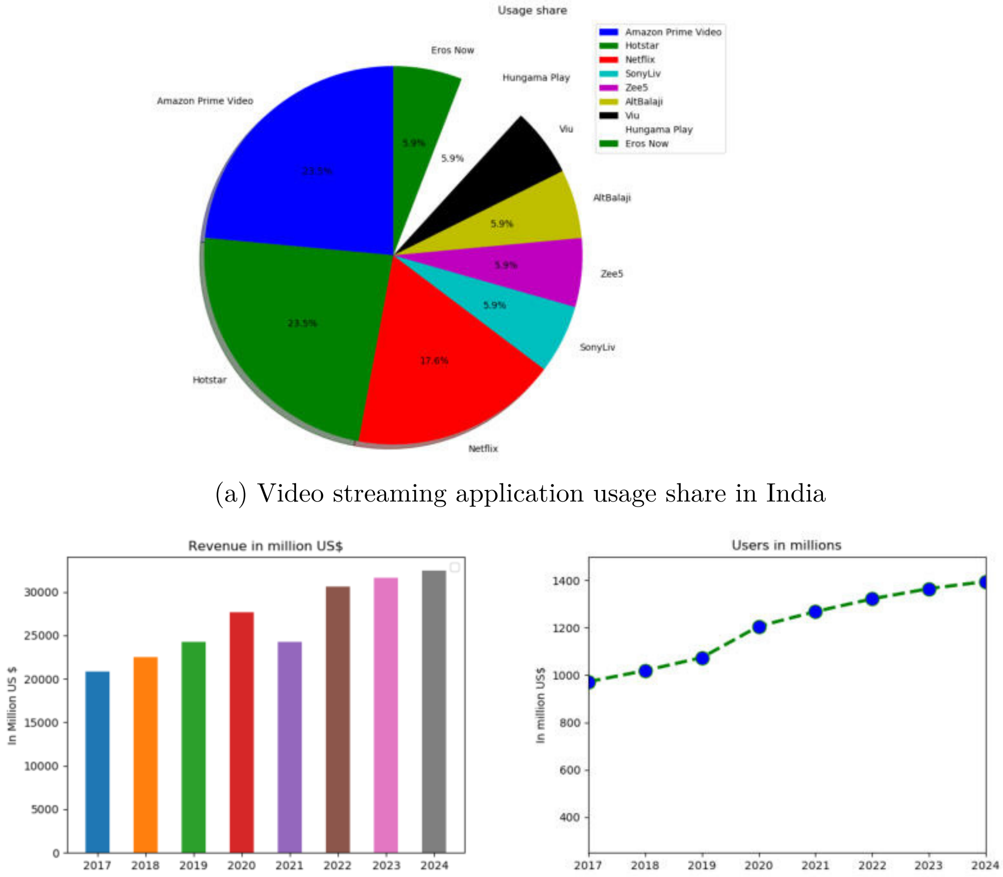
AMS subject classifications. 68M11

1. Introduction. Cyber-Physical Systems (CPS) [1] has arisen as a novel generation of engineered systems seamlessly assimilating the capability of computing, control, communication, and information with physical systems. Information sharing, organization & coordination among physical systems, human beings, and the Internet are often required by CPS. The advancement of mobile communications and embedded technology has made it possible to build large scale CPS consisting of the interconnection of mobile phones. These devices collect information about the surrounding environment at any time anywhere basis through real-time video capture. Sensing applications that are video-based equipped with camera-based mobile phones required to capture, send, and receive real-time videos for CPS is envisioned. Internet is playing a major role in the communication of the captured videos and live or broadcasted videos from mobile phones about the events taking place in the proximity of a user in real-time. In comparison with applications on mobile devices like for downloading and viewing videos, several challenges are faced by video-based CPS. These challenges include abundant broadband network access, relevant content delivery as per the substantial need of the hour, secure communication, unnecessary downloading of software required for inspecting these videos, etc. The adoption of VoD & LVS has advanced implementation methods to overcome these challenges. Video content providers like YouTube, Netflix, Hotstar, Facebook, Instagram, Snapchat, TikTok boasting 300 hours of video content uploaded to their space every minute have adopted different services to achieve the best user experience with

[*] Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India 382481 (17bce161@nirmauni.ac.in).

[†] Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India 382481 (sudeep.tanwar@nirmauni.ac.in).

[‡] Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India 382481 (anuja.nair@nirmauni.ac.in).



(a) Video streaming application usage share in India

(b) Revenue in a million US$ world wide

(c) Users in millions world wide

FIG. 1.1. *Video streaming market analysts reports with detailed comparisons [2]*

video delivery. LVS refers to having no time delay, ability to chat, ask, and respond to questions, etc. For example, live cricket match streaming like Hotstar streaming match where each frame serves to users immediately after it is captured.

On-demand streaming refers to high quality (HD) video and audio playing smoothly at any Internet speed where the video is already framed at different edge locations of the cache server with different resolutions. Considering the current scenario, due to the COVID-19 virus, there will be a direct impact on the video streaming market because almost all educational universities are moving towards online lecture video streaming. In the entertainment industry also, streaming services revenue will grow rapidly as shown in Figure 1.1a [2]. Figure 1.1b [2] indicates how revenue will increase in 2020 as total segment amount will be US$ 27,628M. Along with revenue increase, the number of users will also increase as shown in Figure 1.1c [2] where the number of users is expected to be 1,395.7M by 2024. The services like bandwidth analysis, capturing, uploading a video, encode frame, analysis video frame, decode the frame, convert the frame into different resolutions, distributed storage, user analysis data, and most important continuous delivery of frame in sequence frames to the end-user are required. All of that can be achieved by dividing services in microservices. In this paper, we will discuss how to capture a frame from a large scale user and analyze each frame and serve them to end-user using cloud-oriented services.

1.1. Need of Video Analysis. Video analysis has numerous advantages in today's streaming applications. Video analysis is used to analyze objects in a continuous frame using machine learning and deep learning algorithms. For example, in cricket LVS, we can identify player face reactions or predict ball direction, etc. To achieve the best video analysis system, we need to take care of how fast we can stream data to multiple nodes. Multiple nodes are required because video analysis needs a lot of computational power to analyze which is not reasonable using single thread or single CPU power. Instead, distributed computational power is the need of the hour. That would be a reason why we need Kafka as messaging services. An open-source stream-processing software platform written in Scala and Java coined as Apache Kafka provides high throughput, unified, and low latency platforms for real-time video feeds.

Till now, many surveys, studies, and methodologies are conducted by several authors for VoDS & LVS. For example, Panda *et al.* [3] analyzed capturing video frames from different sources like IP cameras, web cameras, mobile cameras performed using wireless ad hoc networks. Ichinose *et al.* [4] investigates how video analysis can be done using Kafka and Spark. The video frame was converted into the JSON object and the paper defined different configurations for the Kafka producer side and consumer side. Bouge *et al.* [5] analyzed how Kafka is helpful in big data analysis and Liu *et al.* [6] worked on the Internet of Things (IoT) devices for Kafka which helps in analysis video. Big data needs a lot of computational power to analyze fast, at the same time, video and image should not drop their quality. Hence, all Kafka and Spark configuration must be done at the client-side as well as the server-side. Ma *et al.* [7] showed performance evaluation for video and image services provided by different cloud service providers. Huang *et al.* [8] used NGINX for load balancing and Red5 media servers or Amazon web service (AWS) services like media elemental services. Red5 is an open-source media server that is designed in Java and provides different services like Wowza Streaming Engine, Adobe Flash Media, AWS Streaming and Wowza Streaming Support, HTTP Live Streaming (HLS), Flash, WebSockets, and RTSP so that video frames gets delivered in various devices. In mobile devices, there are several parameters like error handling, frame loss, and buffering which was handled by Muthuswamy *et al.* [9] using sliding window protocol and standard H.264 encoding technique.

To send a frame to the user's mobile, first, the server needs to know about user bandwidth, accordingly, the server sends relevant frame resolution frames to end-user mobile. Users can send requests from anywhere worldwide and hence, that frame requires time to reach an endpoint. In the VoDS, a delay in the frame is accepted but LVS requires each frame to be delivered to the user without any delay. Kim *et al.* [10] penned down that based on the usage of different CDNs, user experience i.e. QoE and video streaming cost may vary. All services for LVS & VoDS are on the cloud and hence, there are various techniques and user behaviors to decrease cost and increase QoE. Lee *et al.* [11] investigates video streaming based on different factors like user age, user gender, user watch timing, etc.

1.2. Motivation. CPS is a combination of systems with diverse nature with the purpose to control a physical process. It takes feedback and adapts to new conditions based on real-time input. It is a combination of physical processes, networking, and computation. The advancement of mobile communications and embedded technology has made it possible to build large scale CPS consisting of the interconnection of mobile phones. Sensing applications that are video-based equipped with camera-based mobile phones required to capture, send, and receive real-time videos for CPS is envisioned. Video-based CPS also faces a lot of challenges like abundant broadband network access, relevant content delivery, etc. Streaming of video is possible through either LVS or VoDS. Analyzing a video is required in CPS because if the system needs to be changed as per conditions being met in real-time, what is happening in the video after analyzing it. Since we need real-time services, video analysis should be very fast and Apache Kafka messaging services are need of the hour. It provides high throughput, unified, and low latency platforms for real-time video feeds. Hence, this would be beneficial for CPS.

1.3. Contributions. Through this paper, we present a review of video analysis over the LVS & VoDS video application. We compared different messaging brokers which help to deliver each frame in a distributed pipeline. Our primary focus is to analyze the impact on two message brokers for video analysis and how we can achieve LVS & VoDS using AWS elemental services. In this paper, we also analyzed the Kafka configuration parameter for reliability on full-service-Mode. Based on the above discussion, the following are the significant contributions of this paper:

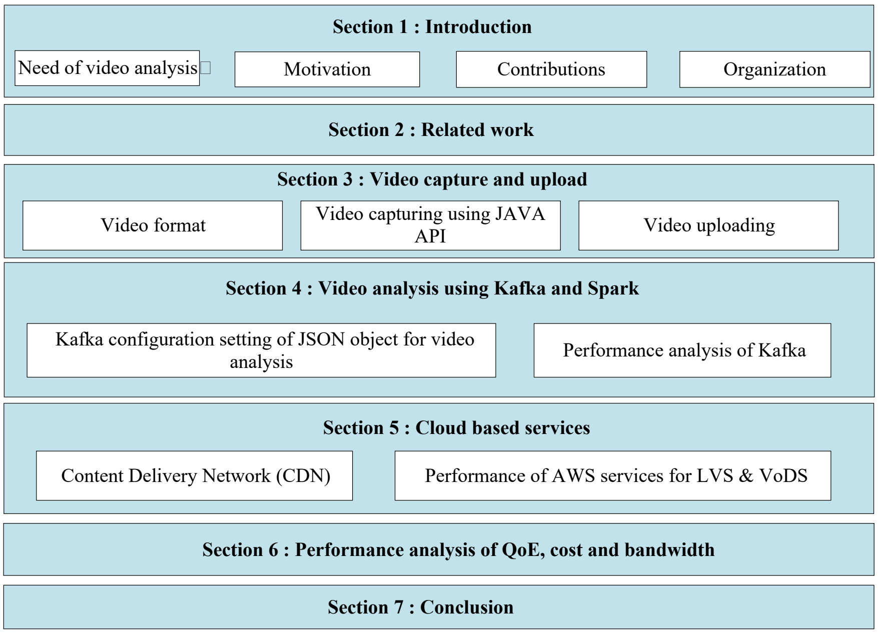

Section 1 : Introduction
Need of video analysis
Motivation
Contributions
Organization
Section 2 : Related work
Section 3 : Video capture and upload
Video format
Video capturing using JAVA API
Video uploading
Section 4 : Video analysis using Kafka and Spark
Kafka configuration setting of JSON object for video analysis
Performance analysis of Kafka
Section 5 : Cloud based services
Content Delivery Network (CDN)
Performance of AWS services for LVS & VoDS
Section 6 : Performance analysis of QoE, cost and bandwidth
Section 7 : Conclusion

Fig. 1.2. *Organization of the paper.*

- We present a scheme that helps to analyze video from the LVS & VoDS. It obtains how the frame can be divided into multiple OpenCV MAT objects and video analysis through Apache Kafka streaming and Spark.
- Video frame stored in the cloud that gets delivered to the end-user by using a content delivery network.
- The proposed scheme also finds out how to increase QoE and manage cost as well as bandwidth.
- The proposed scheme uses different cloud vendors or implementation like AWS, Google Cloud, Azure, and analyze which cloud vendor is best suited for what services and in which region.

1.4. Organization. The structure of the survey is as shown in Fig. 1.2. Table 1.1 lists all abbreviations used in the paper. The rest of the paper is organized as follows. Section 1 gives an introduction to CPS and its connection with video analysis. Section 2 examines related works conducted by several authors for LVS & VoDS. Section 3 presents insight into video capturing and uploading. Section 4 presents the video analysis using Kafka and Spark. Section 5 describes the implementation of different cloud services and corresponding results. Section 6 presents performance analysis concerning QoE, cost, and bandwidth.

2. Related Work. Several surveys and research work has been done in the area of LVS & VoDS. Ichinose *et al.* [4] analyzed real-time video data from multiple cameras and analyzed them using streaming engines and different machine learning and deep learning libraries. The authors used Kafka for streaming and Spark for streaming and analyze video frames. Real-time Transfer Control Protocol (RTCP) can be used to unicast and multicast which is described by Wang *et al.* [16]. The system is divided into two modules, namely, collection system and play system. Video data collection, H.264 encoding, RTP package, RTCP control, cache and preview, and real-time transmission is a part of the collection system. On the other side, the play system consists of data receiving, H.264 decoding, and RTP decoding. The authors used socket technology caching after video capture. The proposed system has strength over bandwidth for smooth video streaming for a stable environment whereas, in case of the fluctuating environment, the system can not predict actual bandwidth of end-user for smooth video streaming. Patel *et al.* [18] introduced how smooth video streaming can be achieved with fluctuating bandwidth. First, Available Bandwidth (ABW) is analyzed, then the average of that ABW is taken and at the end, according to the average ABW, streaming engine sends particular video frame resolutions.

Eduardo *et al.* [12] proposed a distributed architecture for LVS & VoDS which helps to stream large

Table 1.1
List of abbreviations

Abbreviation	**Description**
CPS	Cyber-Physical Systems
VoDS	Video On-Demand Streaming
LVS	Live Video Streaming
QoE	Quality of Experience
QoS	Quality of Service
CDN	Content Delivery Network
AWS	Amazon Web Services
HD	High quality
IoT	Internet of Things
HLS	HTTP Live Streaming
LUT	Look-up-table
AVC	H.264 Advanced Video Coding
HEVC	H.265 High Efficiency Video Coding
MP3	MPEG Audio Layer-3
AAC	Advanced Audio Coding
D-VoD	Distributed video on-demand
JD-VoD	Java distributed video on-demand
AC-3	Audio Coding 3
CDC	Change data capture
RTCP	Real-time Transfer Control Protocol
ABW	Available Bandwidth
AMQP	Advanced Message Queuing Protocol
IaaS	Infrastructure as a Service
ARQ	Automatic Repeat Request
FEC	Forward Error Correction
EC2	AWS Elastic Compute Cloud
S3	AWS Simple Storage Service
HDFS	Hadoop Distributed File System
RDD	Resilient Distributed Dataset
SSM	Spring SpringMVC Mybatis
MSK	Amazon Manages Streaming Kafka
RTMP	Real-time messaging protocol
AAC-LC	Advanced Audio Aoding - Low Complexity
HE-ACC	High Efficiency Advanced Audio Coding
RTT	Round Trip Time
RDBMS	Relational Database Management System
U_{rl}	Camera URL
C_{obj}	Camera object created by OpenCV
M_{rw}	MAT object number of rows
M_{cl}	MAT object number of columns
M_{ty}	MAT object type
J_{obj}	JSON object
T	Topic name
P_r	Kafka producer object
S_c	JSON schema structure for Spark
D_s	Create dataset from stream messages from Kafka
P_{data}	Process data
S_t	Group states
E_x	Existing states
P_p	Event processed data
E_{data}	MAT data from JSON
E_{key}	Camera key from JSON

videos on small networks also. In the video frame, duplicate frames could also be transferred. To overcome this problem, Wu *et al.* [21] used different delivery semantic and compared with other messaging systems like RabbitMQ, Advanced Message Queuing Protocol (AMQP), and Apache Flink. Apache Kafka works in a distributed environment and there are many cloud vendors available that provide Kafka as Infrastructure as a Service (IaaS). Wu *et al.* [22] compares Kafka performance on different cloud vendors. In Spark, data

Table 2.1

A relative comparison of state-of-the-art video analysis approaches

Author	**Approach**	**Objective**	**Year**	**Aplication**	**Pros**	**Cons**
Batista *et al.* [12]	D-VoD server and JD-VoD server	Distributed architecture for LVS & VoDS distribution	2005	Distributed architecture for LVS & VoDS distribution	Processing of large videos on small networks	Not scalable.
Yi *et al.* [13]	Look-up Table (LUT) technique.	Decoding technique by means of converting input video stream into bit stream.	2015	Format stream in Little-Endian Systems	3.49% faster decoding and 11.45% reduction in bit extraction time	Can only be applied to H.264 video application
Machida *et al.* [14]	Socio-ICT model	Timely detect abnormal events and swiftly deliver alerts to security agencies	2015	Resilient video surveillance service for the purpose of ensuring safety	System is more resilient	Costly and requires accurate gadget
Li *et al.* [15]	AWS services for transcoding	Transcode same video into different formats and decrease QoS violation of video streams	2016	Deliver multiple output videos via progressive download or adaptive bit rate	Major computational part resides at cloud and cost-efficient and QoS	Does not work on LVS
Wang *et al.* [16]	Mobile captured, media streaming engine	LVS system that is mobile based on streaming media technology	2016	Used for smaller sized organization	High transmission efficiency	Socket technology used for capturing and caching
Su *et al.* [17]	Femtocaching, VoDS, QoE.	Wireless traffic demand driven by VoD streaming and LVS	2017	Single user architecture	Playback duratio affected	Need a large scale edge location
Patel *et al.* [18]	Available bandwidth prediction, adaptive video streaming	Video streaming over fluctuating bandwidth	2017	Can be suitable for VoD	Good for wireless networks	Very complicated to decide which packet belongs to which base at client side
Kanrar *et al.* [19]	Viewer driven session based multi-user model	Bandwidth utilization	2017	Peer-to-peer network where bandwidth is consumed at every session	Services like pause, move slow, rewind, skip some number of frames	Works only on mesh networks
Bhole *et al.* [20]	Stream data, Apache Kafka, Cryptocurrency	Big data streaming for Cryptocurrency	2018	Tracking of cryptocurrency	Web based tracking modules	Locally tested
Liu *et al.* [6]	Spring SpringMVC Mybatis (SSM) + KAFKA	IoT middleware message service	2019	Used in IoT tracking devices	High throughput, distributed, fault tolerant	Lack of pace, issue with message tweaking
Wu *et al.* [21]	Streaming processing, Apache Kafka	Impact of all kinds of configuration parameter on the reliability of Kafka	2019	One-to -many communication streaming	Management on duplicate message control	Cluster-Service mode
Wu *et al.* [22]	Input configuration parameter	Queueing based packet flow model to predict performance metrics of Kafka	2019	D-stream data configuration parameter	Performance-based configuration	Configuration only for cloud vendors not for cluster-service mode
Tu *et al.* [23]	Map reducer, Spark streaming, big health data	Medical streaming data architecture for big data	2019	Hospital data analysis	Real-time data processing for daily health	MySQL vertical scaling or scale-up
Uddin*et al.* [24]	MMLSpark, OpenCV, HDFS	Distributed video analytic for intelligent video surveillance	2019	Several offline and online surveillance videos	Ensures scalability and fault tolerance	Security and privacy of cloud is not addressed
Mahapatra *et al.* [25]	Spark pipelines, IoT mashap tools, graphical tools, and streaming analysis flow base	Reduce complexity for pipeline data stream	2020	Graphics sorting	Spark pipeline D-stream flow	Re-usable model for persisting external file system
Shabrina *et al.* [26]	Using different cloud vendors	Improving of QoS	2020	Can be used at different edge locations with different cloud vendors	Throughput averaged to 3990.4 KB/S	Geo location may not be close to all target locations, adds complexity to your website for deployment procedures
Dongen *et al.* [27]	Apache Spark, structured streaming, Apache Flink, Apache Kafka, Kafka streams, distributed computing	Single pipeline for data streaming	2020	Large user scale	High throughput	Single Kafka broker
Anveshrithaa *et al.* [28]	Apache Spark, Long Short-Term Memory, Kafka	To predict traffic flow information	2020	Real-Time Vehicle Traffic Analysis	Reduces travel time, energy and cost	Hyperparameter optimization not been carried out

comes from across different platforms like MySQL, Kafka, Flink, RabbitMQ, etc. Tu *et al.* [20] analyzed how Spark architecture will perform with varying input stream data, for example, they have used health monitoring data for analyzing each input stream engine. Yi *et al.* [13] presented a decoding method for compressed video stream which is 3.59% faster than conventional method. Shabrina *et al.* [26] had used AWS service to analyze Quality of Service (QoS) which was proven cost-efficient for HLS. To improve the QoS of a video stream, CDN is the main factor for analysis. The authors in [26] analyzed the mechanism of HLS streaming related to CDN implementation. Su *et al.* [17] investigated the use of caching policy for VoD service demand. Caching policy helps in playback duration but it also requires a large scale edge location for world wild edge location. Kanrar *et al.* [19] presented a session-based multi-user model for bandwidth utilization that helps in a peer-to-peer network where bandwidth is required at any session but it works on only mesh network.

Liu *et al.* [6] proposed Kafka role in IoT based application. IoT services interconnect through many telecommunication networks and the Internet which connects IoT services point to point. There are many to many and one to one relationships between IoT services and hence, middleware plays a major role here which helps to send data to different services according to their relationship. The middleware services must be efficient and reliable message services and hence, Kafka is used as a middleware as a message broker. All IoT services are tested in multiple nodes Kafka. The Spring MVC has been used for Kafka middleware as Spring cloud provides microservices which makes each middleware service as independent service, so that, flexibility using different technology and scalability can be leveraged. Bouge *et al.* [5] analyzed Kafka in different computational parameters for big data streams. They prescribed four phases namely data collection, ingestion, processing, and storage. The main contribution of this paper was that ingestion performance can impact the overall stream processing. Video analysis is a major factor when it comes to smart city surveillance systems. Machida *et al.* [14] introduces a socio-ICT model which consists of a dynamic and queuing model. The platform enables dynamic load changes offloading video analysis. Social simulation to analyze the causal relationship between different parameters like arrival rate increases while the exit rate remains unchanged and the congestion level increases due to increased face drop rate. Concerning one video frame to be served in different resolutions, Li *et al.* [15] introduced transcoding for LVS & VoDS. The authors stored numerous versions of the same video in advance called pre-transcoding. Transcoding services also factorize for cost-efficient video transcoding using heterogeneous cloud services. Table 2.1 shows a comparison between existing approaches in video analysis techniques.

3. Video capture and upload. This section explores various video formats. We have considered the heterogeneity/complexity of different codec techniques used to perform video analysis. Table — depicts the comparative study of such techniques along with their pros and cons. Also, video capturing using JAVA API and video uploading using OpenCV is discussed in brief.

3.1. Video format. A video format consists of two parts. First is the format and codec being the second one. A format is a standard set of rules for storing containers, codecs, metadata, and folder structure. In the market, different format exists like MP4, AVI, WMV, etc. Each different format factor is arrangements of information within a container like a video stream, audio stream, metadata (bit rate, device resolution, subtitle, time of creation). Codec, which is part of video metadata, is a combination of CODER and DECODER. It encodes the video and audio stream making it smaller and easier to manage. At the end-user, the device decodes the video using a video player which also suggests codec. There are several video codec available in the market but mainly there are 3 codecs that almost all devices support like H.264 Advanced Video Coding (AVC), H.265 High-Efficiency Video Coding(HEVC) and VP9. The main audio codec is MPEG Audio Layer-3 (MP3), Advanced Audio Coding (AAC), and Audio Coding 3 (AC-3).

3.2. Video capturing using JAVA API. According to Panda *et al.* [3] Java API helps to capture videos from different sources like IP camera, WebCam, mobile video camera, etc. The API helps in encoding and compressing video and later using a multipart (It is a container that holds the whole body part of the capture video frame) streaming technique to transport the video to the Kafka streaming engine. The whole operation is divided into independent microservices.

Ad-hoc networks are used for capturing large scale video from different users. While the user capturing a video, the wireless link gets repeatedly broken and re-established again and again resulting in the loss of some

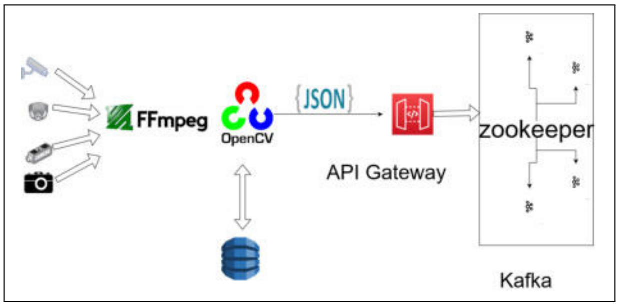


Fig. 3.1. *Video Capturing using OpenCV*

of the frames. The authors have used Automatic Repeat Request (ARQ) and Forward Error Correction (FEC) which is modified to reduce high transmission error. Sender side video is encoded by Java's encoder which is done at the client-side. The streaming is transferred to a particular route by different routing protocols. At the receiver side, the video streams are put in sequence after the video is decoded for different platforms. But, if we have a mobile device or any browser for display, it requires a lot of computational power for decoding, and hence, decoding work is done at the cloud. AWS Elastic Compute Cloud (EC2) helps to create a server that takes an encoded frame. AWS cloud helps us to decode the frame into different formats. Many different cloud services will help to compress the video to reduce the size of the video.

3.3. Video uploading. Video uploading tasks are done by the user or video streaming provider. For example, on Facebook, user uploads video whereas in Netflix, admin uploads video, and different subscribed users are endpoints, hence, it is termed as VoDS. The Video stream collector works with different user's video uploading at large scale OpenCV video processing libraries to help convert a video stream into frames. OpenCV stores all frames and images into MAT objects that the object is serialized to and can transfer to Kafka producer to deliver it to different consumers for Spark machine learning library for video analysis. The MAT object is transferred in JSON format with a user ID and camera ID as shown in Figure 3.1. After Spark operation is performed, all analyzed video frames are transferred to secure storage like AWS Simple Storage Service (S3) or Hadoop Distributed File System (HDFS).

4. Video Analysis using Kafka and Spark. It is very difficult to perform large-scale real-time data processing in the cloud because of the computational complexity while performing data analytics. Data within a distributed network transfers at a very fast rate. For video analysis, we have selected Kafka because it provides a fast, fault-tolerant, scalable messaging system and is often used in real-time data streaming architectures providing real-time analytics. Following are the merit reasons for using Kafka to perform the video analysis:

- Distributed environment with publisher and subscriber architecture.
- Provide up to seven-day storage facility.
- Receive messages in order with different partitions with the same topic (support strict order).
- Support multiple thread nodes for multiple consumer data deliveries.
- Commit log for retrieve data after a failure.
- High throughput.
- Massive scale data support.
- Wider use case of event-driven microservices, log store, streaming, event sourcing, change data capture (CDC), enterprise data pipeline.

Ichinose *et al.* [4] proposed a framework that measures the data analysis throughput which varies depending on the number of brokers and the number of nodes for consumers. Here, video analysis needs continuous transfer of large amounts of data for computation in the cloud. There are many deep learning frameworks like Cafe, TensorFlow, Chainer, etc. A sequential process is followed wherein the client receives data from Kafka with Spark, executes its machine learning library and python program, and later Chainer is called upon. Data is stored in the form of a Resilient Distributed Dataset (RDD) in Spark. Kafka configuration is divided into two subsections, first being single node and second being multiple node configuration as stated below:

- *Single Node:* In this, one thread is assigned to one customer. The number of machine cores used is 8 cores setting several consumers. Thus, the number of machine cores enables efficient processing.
- *Multiple Node:* Configuration of 1 to 2 consumer nodes means that throughput increases by 1.8 times but configuration of 2 to 4 consumer nodes, the improvement is by 1.4 times.

The requirements of Kafka's reliability relies on different use cases. In the case of video streaming, no duplicate messages should be delivered to the endpoint. For this use case, Kafka provides three delivery semantics. The first approach is at-most-once semantics where the producer sends messages continuously without waiting for any acknowledgment from consumers. Thus, in this case, the message delivered might be duplicate or message may be lost. The second semantics is at-least-once where the producer waits until an acknowledgment from the servers. Sometimes this approach might lead to duplicated messages. The third approach is exactly-once semantics which guarantees where all messages will be delivered without duplication. For our approach, we have used exactly-once semantics.

In this paper, Spark SQL is used for grouping the same camera ID frame for analysis. Algorithm 1 shows steps of video capturing and converting frames into JSON objects. In the algorithm, an input is a Camera URL and output is a JSON object. Prerequisite is one must have configured parameters for Kafka producer which is mentioned below and the external OpenCV library must be added in the Java Spring project. If one's source of the video is not a device camera, then one can add a video's actual path instead of URL. According to one's need, one can change URL parameters in the algorithms. These algorithms are designed for both LVS & VoDS.

4.1. Kafka configuration setting of JSON object for video analysis. Maintaining the order of messages in a single partition is mandatory while video analysis is performed using JSON objects. Kafka helps to maintain order by putting key value in the JSON object while producing data. To store large messages, some configuration needs to be changed on the server-side. Firstly, configure the Properties file message.max.bytes and secondly replica.fetch.max.bytes at server-side and for consumer side configuration is max.partion.fetch.byte and max.poll.records. These all configurations help to transfer JSON objects to the Spark for video analysis continuously. Some parameters should be taken care of while performing the configuration of Kafka for computing a large amount of data. The following parameters need to be set for the consumer side and producer side configuration.

1. Kafka configuration parameter for stream collector
   (a) kafka.asks = all
   (b) kafka.retries = 1
   (c) kafka.batch.size = 20971520
   (d) kafka.linger.ms = 5
   (e) kafka.compression.type = gzip
   (f) kafka.max.request.size = 2097152
2. Kafka configuration parameter for stream processor
   (a) kafka.max.partition.fetch.bytes = 2097152
   (b) kafka.max.poll.records = 500

At stream collector side, kafka.ask = all denotes when the producer gets acknowledgment from all in-sync replicas to the leader. kafka.retries = 1 indicates the number of retries if any broker goes down. kafka.batch.size means the total bytes of a message to collect before sending it to the producer. kafka.linger.ms signifies letting the producer know about the waiting time to particular ms in the expectation of more records for arrival. GZIP is a type of compress file that compresses the file to a smaller size and is best for faster network transfer. kafka.max.request.size denotes the total size of the record. A stream processor, kafka.max.partition.fetch.byte indicates the maximum amount of data fetch from per partition. kafka.max.poll.records signify letting the consumer know about the maximum number of records returned in a single consumer call for poll() function.

Figure 4.1 describes how to frame data to JSON object flows through Kafka and Spark with leveraged use of OpenCV for analysis. Some of the sources of video can be IoT devices also. Subscription system middleware is used to store IoT captured data into the database. Most IoT devices are required to implement a high-

Algorithm 1 Collect Stream

```
Input: Camera or File $\mathrm{U_{rl}}$
Output: Pass JSON object to Kafka stream using $\mathrm{P_r}$
Initialization: Create MAT object

procedure JSONCOLLECTIONSTREAM($\mathrm{U_{rl}}$)
    while Thread.Run() do
        if StringUtils.isNumeric($\mathrm{U_{rl}}$) then
            $C_{obj}$ = VideoCapture($\mathrm{U_{rl}}$)
        else
             Throw Exception
        end if
        if $\mathrm{C_{obj}}$.isOpen() then
            Thread.Sleep(500)
            if $\mathrm{C_{obj}}$.isOpen() then
                Throw Exception
            else
                while $\mathrm{C_{obj}}$.read(MAT) do
                    $\mathrm{M_{cl}}$=getCols(MAT)
                     $\mathrm{M_{rw}}$=getRow(MAT)
                    $\mathrm{M_{ty}}$=getType(MAT)
                    $\mathrm{J_{obj}}$=JSONOBJECT( $\mathrm{M_{cl}}$,$\mathrm{M_{rw}}$, $\mathrm{M_{ty}}$)
                    $\mathrm{P_r}$.send($\mathrm{J_{obj}}$,T) // Send Stream to KAFKA
                end while
            end if
        else
            while $\mathrm{C_{obj}}$.read(MAT) do
                $\mathrm{M_{cl}}$=getCols(MAT)
                 $\mathrm{M_{rw}}$=getRow(MAT)
                $\mathrm{M_{ty}}$=getType(MAT)
                $\mathrm{J_{obj}}$=JSONOBJECT( $\mathrm{M_{cl}}$,$\mathrm{M_{rw}}$, $\mathrm{M_{ty}}$)
                $\mathrm{P_r}$.send($\mathrm{J_{obj}}$,T) // Send Stream to KAFKA
            end while
        end if
    end while
end procedure
```

concurrency, low latency message system. Liu *et al.* [6] used SSM framework. Spring is an open-source Java framework that provides a lot of functionality for dealing with large scale data, for example, Spring-Kafka connector, Spring JPA, Spring Cloud, and Spring CloudStream. For middleware, the database needs a fast key-value in memory and hence, the Redis database is used to store captured data set with key-value pairs. Redis is an open-source, in-memory data structure store, handed as a database, message broker, and cache. Redis supports abstract data structures like bitmaps, strings, lists, sets, sorted sets, maps, etc.

Algorithm 2 shows how Kafka consumers fetch data and put streams into Spark stream engines. Spark uses MapFunction() function to treat each stream data individually. Spark helps to group stream data using the same Camera ID and makes state from the group and pass the state group data into the Event function. The Event() function is VideoFrameDetection in Algorithm 2. Instead of VideoFrameDetection, one can use the Event() function which analyzes each frame individually and writes its own algorithm for the Event() function. groupByKey() function creates a group using Camera ID from JSON stream. mapGroupWithState() function creates a state for the group and serves each data from each group for Event() function.

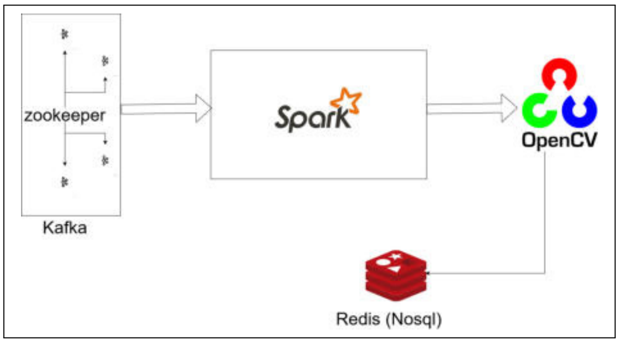


Fig. 4.1. *Kafka streaming and Spark engine*

Algorithm 2 Process Stream

Input: E_{data} will be input from kafka
Output: Video Analysis Operation
Initialization: SparkSession Object

```
procedure ProcessStream($E_{data}$,$E_{key}$)
    $S_c$=DataType.createStructType()
    while True do
        $D_s$=$D_s$.groupByKey(MapFunction($E_{data}$,$E_{key}$)
        $P_{data}$=$D_s$.mapGroupWithState($E_{data}$,$E_{key}$))
        if $S_t$.exist() then
            $E_x$=$S_t$.get()
        else
            $P_p$=VideoFrameDetection($E_x$,$K_e$,$V_v$)
            if $P_p$!=Null then
                Successful Operation
            else
                Error In Operation
            end if
        end if
    end while
end procedure
```

4.2. Performance analysis of Kafka. In video analysis architecture, it is evident that very high throughput matters for sending video frame stream into distributed networks. In this section, we will discuss Kafka's performance and test results that we encountered. Also, a comparison with other message brokers like RabbitMQ is represented here. Figure 4.2 shows the test results of different resolutions from data tested on Kafka and RabbitMQ. A total of 5379 frames are projected from a three-minute MKV video with different resolutions like 640x480, 720x480, and 1080x720. In Figure 4.2, the x-axis represented even number is Kafka's data and that with an odd number is RabbitMQ's data. You will see that graph of Kafka drop frame rate is quite negligible because Kafka is mainly designed for distributed publishers and subscriber architecture. In our case, Kafka broker is a multi-threaded node, and hence, in-case if any broker fails, then Kafka has supportive fault tolerance management services. Secondly, Kafka can highly maintain order from different partitions with the same topic with the key. Henceforth, the reason why RabbitMQ frame drop average rate is high because RabbitMQ is a single-threaded message queue.

RabbitMQ guarantees to deliver messages but not in order. RabbitMQ does not provide good enough fault tolerance service compared to Kafka because Kafka uses internal storage to store messages. We can store

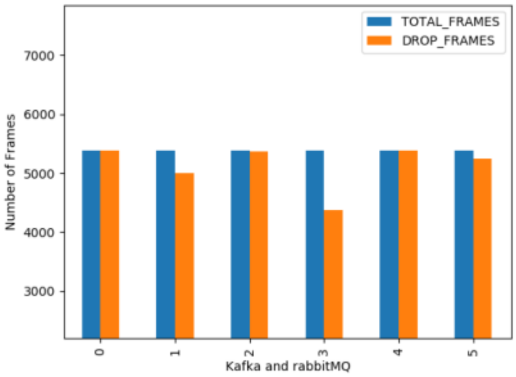


Fig. 4.2. *Analysis of Kafka and RabbitMQ*

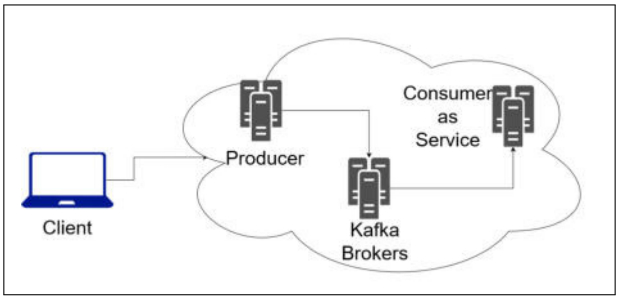


Fig. 5.1. *Full-Service-Mode*

a particular message for a maximum of 7 days using Kafka. In case of any failure, the Zookeeper maintains a pointer for the commit log. We can retrieve data with the help of the pointer. Whereas, in RabbitMQ, queue services discard data if data goes out of the queue. Single thread services in RabbitMQ proves to be a disadvantage for video analysis architecture as it architects need of a distributed environment for large scale services.

Kafka provides publisher and subscriber architecture, and hence, one can serve a stream to multiple consumers with the same order but the same cannot be done on RabbitMQ because, for each user, it needs to maintain queue services. If scaling broker > 3 becomes complicated, it can have a negative performance impact on RabbitMQ. In case, one wants to use RabbitMQ for streaming, then the usage of Redis services with RabbitMQ is mandatory.

5. Cloud based services. Many cloud service vendors provide Kafka as Infrastructure as a Service (IaaS), Wherein cloud vendors provide different resources for Kafka like CPU, storage, RAM over the Internet, which is on payment basis as per the use. The main advantage to move Kafka on the cloud turns out to be automatic scaling up and down of resources according to their requirements. Currently, many cloud vendors like Amazon Manages Streaming Kafka (MSK), Apache Kafka HDInsight provided by Microsoft Azure, Event Streams by IDM provide Kafka services. In this paper, we have used MSK for the Kafka cluster. Two different modes are provided by cloud service vendors. AWS offers a full-service mode where the Kafka cluster and producer resides at the cloud server and the consumer being another service of cloud as shown in Figure 5.1. Another mode is a cluster-service mode where Kafka broker resides at cloud service but producer and consumer reside at local devices like standalone workstations or mobile devices. We have inherited the full-service mode concerning our research work.

Cloud helps in various functions of video streaming like encoding and decoding, media streaming engine, CDN, etc. Five-part development involves a content server, reverse proxy server, load balancing server, streaming server, and storage for VoD clients on smart devices. The content server is responsible for managing video resources on the storage side like a media source database and serving that media source to the end-user in browser information. Load balancing and reverse proxy cache servers are responsible for processing requests faster. NGINX is a webserver which used as a reverse proxy server and load balancer. For streaming media, the Red5 server is used which is an open-source media server that provides LVS services over the cloud. AWS media elemental streaming service or Wowza streaming engine can also be used for the same purpose. In the end, cloud storage technology could use either HDFS or AWS S3.

Content servers were divided into two parts. The first is to give navigation information toward client devices' distributed server and second is to manage user resources and video information and upload file media. Also, for the best caching performance with NGINX, CDN can be used. Users can get a cached video from a nearby edge location. Continuous video streaming with standard encoding technique H.264 and sliding window protocol for continuous frame delivery in mobile is required. H.264 is a compression technology and media streaming engine that helps to convert video streams into H.264 format which is a globally accepted format for almost all devices. AWS media elemental services help to encode video into H.264, RTMP, and Apple HLS. The following explains the points to be taken care of in video streaming.

- *Buffer:* It helps to store incoming video frames and delivers the buffered frame without any delay. It helps in the continuous streaming of video. Buffers and timers are controlled by real-time messaging protocol (RTMP) players. RTMP contains its buffer that holds some beforehand frame in case of any network error and network failure. Thus, buffer helps in fault tolerance from frame loss.
- *Timer:* It helps to synchronize within media streaming servers and mobile devices. The time arrival of the video frame is solved using the same. Synchronization can also be done by timer while decoding.
- *Decoder:* It receives video from the timer section and converts the encoded video into a particular format that a receiver mobile supports. Decoder modules are in-built in the RTMP player.
- *RTMP player:* The player built using the android platform helps to perform LVS & VoDS. It contains a timer and a decoder section. It can help LVS by a buffer frame. HTML embed tag is used to embed URLs of media streaming servers and RTMP automatically starts playing from the initial default frame in the first queue manner. The streaming frame pointer data is stored in the in-memory database Redis. It is used for low latency streaming and requires no buffer.
- *Apple HLS:* This protocol uses adaptive bitrate streaming and supports almost all major streaming devices like browsers, android devices, and operating systems. For audio codecs, Apple HLS helps to stream Advanced Audio Coding - Low Complexity (AAC-LC), MPEG Audio Layer-3 (MP3), High-Efficiency Advanced Audio Coding (HE-ACC+V1&V2). For video codecs, it uses H.265 and H.264.
- *Bandwidth predictor:* But at first, the server requires bandwidth of user for frame resolution that should be done every particular second. It uses Round Trip Time (RTT) sent by the user device. At the client-side, the sliding window protocol is used to calculate the RTT.

5.1. Content delivery network. CDN helps to cache frames from a video and store all frames at a nearby location of the user that we call as an edged location server. The cloud vendors provide different QoE among varying users. Different regions have different scalability, cost, and cache hit rate. AWS CloudFront provides better quality of experiences than Google Cloud CDN for ASIA, AUSTRALIA, and NORTH AMERICA. Table 5.1 shows test results from different AWS CloudFront CDN edge location. It projects that South America has more frame delay around $>$ 5000 msec and hence, can also use another CDN provider to deliver the same content delivered by AWS.

Google Cloud CDN performed better than AWS in South America and Europe. It has been observed that if two users are using the same video streaming link at a time, a little latency is observed which leads to QoE drop. To conclude that QoE drop is not related to latency, different factors are responsible such as low-level computational power for encoding and decoding. If video streaming provides the worldwide area, then we can use different cloud CDN vendors for delivering the best user experience with video. Figure 5.2 shows AWS media services that help in resizing each frame into different resolutions and store it on S3. AWS CloudFront helps to deliver frames using CDN edge servers.

5.2. Performance of AWS services for LVS & VoDS. Figure 5.4 depicts VoDS with different frame resolution times required to get delivered to the cloud and time required to transcode that video frame into

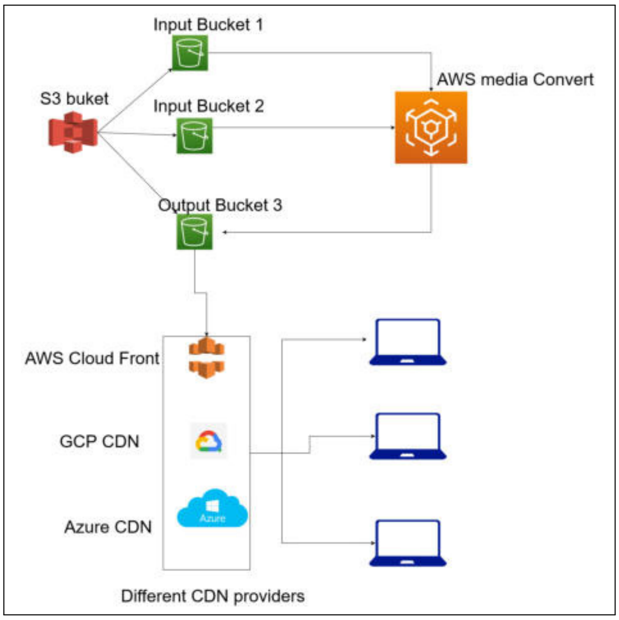


Fig. 5.2. *AWS S3 to different CDN*

Table 5.1
Different CDN edge locations

Edge Location	**Monitoring Time**	**Duration (msec)**	**Status**
Madrid	04/13/2020 06:49:50 PM	30359	S
Beijing	04/13/2020 06:49:51 PM	33262	S
Sydney	04/13/2020 06:49:50 PM	30136	S
Copenhagen	04/13/2020 06:49:50 PM	22224	S
Seattle	04/13/2020 06:49:49 PM	14610	S
San Francisco	04/13/2020 06:49:50 PM	79491	F
Mumbai	04/13/2020 06:49:50 PM	30813	S
Warsaw	04/13/2020 06:49:50 PM	28574	S
Paris	04/13/2020 06:49:50 PM	22874	S
Johannesburg	04/13/2020 06:49:51 PM	30971	S
Buenos Aires	04/13/2020 06:49:50 PM	30156	S
Shanghai			F(Timeout)
Amsterdam	04/13/2020 06:49:50 PM	30208	S
Dallas	04/13/2020 06:49:49 PM	8894	S
Brisbane	04/13/2020 06:49:50 PM	30176	S
Denver	04/13/2020 06:49:49 PM	10664	S
Frankfurt	04/13/2020 06:49:50 PM	22362	S
Montreal	04/13/2020 06:49:49 PM	7106	S
Hong Kong	04/13/2020 06:49:50 PM	30211	S
Tokyo	04/13/2020 06:49:50 PM	30142	S
N. Virginia	04/13/2020 06:49:49 PM	10609	S
Washington DC			F(Timeout)
Miami	04/13/2020 06:49:49 PM	17542	S
London	04/13/2020 06:49:50 PM	18354	S
New York	04/13/2020 06:49:49 PM	10354	S

different resolutions. Here, video frame are divided into two formats MP4 and HLS and both frames contain 4 different resolution videos 640x360, 960x540, 1280x720, 1920x1080. Figure 5.5 depicts LVS time taken by frame to get delivered to end-users. This live stream transcode are present in 5 different resolutions 640x360, 640x480, 720x480, 1080x720, 1920x1080. Note that we had used AWS Elemental Media Convert and AWS Elemental Media Live and AWS Elemental Media Package. Figure 5.3 represents whole system architecture.

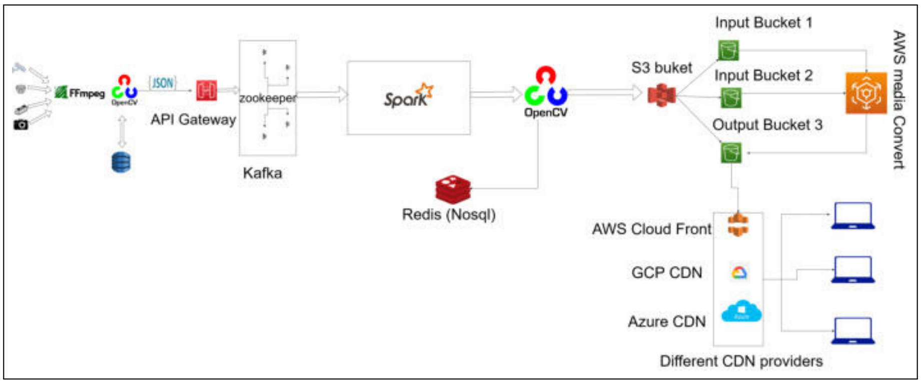


Fig. 5.3. *System Architecture*

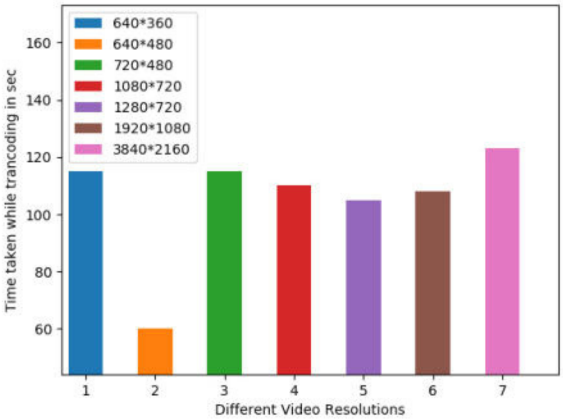


Fig. 5.4. *Time required to transcode VoDS video streaming using AWS Elemental MediaConvert*

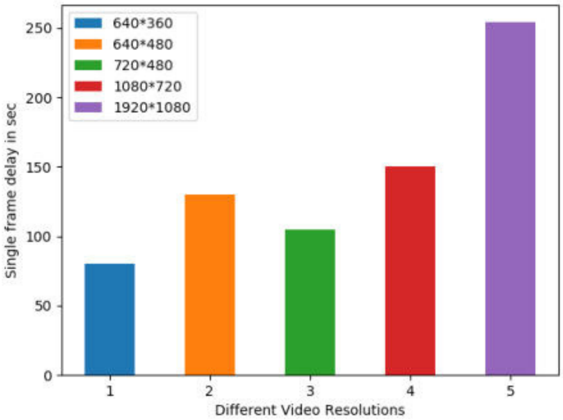


Fig. 5.5. *Time required to deliver single frame in LVS using AWS Elemental MediaLive*

We can see that time taken while transcoding LVS is less as compared to that of in VoDS. LVS is always more important than VoDS because one cannot miss out a frame in LVS due to live streaming. Hence, the QoE parameter is much improved.

6. Performance of Analysis of QoE, Cost and Bandwidth. According to CISCO visual network index, 80-90% of traffic in 2021 will be video-based. Cost, bandwidth, and QoE will be analyzed using different parameters. The parameter is age, viewing habits of the user, early leaving, steady user, viewing pattern of the different users at different times, viewing the content in different devices.

In this system architecture, we have placed a database for video analysis and user observation data for further QoS analysis. For example, we can store the user's different parameters which help to analyze the user's viewing pattern. The first parameter is "who is watching the video" points to the user profile. We can store user profiles that serve as the best video according to user viewing history. The second parameter is "when do users watch video" which means in the database we can store user viewing time with which category show they watch, for example, kids watch cartoon shows during the day, and this, we can suggest cartoon shows during the daytime. The third parameter is "which content is popular within an area". For this, we can acquire edge server location according to popular content being watched there. The fourth parameter is "status of the user for particular video" which denotes video summarization and helps to analyze which parts in video are most popular or user views most of the time. The last parameter is "user's devices", for example, video suggestions might have differences between TV and personal devices like mobile phones and laptops.

Data set for the above operation (D, H, M, S) divide data into 4 columns separated as Display, Hours, Minute, and Seconds. Data set can be created in MongoDB and Cassandra NoSQL database and a second table for profile information which can be stored into the Relational Database Management System (RDBMS). The user table should have 3 columns UserId, Gender, and Age. Spark for data and Zeppelins for data analysis can be used. We conclude that different users impact on QoE, bandwidth, and cost analysis.

To understand the impact of age effect on the server, let us take an example. Two major video content delivery platforms i.e. Netflix and YouTube have different region demands. For example, the Netflix series i.e. House of Cards is very popular in the USA's generation age group of 18-32, whereas, Sacred Games is popular in India's generation age group of 20-30. Thus, Netflix can cutoff costs by not uploading shows like Sacred Games and only upload shows contained among the demand region's edge location. This will decrease cost on the server and Netflix can distribute the load in a nearby location most of that time duration. One can also analyze viewing content as per gender group, to recognize the shows of Netflix that is popular in a particular gender.

Content centric viewing pattern is also different for videos. Some videos are popular among the kids and they are most active during the day time. Hence, we can make different microservices for kid's shows. This can be scaled up and down depending on most likely viewing time. Video browsing behavior also has several categories like watching numerous videos within an hour or watching only interesting part of a video or watching content for a short duration only. In the end, all these parameters help to create a digital marketing strategy for a particular show on a particular platform.

7. Conclusion. This paper presents the possibility of LVS & VoDS video analysis with the OpenCV library of Java webcam capture API to encode video. CPS is a combination of physical processes, networking, and computation and requires real-time captures and suggestions. The presence of CPS and its applications demands on-time video streaming with maintained high quality. Video streaming needs to be very quick from the camera to the cloud and from the cloud server to end-users for streaming video frames quickly. We have a comparative analysis of Kafka and RabbitMQ messaging services and found that drop rate in Kafka is very negligible as compared to RabbitMQ. Hence, we have used the Kafka streaming engine for sending video frames into JSON objects to different consumers. As video analysis needs a fast streaming data frame, therefore, Spark provides that fast streaming platform for analysis. For analysis, the Spark MLlib library is used. In the end, to store all video frames on to distributed storage, we have used AWS S3. To deliver video frames to end-users with the best experience, we should make use of a suitable content delivery network providing good results to users. This paper also devises different factors for QoE. We saw that time taken for transcoding frames in LVS was lesser as compared to that of VoDS and hence giving a better QoE to user. It makes the user to experience the video streaming without any breaks. Single architecture helps to achieve different tasks to analyze video frames. Video streaming is proven to be growing rapidly since a decade in the every field thus, quality parameters such as QoS, QoE, etc. play a major role for a user's convenience.

REFERENCES

[1] J. Lee, B. Bagheri, and H.-A. Kao, A cyber-physical systems architecture for industry 4.0-based manufacturing systems, *Manufacturing Letters*, vol. 3, pp. 18 – 23, 2015.

[2] Video streaming - statista, `https://www.statista.com/outlook/206/119/video-streaming--svod-/india`, accessed: 2020.

[3] A. Kathuria and S. N. Panda, Video capturing and streaming over ad-hoc networks, in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017, pp. 379–382.

[4] A. Ichinose, A. Takefusa, H. Nakada, and M. Oguchi, A study of a video analysis framework using kafka and spark streaming, in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 2396–2401.

[5] P. Le Noac'h, A. Costan, and L. Bougé, A performance evaluation of apache kafka in support of big data streaming applications, in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 4803–4806.

[6] Z. Yuan, B. Pang, Y. Du, X. Liu, J. Yao, and C. Kong, Design and implementation of internet of things message subscription system based on kafka, in *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)*, 2019, pp. 603–606.

[7] Y. Xue, H. Zhang, and H. Ma, Performance evaluation of image and video cloud services, in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2018, pp. 733–741.

[8] Z. Liu, Q. Wang, J. Huang, Y. Wu, Y. Wang, X. Jia, and H. Chen, Cloud-based video-on-demand services for smart tv, in *2017 Seventh International Conference on Information Science and Technology (ICIST)*, 2017, pp. 81–84.

[9] S. P. Tamizhselvi and V. Muthuswamy, Adaptive video streaming in mobile cloud computing, in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1–4.

[10] C. Wang, A. Jayaseelan, and H. Kim, Comparing cloud content delivery networks for adaptive video streaming, in *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, 2018, pp. 686–693.

[11] S. Rahman, H. Mun, H. Lee, Y. Lee, M. Tornatore, and B. Mukherjee, Insights from analysis of video streaming data to improve resource management, in *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*, 2018, pp. 1–3.

[12] C. E. C. F. Batista, T. L. Samilto, L. E. C. Leite, G. L. de Souza, and G. E. da Silveira, Big videos on small networks. a hierarchical and distributed architecture for a video on demand distribution service, in *2005 1st International Conference on Multimedia Services Access Networks, 2005. MSAN '05.*, 2005, pp. 15–19.

[13] K. Yi, Y. Lee, and J. Joo, A fast video decoding technique by means of converting input video stream into forward-oriented format stream in little-endian systems, in *2015 7th International Conference on Multimedia, Computer Graphics and Broadcasting (MulGraB)*, 2015, pp. 15–18.

[14] F. Machida, M. Fujiwaka, S. Koizumi, and D. Kimura, Optimizing resiliency of distributed video surveillance system for safer city, in *2015 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2015, pp. 17–20.

[15] X. Li, M. A. Salehi, and M. Bayoumi, High performance on-demand video transcoding using cloud services, in *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, 2016, pp. 600–603.

[16] J. Wang, W. Xu, and J. Wang, A study of live video streaming system for mobile devices, in *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)*, 2016, pp. 157–160.

[17] H. Su, S. Han, and C. Yang, Caching policy optimization for rate adaptive video streaming, in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016, pp. 713–717.

[18] K. Patel and G. Panchal, Smooth video streaming in bandwidth fluctuating environment, in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 79–83.

[19] S. Kanrar and N. K. Mandal, Approximation of bandwidth for the interactive operation in video on demand system, in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2017, pp. 343–346.

[20] B. R. Hiraman, C. Viresh M., and K. Abhijeet C., A study of apache kafka in big data stream processing, in *2018 International Conference on Information , Communication, Engineering and Technology (ICICET)*, 2018, pp. 1–3.

[21] H. Wu, Research proposal: Reliability evaluation of the apache kafka streaming system, in *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2019, pp. 112–113.

[22] H. Wu, Z. Shang, and K. Wolter, Performance prediction for the apache kafka messaging system, in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2019, pp. 154–161.

[23] Y. Tu, Y. Lu, G. Chen, J. Zhao, and F. Yi, Architecture design of distributed medical big data platform based on spark, in *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2019, pp. 682–685.

[24] M. A. Uddin, A. Alam, N. A. Tu, M. S. Islam, and Y.-K. Lee, Siat: A distributed video analytics framework for intelligent video surveillance, *Symmetry*, vol. 11, no. 7, 2019. [Online]. Available: `https://www.mdpi.com/2073-8994/11/7/911`

[25] Mahapatra, Tanmaya and Prehofer, Christian, Graphical flow-based spark programming, *Journal of Big Data*, vol. 7, no. 4, 2020.

[26] W. E. Shabrina, D. Wisaksono Sudiharto, E. Ariyanto, and M. A. Makky, The qos improvement using cdn for live video

streaming with hls, in *2020 International Conference on Smart Technology and Applications (ICoSTA)*, 2020, pp. 1–5.

[27] G. van Dongen and D. Van den Poel, Evaluation of stream processing frameworks, *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 8, pp. 1845–1858, 2020.

[28] S. Anveshrithaa and K. Lavanya, Real-time vehicle traffic analysis using long short term memory networks in apache spark, in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–5.

Edited by: Anand Nayyar
Received: May 15, 2020
Accepted: Jul 14, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 497–514, DOI 10.12694:/scpe.v21i3.1770

DESIGN AND MODELING OF RESOURCE-CONSTRAINED IOT BASED BODY AREA NETWORKS

SANIYA ZAHOOR AND ROOHIE NAAZ MIR*

Abstract. Due to the recent advancement and development of sensing, wireless, and communication technologies, there has been a shift in attention towards Body Area Networks (BANs). One of the most important services of BAN is the remote monitoring of patients, enabling doctors to observe, diagnose, and prescribe the patients without being physically present. Various vital signs are being monitored by body sensing devices installed inside, on or off the body of patients, but most of these devices are constrained in terms of resources such as storage, processing, bandwidth, and energy due to their smaller size. This paper aims at highlighting the key findings related to BAN applications, constrained resources, and various resource management techniques. The paper also presents the design and modeling of a resource-constrained BAN system and discusses the various scenarios of BAN in the context of resource constraints. It further proposes an Advanced Edge Clustering (AEC) approach to manage the resources such as energy, storage, and processing of BAN devices while performing real-time data capture of critical health parameters and detection of abnormal patterns. The comparison of the AEC approach is done with the Stable Election Protocol (SEP) through simulations and empirical data analysis. The results show an improvement in energy, processing time and storage requirements for the processing of data on BAN devices in AEC as compared to SEP.

Key words: Internet of Things, Edge Computing, Body Area Networks, Body Sensors

AMS subject classifications. 68M11

1. Introduction. Internet of Things (IoT) is a network of self-configuring objects (devices) that exchange data by interacting with the environment [1]. The main aim of IoT is to form a network of day-to-day life objects and make them programmable using wireless and sensor technologies, and pervasive connectivity. The pervasiveness of IoT eases everyday activities such as data exchange by devices while sensing, reacting to events in the application environments; one of the most important sectors is the healthcare industry [2][3]. With the advance of wireless and sensor technologies, there has been more penetration of IoT devices in IoT enabled healthcare applications. The Body Area Networks is one of IoT enabled healthcare applications that employs body sensors inside, on or off the body of patients for remote monitoring of patients [4] [5]. In BAN applications, the patients are being monitored by using IoT based wireless sensors that sense and transmit the data to the Personal Digital Assistant device (edge) for further processing and storage. Various vital signs are being monitored by body sensors while offering flexibility to patients to move. The edge raises the alarms in case of any abnormality in the physiological data.

The world of BAN is broad and multifaceted and one may even find it complex due to the plethora of applications that it encompasses. From non-medical applications to medical applications of BAN, the world is all set to undergo a shift in IoT healthcare. According to recent reports in Gartner, there will be an increase in 19% BAN devices by 2021, out of which 42% will constitute medical IoT devices and rest will constitute the non-medical IoT devices. Several smaller sized IoT devices are used in BAN for patient monitoring e.g., EMG, ECG, EEG, Fingertip Pulse Oximeter, Inertial Measurement Unit, Blood Pressure, Accelerometer, Temperature sensors, Body Humidity, etc [6]. Due to the smaller size of these devices, most of these devices are constrained in terms of resources such as processing, energy, storage, and bandwidth [7]. In addition to this, the BAN devices in such systems cannot be controlled dynamically leading to the transmission of unnecessary readings which causes further wastage of resources especially energy, storage, processing, and bandwidth.

Due to the limited resources in BAN, there are open challenges at different levels of hardware design and software development, as such research is being carried out in finding resource-efficient algorithms and protocols

[*] Department of Computer Science Engineering, National Institute of Technology Srinagar, India. (saniyazahoor@nitsri.net).

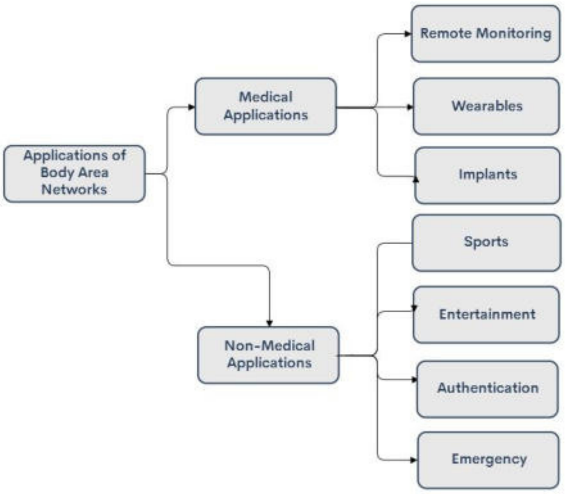


FIG. 2.1. *Applications of Body Area Networks*

that can store, process, and transfer the data with optimized resource management. There has been research in BAN resource management through data compression [8], data fusion [9], communication [10], network topology [11], machine learning techniques [12], and clustering approaches [13]. The main goal of all is to have seamless remote monitoring of patients by proper management of resources. However, none of the papers address resource constraints as a whole, which gives us the motivation to design, model, and implement a resource-efficient BAN system that could take into account more of the resource parameters.

This paper highlights the key findings related to BAN applications, constrained resources, and various resource management techniques. The paper also presents the design and modeling of a resource-constrained BAN system and discusses the various scenarios of BAN in the terms of resource constraints It further proposes an Advanced Edge Clustering approach for remote monitoring of patients with proper management of resources. In addition to this, an anomaly detection algorithm is also proposed that gives BAN the capability to detect false alarms. The comparison of the AEC approach is done with SEP [14], and the results are presented in the form of graphs and necessary explanation.

The organization of the paper is: Section 2 discusses literature survey, section 3 gives design and modeling of Body Area Networks, Section 4 discusses the evaluation of resource consumption in BAN scenarios, Section 5 presents proposed Advanced Edge Clustering approach, section 6 discusses simulation setup and evaluation and section 7 gives the conclusions.

2. Literature Survey.

Over the last few years, there has been research on Body Area Networks highlighting the key findings related to applications, resources and constraints of BAN, and various techniques to address these constraints [15][16] [17], as discussed below.

2.1. Applications of BAN.

From non-medical applications of BAN such as social networking, gaming, etc to medical applications of BAN such as body sensors for monitoring our health, the world is all set to undergo a shift in IoT healthcare. There are numerous applications of BAN which are categorized into medical and non-medical applications as shown in Figure 2.1.

2.1.1. Medical Applications.

BAN has wide applicability in medical fields. Based on the devices used in medical applications, we have categorized BAN devices into remote monitoring devices, wearable devices, and implant devices as discussed below.

- Remote monitoring BAN devices allow us to keep track of patient's health parameters and provide real-time feedback to the patients at home [18]. The intent of remote monitoring by BAN devices is to provide affordable remote healthcare to people at home ,and avoid hectic interactions and travel

Table 2.1
Classification of studies carried out in BAN applications

Category	Ref.	Focus	Types	Constraints
Medical Applications	[18]	Monitoring heart rate, body temperature, respiration rate, etc	Remote Monitoring	Communication Interference
	[19]	Monitoring of heart rate	Remote Monitoring	Communication Interference
	[20]	Exercise monitoring	Remote Monitoring	Delay
	[21]	Temperature and blood pressure monitoring (ASNET)	Remote Monitoring	Security
	[22]	Monitoring activities of soldiers in battlefield	Wearable	Security, Privacy and authentication
	[23]	Monitoring health of Athlete	Wearable	Storage and Processing, Motion Interference
	[24]	Sleep monitoring	Wearable	Storage and Processing
	[26]	Monitoring cardio-vascular diseases and other abnormalities of physical	Health Implants	Environmental challenges
		Cancer detection and tumor diagnosis by sensors	Implants	Environmental challenges
Non-Medical Applications	[27]	Body movement and temperature of trainee	Sports	Environmental Inferences, Security
	[26]	Gaming by body gestures, hand movements, etc	Entertainment	Environmental Factors
	[33]	Biometric parameters (Fingerprint, hand geometry, retina recognition, etc)	Authentication	Environmental Factors
	[34]	Fire alarms, notification management	Emergency	Heterogeneity

to healthcare institutions and hospitals. In addition to this, remote monitoring of patients helps in the continuous assistance of the patient's health condition by timely sending of data to the doctors. Therefore, doctors find it easy to follow up patients either by video conferencing or phone calls. Considerable work in this includes real-time exercise monitoring [19], monitoring health parameters such as temperature and blood pressure [20][21], etc.

- Wearable BAN devices are normally attached to body surface using straps. Considerable work in this includes activity monitoring of soldiers at battlefield [22], monitoring of athletes [23], BAN based sleep monitoring [24], etc.
- Some BAN devices are implanted inside the body e.g., visual implants for recovering retinal degeneration problems of a patient [25], while others are used to monitor various abnormalities of physical health e.g., cardiovascular diseases, cancer monitoring [26], etc.

2.1.2. Non-Medical Applications. The applications of BAN are not confined to medical applications only but also used in various non-medical applications [27]. With the advance in BAN technology, it is widely used in many non-medical application areas such as sports, entertainment, emergency, and authentication as discussed below.

- BAN is mainly used in the field of sports for motion recognition and physiological status detection, which helps sportspersons with the correction of postures and improvement of skills. Considerable work in this includes heart rate measurement system [28], BAN based kinematic analysis of swimming strokes [29], etc .
- BAN is also used in entertainment such as social networking, gaming, get together with friends, making phone and video calls, etc [26].
- BAN finds its applicability for authentication purposes such as face detection [30], fingerprint [31], iris recognition [32], etc [33].
- BAN devices are useful in the detection of household smoke, fire, poisonous gases, etc [34]. It plays an important role to make workplace and homes safer to stay in.

Table 2.1 shows the classification of studies carried out to evaluate the applications of BAN.

2.2. BAN Resources and Constraints. The Body Area Network typically consists of a collection of low-power, miniaturized, lightweight devices with wireless communication capabilities that operate in the

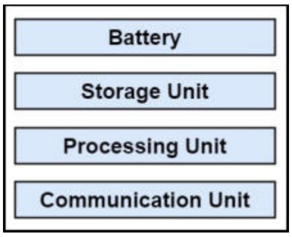


Fig. 2.2. *Resources in BAN Device*

proximity of a human body. An important aspect of these devices is its resources. As shown in Figure 2.2, a typical battery-operated BAN device possesses energy, storage, processing, and bandwidth as its resources [35]. Due to the smaller size of BAN devices, there are resource constraints in such applications. The resource constraints and related issues faced in Body Area Networks are discussed as [36]:

- Battery: Body sensing devices have limited energy which becomes one of the main constraints of BAN applications. Various solutions such as compression techniques sleep modes, low energy consumption hardware, etc are major design issues in BAN applications.
- Computation: Body sensing devices possess limited computing capabilities i.e., less processing power and storage which becomes challenging to store and process the data sensed by body sensors.
- Communication: BAN environments have limited communication and connectivity. As the number of body sensors increases, the number of transmissions increases which in turn demands more connectivity. Thus, limited communication and connectivity are major challenges in resource management.
- Security: In BAN applications, the transmitted data from the body sensors to the edge should be accessed only by authenticated entities. Any tampering in the information can cause serious issues such as a patient's death. As such, there is a need of securing this information, but implementing any security algorithm incurs resource overhead leading to more energy consumptions.
- Environmental Interference: There are many challenges faced by BAN applications due to environmental factors e.g., movement of the body can lead to a change in position of BAN devices, leading to path loss and incomplete reception of data in BAN applications, thus causing resource overheads.
- Heterogeneity: In BAN environments, there is heterogeneity in data types and measurements, leading to high complexity and resource overheads in such applications. Management of heterogeneity at the hardware and software levels is one of the major challenges in BAN applications.

2.3. Resource Management Techniques. Due to limited resources in BAN, there are open challenges at different levels of hardware design and software development, as such research is being carried out in finding resource-efficient algorithms and protocols that can store, process, and transfer the data as per application requirements and with optimized resource management. Most of the work has been carried out through data compression, data fusion, communication, network topology, machine learning techniques, and clustering approaches. The main goal of all is to have seamless remote monitoring of patients by proper management of resources.

2.3.1. Compression. BAN devices are mostly constrained in terms of energy, storage, processing, and communication. There have been recent studies in compression techniques that aim to reduce energy consumptions on BAN devices by reducing the data size to be transmitted from BAN devices to the sink (edge node) instead of transmitting to the sink directly. Considerable work in this include compression algorithm that aims to reduce the number of data transmissions and enhances real-time experience [37], Joint orthogonal matching pursuit that aims to reduce the number of transmissions, and storage and processing of data [8], compressive sensing that aim to reduce data rate [38], DFT method that aims to compress data [39], etc. There has been work on compression techniques in ECG monitoring as well and these include a compression algorithm based on WT that aims to reduce frame size and low delay [40], Quadratic compression algorithm that aims to reduce energy consumptions [41], etc.

Limitation: Most of the work in compression techniques focuses on optimizing energy utilization. No attention is paid towards other resources such as storage, processing, and bandwidth.

2.3.2. Data Fusion. In Body Area Networks, the body sensors sense several physiological data about the patient or elderly, and while sending the data from the sensors to edge, BAN face many challenges at the node and network levels. Data manipulation is one of the main challenges faced in BAN and to address this, various data fusion mechanisms are employed that combine data from multiple body sensors into more accurate information than an individual body sensor [42]. Considerable work in this includes data fusion techniques that achieve effective noise filtering and accurate inferences [9], data fusion mechanisms to reduce energy consumptions, minimize data redundancy and increase network lifetime [43], etc.

Limitation: Most of the work in data fusion techniques focuses on optimizing energy usage and little attention is paid towards storage, processing, and bandwidth as a resource.

2.3.3. Network Communication. In BAN applications, a typical node is characterized as a non-rechargeable and resource-constrained device. Due to the limited transmission range of these devices, an Edge-IoT network is formed wherein edge nodes are deployed hierarchically to process the real-time data in such applications. In many BAN applications such as implants, body sensors are deployed inside the human body, as such the electricity of the body affects the communications, that take place among the sensors or from sensors to edge. There is research going on sensor and communication technologies to address these issues such as radio technologies and other communication protocols in BAN applications [44], data communication to enhance Quality of Service and security [10], adaptive multi-hop routing protocol to improve network lifetime for multi-hop wireless body area network [45], medical implant communications service suitable for low data rate networks [46], etc.

Limitation: Most of the work in this focus on optimizing bandwidth,energy utilization and security. Little attention is paid towards storage and processing.

2.3.4. Topology. Deployment of BAN sensors is possible in several different topologies; the simplest topology is the single-hop star where every node communicates its sensed data to the sink node directly. The most commonly used topology in BAN is star topology [47][48]. This topology simplifies the network design but the disadvantage is that it is less robust and scalable. For larger networks, multi-hop routing is necessary, which depends on the placement or arrangement of body sensors in a network. The multi-hop routing drains less energy than direct routing but it increases the network delay, therefore, direct communication is better in scenarios where nodes are closer to edge nodes. Currently, research is going on multi-hop networks, most of the research focuses on energy acquisition, human body channels, etc to give better BAN services [11]. In BAN applications, the nodes can be deployed in a structured or randomized manner depending on the application requirement [49]. In structured deployment, nodes are placed at a fixed spot and routing paths are predetermined. In randomized deployment, nodes are scattered randomly. In most of the BAN applications, nodes are mobile which a major challenge in such environments becomes. An efficient node deployment scheme is needed to reduce the complexity caused due to mobility. Further, there are challenges in node deployment such as redundant data, energy consumption, delay, storage, coverage, etc that need to be taken care of in such applications.

Limitation: Most of the work in this focuses on optimizing energy utilization and delay. Little attention is paid towards other resources.

2.3.5. Machine Learning. Machine learning techniques are foreseen to transform healthcare by completing tasks with lesser delay and greater accuracy by using few resources of BAN devices. Machine learning techniques such as genetic algorithm [50], fuzzy logic [51], KNN [52], SVM [53], decision tree[54], neural network [55], etc are used for feature extraction and building decision models for prediction. The data analysis and prediction, done by machine learning algorithms, helps to make data-driven decisions, predict outcomes, and detect anomalies, which are useful for end-users. Considerable work in this includes leg motion classification with artificial neural networks [56] [57], human activity recognition system based on acceleration and vital sign data [58], real-time continuous glucose monitoring [59], etc.

Limitation: Most of the work in machine learning techniques focuses on optimizing delay. Little attention is paid towards optimizing resources.

Table 2.2
Summary of literature survey

Summary	Related Work	Benefits
Compression	Compression Algorithm [37]	Reduced Number of data transmissions
	Joint Orthogonal Matching Pursuit [8]	Reduced Number of data transmissions, Improved Storage and Processing of data
	Compressive Sensing [38]	Reduce Data Rate
	WT Compression Algorithm [39]	Reduce Frame Size Low Delay
	Quadratic Compression Algorithm [40]	Low Energy Consumptions
Data Fusion	Data Fusion in BAN [9]	Effective Noise Filtering, Accurate Inferences
	Data Fusion Mechanisms [43]	Reduced Energy Consumptions, Increased Network Lifetime, Low Data Redundancy
Network Communication	Data communication [45]	Enhanced Quality of Service Security
	Medical Implant Service [46]	Low data rate networks
Topology	Star BAN [47][48]	Less delay
	Multi-hop BAN [11]	Low energy consumption
Machine Learning	Leg motion classification with artificial neural networks [56]	Less delay
	Human activity recognition based on acceleration and vital sign data [58]	Less delay, Better Accuracy
Clustering	LEACH [61], SEP[14], etc	Improves energy utilization

2.3.6. Clustering Approaches. Clustering approaches are considered as one of the effective mechanisms to manage resources in resource-constrained IoT applications. In clustering, the IoT network is divided into clusters with each having a cluster head (CH) to reduce the consumption of resources. Considerable work in this includes data aggregation guaranteeing low communication and storage costs in IoT [60], a cross-layer data aggregation scheme increasing energy efficiency [13], etc. There has been a lot of work in clustering on data aggregation protocols for Body Area Networks such as Low-Energy Adaptive Clustering Hierarchy (LEACH) [61], Anybody [62], SEP [14], etc. LEACH is the simple clustering protocol in which nodes elect cluster Head based on a pre-defined probability while the other nodes join the closest cluster head [61]. Another data gathering protocol is 'Anybody' that uses clustering to reduce the number of direct transmissions to remote base stations [62]. There has been a lot of work in BAN resource management via SEP [14]. In SEP, a fraction 'm' advanced nodes in a total of 'n' nodes are provided with an additional energy factor, thereby increasing the stability period due to advanced nodes. The disadvantage is since advanced nodes become more frequently CHs, the energy of advanced nodes becomes less than the normal nodes. To overcome this, many modified versions of SEP have been introduced to achieve resource savings than the traditional one to some extent [63][64].

Limitation: There has been work in clustering approaches in IoT but most of the work focuses on energy. No attention has been paid towards other resources.

Table 2.2 gives the summary of above discussed research papers in terms of benefits that the BAN applications get. And Table 2.3 gives the summary of above discussed papers in terms of resource parameters that these papers have addressed. It is evident that none of the papers have focused on addressing resource constraints as a whole and therefore, it gives us the motivation to design, model, and implement a resource-efficient BAN system that could take into account more of the resource parameters.

3. Design and Modeling of Body Area Networks. Consider a BAN system where a patient or elderly wears IoT based body sensing nodes (devices), that sense medical parameters such as temperature, blood pressure, sugar level, etc to a central edge that integrates patient's medical data. It then transfers the data to the cloud or backend servers for related diagnosis. Figure 3.1 shows a general BAN system consisting of three levels viz. sensor level, edge level, and the cloud level.

At the bottom level, there are many sensors such as ECG, EEG, EMG, breathing sensor, etc that measure the physiological parameters of a patient or elderly at home. These sensors generate data combined on the Personal Assistance Device (edge level) to generate information. The information on these devices is pushed on the third level i.e. cloud for further processing and analysis to generate more information. For example, the medical data of patients on the edge devices can be merged to create a community of patients giving more

Table 2.3
Summary of literature survey in terms of Resource Parameters

Ref.	Number of Transmissions	Storage	Processing	Delay	Energy	Quality of Services	Security
[37]	✓						
[8]	✓	✓	✓				
[38]	✓						
[39]	✓			✓			
[40]					✓		
[9]						✓	
[43]					✓		
[45]						✓	✓
[46]	✓						
[47]							
[48]				✓			
[11]					✓		
[57]				✓			
[58]				✓			
[61]					✓		
[14]					✓		

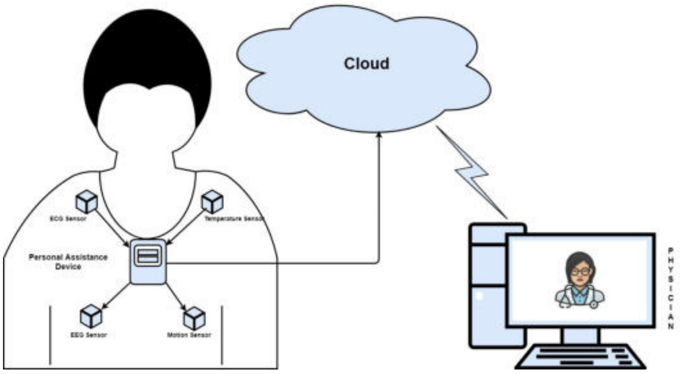


Fig. 3.1. *A General BAN System*

information about the patient's status in the entire city.

3.1. BAN Entities. In this section, we discuss the basic BAN entities such as body sensing nodes, edge and cluster.

3.1.1. Body sensing Node. A BAN network consists body sensing nodes, defined as a four tuples, $B_i = \{B_{id}, B_{st}, d_i, R_i\}$, where

- B_{id} represents the unique ID of body sensing node, B_i,
- B_{st} represents the status of body sensing node, B_i,
- d_i indicates the data items that a node senses,
- B_{Ri} represents the resource status of a body sensing node, B_i.

3.1.2. Edge. An edge node, E_i in BAN application is defined as two tuples, $E_j = \{E_{id}, R(E_j)\}$, where,

- E_{id} is the ID of edge, and
- $R(E_j)$ gives the resource status of the edge.

3.1.3. Cluster Formation. A cluster, C_i corresponds to a logical boundary comprising of several localized body sensing nodes, represented as three tuples, $C_i = \{C_{id}, B[U], E_{id}\}$, where,

- C_{id} is the ID of the cluster,
- $B[U]$ is the non-empty ID array of size U that stores IDs of all corresponding body sensing nodes,
- U is a dynamic value that is decided by the number of body sensing nodes present in a particular cluster, and
- E_{id} is the ID of the edge.

3.2. Data Model. The body sensing nodes,B_i are dedicated to monitor a patient (P_1), which transfers the data to the edge (E_j). To ensure secure communication in BAN applications, the body sensing nodes and personal edge will have unique IDs D_{idn} and E_{id} respectively. The state of body sensing node is denoted by a Boolean $B_{st} = \{0, 1\}$, where the values 0 and 1 symbolize the inactive and active states respectively. When the state of body sensing node is active, it will be able to sense and monitor the health parameters about a patient in the form of data items. These data items together constitute the patient's medical data.

At body sensor level, we have:

$$B_i \to d_i \quad Where 1 <= i <= n \tag{3.1}$$

At BAN level, we have:

$$\{B_1, B_2, B_3, ..., B_n\} \to \{d_1, d_2, d_3, ..., d_n\} \tag{3.2}$$

3.3. Resource Model. In BAN, a workload encompasses resources such as storage, processing, energy and network bandwidth requirements. It becomes essential to monitor the resources for their better allocation to accommodate this application workload. For each body sensing device B_i, the resources are denoted as:

$$R_i = \{E_i, P_{ri}, S_i, B_{wi}\} \tag{3.3}$$

where E_i represents energy, B_{wi} represents bandwidth,S_i represents storage, and P_{ri} represents processing of a body sensing node.

If $\{R_1, R_2, R_3, ..., R_n\}$ denotes the status of monitored resources of body sensing nodes $(B_1, B_2, B_3, ..., B_n)$ such as energy $(E_1, E_2, E_3, ..., E_n)$, bandwidth $(B_{w1}, B_{w2}, B_{w3}, ..., B_{wn})$, processing$(P_{r1}, P_{r2}, P_{r3}, ..., P_{rn})$ and storage $(S_1, S_2,S_3, ..., S_n)$ respectively. Then at the system level, the resources can be represented as:

Total energy, E_{P1} is given as:

$$E_{P1} = \sum_{i=1}^{n} E_i \tag{3.4}$$

Total processing power, P_P1 is given as:

$$P_{P1} = \sum_{i=1}^{n} P_{ri} \tag{3.5}$$

Total bandwidth, B_{P1} is given as:

$$B_{P1} = \sum_{i=1}^{n} B_{wi} \tag{3.6}$$

Total storage, S_{P1} is given as:

$$S_{P1} = \sum_{i=1}^{n} S_i \tag{3.7}$$

The total resources (R_{P1}) of BAN is given as:

$$R_{P1} = \{E_{P1}, E_{P1}, B_{P1}, S_{P1}\} \tag{3.8}$$

3.4. Allocation Model. At node level, the value of each resource is given as:

$$E_i = W_i; P_{ri} = X_i; S_i = Y_i; B_{wi} = Z_i; \tag{3.9}$$

where W_i, X_i, Y_i, Z_i denotes the maximum value for energy, processing, storage and bandwidth respectively. Once the workload is accommodated by proper allocation of resources, the value of resources becomes:

$$\begin{aligned} E_i &= W_i - w_i; \\ P_{ri} &= X_i - x_i; \\ S_i &= Y_i - y_i; \\ B_{wi} &= Z_i - z_i; \end{aligned} \tag{3.10}$$

where w_i, x_i, y_i, z_i denotes the consumed values for energy, processing, storage and bandwidth respectively. Considering the constraints of BAN, the resource allocation can be taken as an optimization problem that maximizes the resource availability. Therefore, at the node level, we have:

$$\text{Maximize } E_i P_{ri} S_i B_{wi} \tag{3.11}$$

Subject to constraints

$$\begin{aligned} E_i &<= W_i; \\ P_{ri} &<= X_i; \\ S_i &<= Y_i; \\ B_{wi} &<= Z_i \end{aligned} \tag{3.12}$$

The above BAN model can be extended to community level which can result in the creation of BAN grids, favorable in smart city IoT applications.

4. Evaluation of resource consumption in BAN Scenarios. The possible scenarios which arise in BAN applications, depending on whether the BAN entities are stationary or mobile, include stationary body sensing nodes and edge, stationary body sensing nodes and mobile edge, mobile body sensing nodes and stationary edge, and body sensing nodes and edge.

4.1. Stationary body sensing nodes and edge. In BAN applications with stationary body sensing and edge nodes, the location of all nodes is pre-determined. Let $\{B_{11}, B_{12}, , B_{1i}, B_{21}, B_{22}, , B_{2j}, , B_{n1}, B_{n2}, , B_{nk}\}$ be the set of stationary body sensing nodes and $E_1, E_2, ..., E_j$ be the set of stationary edge nodes such that a fixed number of body sensing nodes are associated to a particular edge node in each cluster i.e;

$$\begin{aligned} \{B_{11}, B_{12}, ..., B_{1i}\} &\in E_1, \\ \{B_{21}, B_{22}, ..., B_{2j}\} &\in E_2, \\ &..., \\ \{B_{n1}, B_{n2}, ..., B_{nk}\} &\in E_j \end{aligned} \tag{4.1}$$

Let $\{(x_{11}, y_{11}), (x_{12}, y_{12}), (x_{13}, y_{13}), ..., (x_{1i}, y_{1i}), (x_{21}, y_{21}), (x_{22}, y_{22}), (x_{23}, y_{23}), ..., (x_{2j} \ , y_{2j}), , (x_{1i}, y_{1i}), ...$ $(x_{n1}, y_{n1}), (x_{n2}, y_{n2}), (x_{n3}, y_{n3}), ..., (x_{nK}, y_{nK})\}$ denotes the positions of body sensing nodes $\{B_{11}, B_{12} ..., B_{1i},$ $B_{21}, B_{22}, ..., B_{2j}, ..., B_{n1}, B_{n2} ..., B_{nk}\}$ respectively and $\{(x_{e1}, y_{e1}), (x_{e2}, y_{e2}), (x_{e3}, y_{e3}), ..., (x_{em}, y_{em})\}$ denotes the positions of edge nodes $\{E_1, E_2, ..., E_m\}$ respectively. For a particular cluster, say C_1, the distance between the body sensing nodes, B_{11} and B_{12} is calculated as:

$$d_{11} = \sqrt{(x_{11} - x_{12})^2 + (y_{11} - y_{12})^2} \tag{4.2}$$

The distance between the body sensing nodes, B_{11} and edge node, E_1 is calculated as:

$$D_{11} = \sqrt{(x_{11} - x_{e1})^2 + (y_{11} - y_{e1})^2} \tag{4.3}$$

Since all the nodes are stationary, the distance between the nodes is definite. At the time of deployment, the size of cluster is stationary i.e., known number of body sensing nodes is associated with a particular edge. The body sensing nodes sense and generate data items $\{d_{i1}, d_{i2}, ..., d_{in}\}$ which is aggregated on the edge, E$_j$, forming a cluster, which is later pushed to the cloud.

For each cluster, C$_i$ we have $\{R(B_{i1}), R(B_{i2}), R(B_{i3}), ..., R(B_{in})\}$ that denotes the status of resources such as energy $\{E_{ei1}, E_{ei2}, E_{ei3}, ..., E_{ein}\}$, bandwidth $\{B_{wi1}, B_{wi2}, B_{wi3}, ... , B_{win}\}$, processing $\{P_{i1}, P_{i2}, P_{i3}, ..., P_{in}\}$ and storage $\{S_{i1}, S_{i2}, S_{i3}, ..., S_{in}\}$ for body sensing nodes $\{B_{i1}, B_{i2}, B_{i3}, ..., B_{in}\}$ respectively. And in each cluster, C$_i$ the resources of edge node, R(E_j) is always higher than that of body sensing nodes $\{R(B_{i1}), R(B_{i2}), R(B_{i3}), \dots, R(B_{in})\}$, i.e.

$$R(E_j) >> R(B_{i1}) \text{ or } R(B_{i2}) \text{ or } R(B_{i3}) \text{ or } \dots \text{ or } R(B_{in}) \tag{4.4}$$

Also, the resources of edge node are sufficient for associated body sensing nodes, i.e;

$$R(E_j) = R(B_{i1}) + R(B_{i2}) + R(B_{i3}) + \cdots + R(B_{in}) \tag{4.5}$$

Considering the resource constraints at the node level, the resource allocation model manages these resources in an optimized way. In each cluster, the edge will monitor the resources of body sensing nodes in such a way that it maximizes the availability of energy, processing power, memory, and bandwidth. Resource consumption occurs mainly due to the processing and transmission of data as per the application requirements. Such BAN applications are less resource-constrained because there is only one-time processing involved in calculating parameters such as the position of nodes, the distance of nodes, etc.

4.2. Stationary body sensing nodes and mobile edge. In BAN applications with stationary body sensing nodes and mobile edge nodes, only body sensing nodes have fixed positions. In such scenario, a fixed number of body sensing nodes are associated to any edge at a particular instance of time is given as:

$$\begin{aligned} B_{11}, ..., B_{1i} &\in E_1 \text{ or } E_2 \text{ or } \dots \text{ or } E_m \\ B_{21} ..., B_{2j} &\in E_1 \text{ or } E_2 \text{ or } \dots \text{ or } E_m \\ &..., \\ B_{n1}, ..., B_{nk} &\in E_1 \text{ or } E_2 \text{ or } \dots \text{ or } E_j \end{aligned} \tag{4.6}$$

Initially, an edge has a fixed location and a fixed number of body sensing nodes are associated with an edge node in a cluster. However, when edge nodes change their locations, the same fixed-sized cluster of body sensing nodes are now assigned to a new edge node depending upon the nearness of distance between an edge node and cluster of body sensing nodes i.e., the edge node closest to the body sensing node will act as the sink for the sensed data. Also in a particular cluster, the distance among the body sensing nodes is always fixed but the distance between the body sensing nodes and edge node needs to be re-calculated every time a new edge node becomes a sink of these body sensing nodes. In each cluster, C_i, body sensing nodes sense and generate data items that create mini-profiles of body sensing node on a neighboring edge. These mini-profiles are aggregated to form a profile, which then can be pushed to the cloud for further storage and analysis. The resource allocation model optimizes the use of resources in such BAN networks. Resource consumption occurs at a higher rate as compared to the above scenario because there are frequent calculations involved in determining the position and distance of mobile edge nodes.

4.3. Mobile body sensing nodes and stationary edge. In BAN applications with mobile body sensing nodes and stationary edge nodes, the resource consumption occurs at a higher rate as compared to above scenarios. This is because of the mobile behavior of body sensing nodes that drains the resources more quickly.

In such scenarios where there are mobile body sensing nodes but stationary edge, the association of body sensing node to the edge nodes is denoted as:

$$\begin{aligned}\{B_{11},...,B_{1i},B_{21},...,B_{2j} \text{ or } ...,B_{n1},...,B_{nk}\} \in E_1,\\ \{B_{11},...,B_{1i},...,B_{21},...,B_{2j} \text{ or } ....,B_{n1},...,B_{nk}\} \in E_2,\\ ...,\\ \{B_{11},...,B_{1i},...,B_{21},,B_{2j} \text{ or } ....,B_{n1},,B_{nk}\} \in E_j\end{aligned} \tag{4.7}$$

i.e., mobile body sensing nodes can be assigned to a particular edge. For each cluster, C_i, the distance between sensors and edge will vary and the distance among body sensing nodes will also vary. Initially, a known number of body sensing nodes are associated with a stationary edge node. But as the body sensing nodes change their locations, any random set of body sensing nodes are assigned to an edge node depending upon the nearness of distance between an edge node and body sensing nodes. The size of the cluster is not predefined and each cluster can either shrink or expand depending on the movement of body sensing nodes. For example, initially $C_1 = \{B_{11}, B_{12}, B_{15}\}$ and $C_2 = \{B_{21}, B_{25}, B_{35}\}$, after few round, $C_1 = B_{11}, B_{12}$ and $C_2 = \{B_{15}, B_{21}, B_{25}, B_{35}\}$.

In such BAN applications where body sensing nodes are moving and edge nodes are stationary, the resource consumption occurs at a higher rate compared to the above two scenarios. Considering such resource constraints at the node level, the resource allocation model manages resources in an optimized way.

4.4. Mobile body sensing nodes and edge. In applications having mobile nodes, the association of body sensing node to the edge nodes is denoted as:

$$B_{11} \text{ or } ... \text{ or } B_{1i} \text{ or } B_{21} \text{ or } ... \text{ or } B_{2j} \text{ or } ... \text{ or } B_{n1} \text{ or } B_{n2} \text{ or } ... \text{ or } B_{nk} \in E_1 \text{ or } E_2 \text{ or } ... \text{ or } E_j \tag{4.8}$$

i.e., any body sensing node can be assigned to any edge node depending upon the nearness of distance between the body sensing node and an edge node. Since nodes are randomly moving, temporary clusters are formed. For each temporary cluster, C_i, the distance between the body sensing node and edge node, and the among the body sensing nodes vary. In these applications, the resource consumption is highest. Thus, it is necessary to monitor these resources for efficient resource allocation that leads to better management of resources in such a way that it maximizes the availability of energy, processing, memory, and bandwidth on body sensing nodes. In this scenario, resource consumption is highest due to the high randomness of nodes.

5. Proposed Advanced Edge Clustering. In the proposed Advanced Edge Clustering approach, an advanced edge is employed that takes into account not only energy but storage and processing as well. The properties of the advanced edge are as under:

- The energy of the edge node is higher than the body sensing nodes such that it lasts till the last node dies,
- The processing power and storage of the edge is higher than the body sensing nodes,
- The distance of the body sensing nodes from the edge is lesser as compared to the distance from edge to the cloud,

In AEC, an advanced edge node, which is a pre-defined node, is used for each cluster. This superior node has higher resources such as processing, energy, and storage than body sensing nodes (as shown in eq. 4.4) and in each cluster, an edge node has sufficient resources to accommodate the workload of sensing nodes (as shown in eq. 4.5). The sensed data is collected by the edge nodes. And as the size of sensed data increases, resources of the edge node become insufficient and therefore data is pushed to the cloud for storage and processing as and when needed.

If the number of BAN applications increases in number, more edge nodes offload their computation to the cloud if physical edge nodes are not sufficient to accommodate the workload incurred by such applications and in that case, we have:

$$\{E_1, E_2, E_3, , E_k\} \rightarrow C_L \tag{5.1}$$

Table 6.1
Simulation Parameters

Parameters	Value
Setup Area (Network Size)	500 * 500 m^2
Number of Body Sensing Nodes for each Patient	4
Number of Edge	1
Initial Energy of Body Sensing Node	200 Joules
Initial Energy of Edge	1000 Joules
Storage of Body Sensing Node	30KB
Storage of Edge	1.5MB
Distance between Body Sensing Node and Edge	10m
Packet Size	512bytes

Besides monitoring the health parameters, the AEC approach uses an anomaly detection algorithm that raises alarm for the emergency team when any abnormal pattern is detected. It seeks to detect abnormal values to reduce false alarms resulting from faulty measurements while differentiating faults from patient health degradation (see Algorithm 1).

Algorithm 1 Proposed Algorithm

```
For each received record x do
Classify x using svm
If Class(x) == "ABNORMAL" then
Find out number of violations (V)
If V > 1: then
Call emergency procedure
else
Predict using regression
If error < threshold: then
Call emergency procedure
else
Declare false alarm
```

If the difference between the current value and the estimated value is larger than the pre-defined threshold for only one attribute, the measurement is considered faulty and is considered as a false alarm. However, if the readings are higher than the threshold, a medical alarm is triggered for the emergency team to react.

6. Simulation set-up and Evaluation. The simulation results are based on the proposed Advanced Edge Clustering approach as discussed in section 5. The simulation parameters are listed in Table 6.1.

The proposed approach has been tested on the Arduino board with body sensors (such as temperature sensor, breathing rate, heart rate sensor, blood pressure sensors, and movement sensor). To facilitate easy access to the microcontrollers, an Arduino integrated development environment is used and for the programming of microcontrollers, Python is used. We have also tested the proposed approach on dataset taken from the Physionet database consisting of three attributes viz., ABPsys, ABPdias, and HR to detect any anomaly in critical health parameters. The variations in physiological parameters are easily traced by the proposed anomaly detection algorithm.

The data has been captured over a period of time using the above setup and we have evaluated the proposed approach on this data as well. Since the proposed BAN system forms a heterogenous environment, therefore, the comparison of our proposed system is done with SEP [14], which is a standard protocol meant for a heterogeneous environment. The performance parameters used to evaluate the resource efficiency of the AEC approach include energy, processing time, and storage.

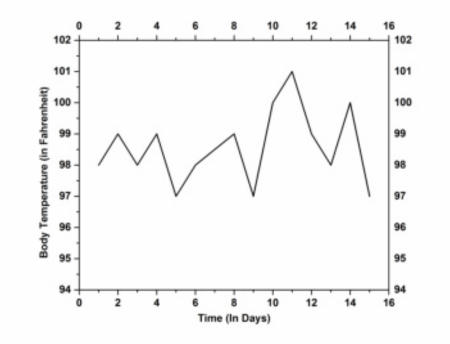


Fig. 6.1. *Body Temperature*

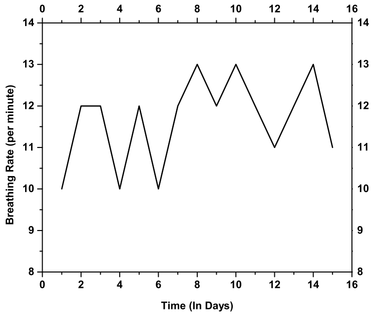


Fig. 6.2. *Breathing Rate*

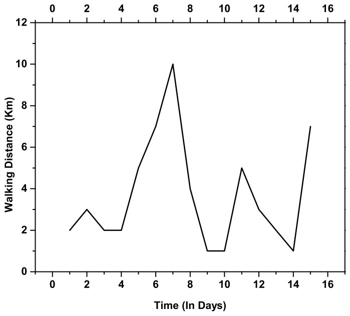


Fig. 6.3. *Walking Distance*

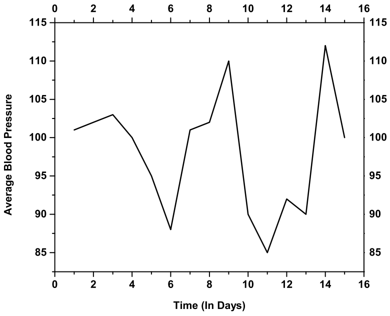


Fig. 6.4. *Average Blood Pressure*

6.1. Data Analysis. In this section, the real-time data capture and analysis is carried out in BAN. The sensed and gathered data from the IoT body sensing nodes were analyzed in the form of graphs. The Figure 6.1, 6.2, 6.3 and 6.4, which is indicative of body temperature, breathing rate, walking distance and blood pressure, gives an indirect method of prediction of diseases in patients.

6.2. Anomaly Detection. The proposed anomaly detection algorithm monitors the critical parameters for the detection of abnormality in a patient. Figure 6.5 shows the variations of heart rate for the monitored patients. The normal values for heart rate are in the range of [60-100] for a healthy human being. Any variation beyond this limit represents the abnormality.

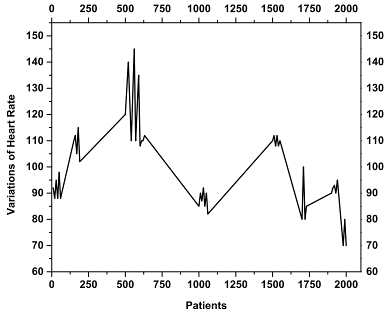
Figure 6.6 shows the variations of blood pressure for the monitored patients and it is observed the blood pressure vary from one individual to another.

Figure 6.7 represents the correlation between monitored health parameters. The proposed anomaly detection algorithm triggers a medical alarm whenever any abnormality is detected in these measurements, otherwise the measurement is considered to be faulty and is discarded without raising any alarm.

6.3. Resource Analysis. The smaller size of BAN devices puts more constraints on resources such as storage, energy, and processing. It is evident from Figure 6.8, 6.9 and 6.10 that the resources of the device are more constrained as the size decrease.

In simulations, a round is the total time required in performing one complete operation of sensing, aggregation, and offloading from body sensing nodes to the edge node. While carrying out the simulations, the proposed AEC operates for a longer time as compared to SEP as shown in Figure 6.11.

Figure 6.12 depicts the storage versus the number of body sensing nodes in AEC and SEP. As the number of body sensing nodes in a cluster increases, the amount of sensed data also increases, and as such it requires sufficient space on the edge node. But as the size of data goes beyond the edge's capacity, the offloading of data from the edge to the cloud for storage becomes necessary. It is evident in AEC, the storage capacity of the edge is more than that of the cluster head in SEP, resulting in less number of offloads.



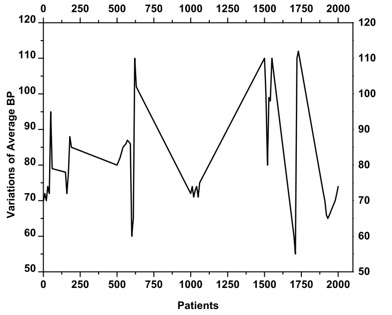


Fig. 6.5. *Variations of Heart Rate*

Fig. 6.6. *Variations of mean Blood Pressure*

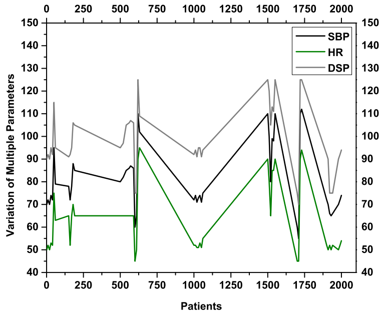
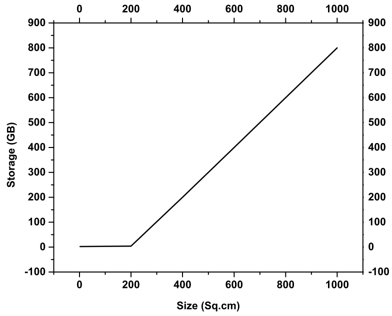


Fig. 6.7. *Variation of Multiple Parameters*



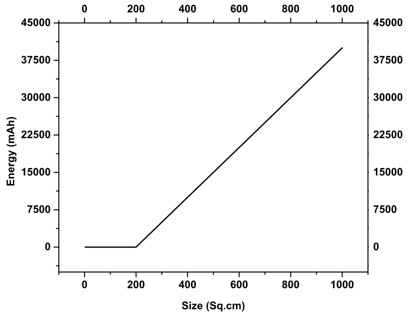


Fig. 6.8. *Storage in IoT Device*

Fig. 6.9. *Energy in IoT Devices*

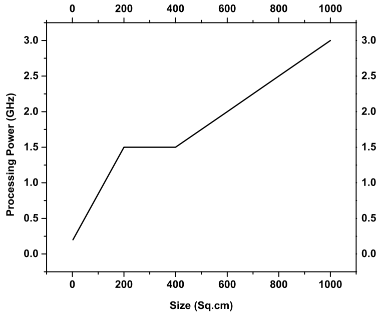


Fig. 6.10. *Processing in IoT Devices*

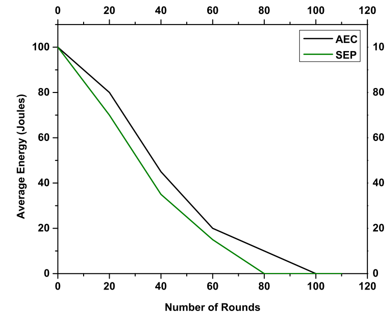


Fig. 6.11. *Energy versus Number of Rounds*

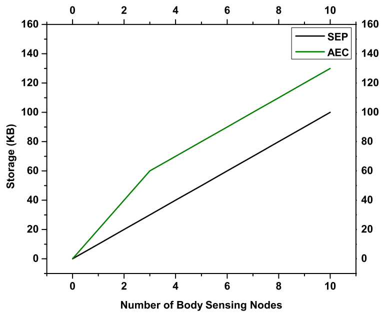


Fig. 6.12. *Storage versus Number of Body Sensing Nodes*

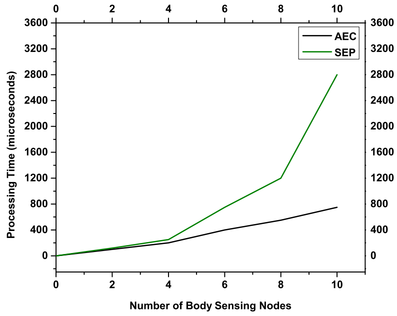


Fig. 6.13. *Processing Time versus Number of Body Sensing Nodes*

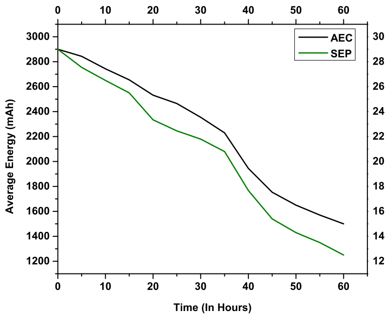


Fig. 6.14. *Energy versus Time*

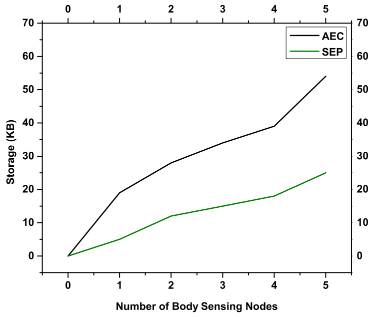


Fig. 6.15. *Storage versus Number of Body Sensing Nodes*

Figure 6.13 depicts the processing time versus the number of body sensing nodes in AEC and SEP. The processing time is more in SEP as compared to AEC because, in SEP, the edge node gets easily overloaded because of the limited processing capability of sensing nodes which results in more processing time.

The proposed AEC approach is also evaluated using empirical data to justify how the proposed approach suits the real-life scenario of BAN. Figure 6.14, 6.15 and 6.16 shows the energy versus time, storage versus the number of body sensing nodes, and processing time versus the number of body sensing nodes in AEC and SEP respectively. It is evident from these graphs that the energy, storage, and processing time are improved in AEC as compared to SEP in this case as well.

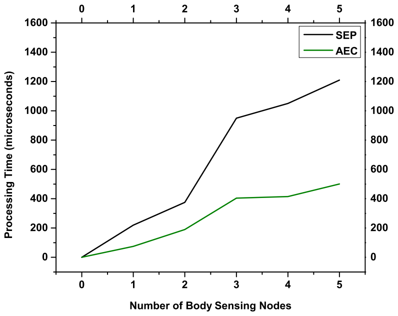


FIG. 6.16. *Processing Time versus Number of Body Sensing Nodes*

7. Conclusions. Body Area Networks is a newly emerging technology in the field of healthcare providing vital care and access not only to patients but to elderly and infants. It allows continuous, autonomous monitoring of patients and keeps the doctor informed of any abnormal fluctuations in the patient's health parameters by sending alarms and emergency services. The remote monitoring allows doctors to serve patients from any location where network connectivity exists.

Although BANs have been a hotspot of research with the emergence of wide range applications, there are open challenges in these. One of the most important challenges is the resource-constrained nature of BAN devices. Several small sized sensor-based wireless devices are used in BAN, the size of which is smaller as compared to other IoT applications. The smaller size of these devices puts more constraint on the use of resources such as energy, processing, bandwidth, and storage. This paper presented the design and modeling of these resource-constrained BAN systems along with its scenarios.

To address the resource limitations in BAN, we proposed an AEC approach that manages energy, storage, and processing time while monitoring the health parameters in patients. The comparison of the AEC approach is done with SEP and the simulation results showed an improvement in energy, processing time, and storage of AEC as compared to SEP. The comparative analysis is carried out on empirical data as well and it also showed an improved resources in AEC as compared to SEP, making the proposed solution suitable to meet the real-life scenarios of BAN.

The future research in BAN will be more interesting because of the pandemic situations like COVID-19 where it will be useful for doctors and paramedics to remotely monitor the patients as a safeguard measure.

REFERENCES

[1] W. Z. KHAN ET.AL, *Industrial internet of things: Recent advances, enabling technologies and open challenges*, *Computers & Electrical Engineering*, 81, pp.106522, 2020.

[2] S. ZAHOOR AND R.N. MIR, *Resource management in pervasive Internet of Things: A survey,Journal of King Saud University-Computer and Information Sciences*, 2018.

[3] F. LIU ET.AL, *Traversing knowledge networks: an algorithmic historiography of extant literature on the Internet of Things (IoT)*, *Journal of Management Analytics*, 4, pp.3-34, 2017.

[4] P. P. RAY, *Edge computing for Internet of Things: A survey, e-healthcare case study and future direction,Journal of Network and Computer Applications*, 140, pp.1-22,2019.

[5] S. BAKER ET.AL, *Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities*, *IEEE Access*, 5, pp.26521-26544, 2017.

[6] M. SRIDHAR ET.AL, *Wireless Body Area Networks: Requirements, Characteristics, Design Consideration, and Challenges*, *In Incorporating the Internet of Things in Healthcare Applications and Wearable Devices*, pp. 67-85, IGI Global, 2020.

[7] M. CHEN ET.AL, *Body area networks: A survey. Mobile networks and applications*, 16(2), pp.171-193, 2011.

[8] F. HU ET.AL, *Design and analysis of low-power body area networks based on biomedical signals*, *International Journal of Electronics*, 99(6), pp.811-822, 2012.

[9] B. ZENG ET.AL, *An energy-efficient data fusion protocol for wireless sensor network*, *In 2007 10th International Conference on Information Fusion*, pp. 1-7, IEEE, 2007.

[10] S. SHOKEEN AND D. PARKASH, *A Systematic Review of Wireless Body Area Network*, *In 2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, pp. 58-62, IEEE, 2019.

[11] M. Nabi Et.al,*A robust protocol stack for multi-hop wireless body area networks with transmit power adaptation*, *In Proceedings of the fifth international conference on body area networks*, pp. 77-83, 2010.

[12] I.H. Witten, *Data Mining: Practical Machine Learning Tools and Techniques (2nd edn)*, *Morgan Kaufmann, Elsevier, San Francisco, CA*, 2007.

[13] Alkhamisi Et.al, *A cross-layer framework for sensor data aggregation for IoT applications in smart cities*, *In IEEE International Smart Cities Conference (ISC2)*, pp. 1-6, 2016.

[14] G. Smaragdakis Et.al,*SEP: A Stable Election Protocol for clustered heterogeneous wireless sensor networks*, *in: Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004)*, 2004.

[15] R. A. Khan and A.S.K. Pathan,*The state-of-the-art wireless body area sensor networks: A survey*, *International Journal of Distributed Sensor Networks*, 14(4), p.1550147718768994, 2018.

[16] S. Al-Janabi Et.al, *Survey of main challenges (security and privacy) in wireless body area networks for healthcare applications*, *Egyptian Informatics Journal*, 18(2), pp.113-122, 2017.

[17] R. Negra Et.al, *Wireless body area networks: Applications and technologies*, *Procedia Computer Science*, 83, pp.1274-1281, 2016.

[18] D.H.Lee Et.al, *Development of a mobile phone based e-health monitoring application Development*, 3(3), 2012.

[19] S. Khan Et.al, *Wireless sensor networks: Current status and future trends*, *CRC press*, 2016.

[20] T. Sheltami Et.al, *Warning and monitoring medical system using sensor networks*, *In The Saudi 18th national computer conference (NCC18)*, pp. 63-68, 2006.

[21] T. Watanabe Et.al, *Tests of wireless wearable sensor system in joint angle measurement of lower limbs*, *In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5469-5472, IEEE, 2011.

[22] S. Ullah Et.al *A comprehensive survey of wireless body area networks*, *Journal of medical systems*, 36(3), pp.1065-1094, 2012.

[23] D. Lewis, *802.15. 6 call for applications in body area networksresponse summary*, 15–08–0407–05-0006, 2008.

[24] N. de Vicq Et.al, *Wireless body area network for sleep staging*, *In 2007 IEEE Biomedical Circuits and Systems Conference*, pp. 163-166, IEEE, 2007.

[25] L. Schwiebert Et.al , *A biomedical smart sensor for the visually impaired*,*In SENSORS, 2002 IEEE*, 1, p. 693-698). IEEE, 2002.

[26] S. Movassaghi Et.al *Wireless body area networks: A survey*, *IEEE Communications surveys & tutorials*, 16(3), pp.1658-1686, 2014.

[27] D.P. Tobón Et.al, *Context awareness in WBANs: a survey on medical and non-medical applications*, *IEEE Wireless Communications*, 20(4), pp.30-37, 2013.

[28] E. Jafer Et.al , *A wireless body area network for remote observation of physiological signals*, *IEEE Consumer Electronics Magazine*, 9, pp.103-106, 2020.

[29] R. Li Et.al , *A A wearable biofeedback control system based body area network for freestyle swimming* , *In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1866-1869, IEEE, 2016.

[30] J. Baikerikar Et.al , *A Home Security System Usings Face Recognition* , *IEEE In Advanced Computing Technologies and Applications* , pp. 303-310, Springer, Singapore, 2020.

[31] J. Muttik and M. Stecher , *Fuzzy fingerprinting of communicating wearables* , *U.S. Patent*, 10,588,005, 2020.

[32] A. Ali and M. F. A. Khan , *Fuzzy An improved EKG-based key agreement scheme for body area networks*, *In International Conference on Information Security and Assurance* , pp. 298-308, Springer, Berlin, Heidelberg, 2010.

[33] K. Revett Et.al, *Biosignals for user authentication-towards cognitive biometrics?*, *In 2010 International Conference on Emerging Security Technologies*, pp. 71-76, IEEE, 2010.

[34] D. Chen Et.al, *Towards a physiological model of user interruptability*, *In IFIP Conference on Human-Computer Interaction*, pp. 439-451, Springer, Berlin, Heidelberg, 2007.

[35] F. C. Delicato Et.al,*The Resource Management Challenge in IoT*, *In Resource Management for Internet of Things*, Springer, Cham, pp. 7-18, 2017.

[36] K. Hasan Et.al,*A comprehensive review of wireless body area network*, *Journal of Network and Computer Applications*, 143, pp.178-198, 2019.

[37] Wu, C.H. and Tseng, Y.C. *Data compression by temporal and spatial correlations in a body-area sensor network: A case study in pilates motion recognition*, *IEEE Transactions on Mobile Computing*, 10(10), pp.1459-1472, 2010.

[38] Z. Charbiwala Et.al, *Compressive sensing of neural action potentials using a learned union of supports*, *In 2011 International Conference on Body Sensor Networks*, pp. 53-58, IEEE, 2011.

[39] M.S. Manikandan Et.al, *ECG signal compression using discrete sinc interpolation*, *In 2005 3rd International Conference on Intelligent Sensing and Information Processing*, pp. 14-19, IEEE, 2005.

[40] B.S. Kim Et.al,*Wavelet-based low-delay ECG compression algorithm for continuous ECG transmission*, *IEEE Transactions on Information Technology in Biomedicine*, 10(1), pp.77-83, 2006.

[41] H. Kim Et.al, *A low cost quadratic level ECG compression algorithm and its hardware optimization for body sensor network system*, *In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5490-5493, IEEE, 2008.

[42] D. Vera Et.al, *Body area networks in healthcare: A brief state of the art*, *Applied Sciences*, 9(16), p.3248, 2019.

[43] K. Lin Et.al, *System Design and Data Fusion in Body Sensor Networks*, *In Telemedicine and E-Health Services, Policies, and Applications: Advancements and Developments*, pp. 1-25, IGI Global, 2012.

[44] A.S. Alzahrani and K. Almotairi, *Performance Comparison of WBAN Routing Protocols*, *In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1-5, IEEE, 2019.

[45] M.M. Et.al, *AMHRP: Adaptive Multi-Hop Routing Protocol to Improve Network Lifetime for Multi-Hop Wireless Body Area Network*, 2019.

[46] K.Y. Yazdandoost and R. Kohno, *Wireless communications for body implanted medical device*, *In 2007 Asia-Pacific Microwave Conference*, pp. 1-4, IEEE, 2007.

[47] J.S. Yoon, Et.al, *PNP-MAC: Preemptive slot allocation and non-preemptive transmission for providing QoS in body area networks*,*In 2010 7th IEEE consumer communications and networking conference*, pp. 1-5, IEEE, 2010.

[48] B. Otal Et.al, *Energy-efficiency analysis of a distributed queuing medium access control protocol for biomedical wireless sensor networks in saturation conditions*, *Sensors*, 11(2), pp.1277-1296, 2011.

[49] V. K. Pandey Et.al , *Applicability of Structured and Unstructured Communication Techniques in Emergency Environment*, *In International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, pp. 111-124, Springer, Singapore, 2020.

[50] K.C. Gouda Et.al, *A GA-Based Intelligent Traffic Management Technique for Wireless Body Area Sensor Networks*, *.. In Nature Inspired Computing for Wireless Sensor Networks* , pp. 57-75, Springer, Singapore, 2020

[51] M. Cicioğlu and A. Çalhan , *Energy-efficient and SDN-enabled routing algorithm for wireless body area network* , Computer Communications., 2020.

[52] O. Salem Et.al*Online anomaly detection in wireless body area networks for reliable healthcare monitoring*,*IEEE journal of biomedical and health informatics*, 18, pp.1541-1551, 2014.

[53] S. Agarwal and G. N. Pandey, *SVM based context awareness using body area sensor network for pervasive healthcare monitoring*, *In Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia* , (pp. 271-278), 2010.

[54] R. Latif Et.al, *Analyzing feasibility for deploying very fast decision tree for DDoS attack detection in cloud-assisted WBAN*, *In International Conference on Intelligent Computing*, pp. 507-519, Springer, Cham, 2014.

[55] F. Wang Et.al, *A human body posture recognition algorithm based on BP neural network for wireless body area networks*, *China Communications*, 13, pp.198-208, 2016.

[56] R.J. Quinlan, *C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA*, 1993.

[57] B. Ayrulu-Erdem Et.al, *Leg motion classification with artificial neural networks using wavelet-based features of gyroscope signals*, *Sensors*, 11, pp. 1721–1743, 2011.

[58] O.D. Lara *Centinela:A human activity recognition system based on acceleration and vital sign data*, *Pervas. Mob. Comput.*, 8, 717–729, 2012.

[59] Juvenile Diabetes Research Foundation Continuous Glucose Monitoring , *Study Group JDRF randomized clinical trial to assess the efficacy of real-time continuous glucose monitoring in the management of type 1 diabetes: research design and methods. Diabetes Technol Ther*, 10, pp. 310-321, 2008.

[60] Liu Et.al, *A novel trust-based secure data aggregation for internet of Things*, *In IEEE 9th International Conference on Computer Science & Education*, pp. 435-439, 2014.

[61] W. Heinzelman Et.al, *Energy-Efficient Communication Protocols for Wireless Microsensor Networks*, *In Proceedings of Hawaiian International Conference on Systems Science*, January 2000.

[62] T. Watteyne Et.al, *Anybody: A self-organization protocol for body area networks*, *In Second international conference on body area networks (BodyNets), Florence, Italy*,pp. 11–13, 2007.

[63] P. G. V. Naranjo Et.al, *P-sep: a prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks*, *The Journal of Supercomputing, vol. 730*, no. 2, pp. 733–755, 2017.

[64] A. Femi Et.al,*An Enhanced Stable Election Protocol for Clustered Heterogeneous WSN.*

Edited by: Anand Nayyar
Received: May 15, 2020
Accepted: Jul 6, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 515–541, DOI 10.12694:/scpe.v21i3.1782

SECURITY CHALLENGES IN FOG AND IOT, BLOCKCHAIN TECHNOLOGY AND CELL TREE SOLUTIONS: A REVIEW

NEELAM SALEEM KHAN[*] AND MOHAMMAD AHSAN CHISHTI[†]

Abstract. As the IoT is moving out of its early stages, it is emerging as an area of future internet. The evolving communication paradigm among cloud servers, Fog nodes and IoT devices are establishing a multilevel communication infrastructure. Fog provides a platform for IoT along with other services like networking, storage and computing. With the tremendous expansion of IoT, security threats also arise. These security hazards cannot be addressed by mere dependence on cloud model. In this paper we present an overview of security landscape of Fog computing, challenges, and, existing solutions. We outline major authentication issues in IoT, map their existing solutions and further tabulate Fog and IoT security loopholes. Furthermore this paper presents Blockchain, a decentralized distributed technology as one of the solutions for authentication issues in IoT. We tried to discuss the strength of Blockchain technology, work done in this field, its adoption in COVID-19 fight and tabulate various challenges in Blockchain technology. At last we present the Cell Tree architecture as another solution to address some of the security issues in IoT, outlined its advantages over Blockchain technology and tabulated some future course to stir some attempts in this area.

Key words: Cyber Physical Systems (CPS), Internet of Things (IoT), Certification Authority (CA), Public Key Infrastructure (PKI), End user (EU), Attribute Based Encryption (ABE)

AMS subject classifications. 68M10, 68M14

1. Introduction. While trying to meet our requirements and provide ease and add value in our daily routine activities devices are becoming more ubiquitous. All the devices like those used in industrial automation, households, and smart city framework are now interlinked with the web. From the way we make purchases to the way we drive and even how we get energy for our houses, the IoT is changing much about the world we are living in. As IoT is advancing with time, Cyber-physical System (CPS), mobile internet, several objects, like machines, people, and things are linked into an information zone anytime in any place [7]. Sensors and Sophisticated chips that transmit valuable data are implanted in the physical things that encompass us. A common IoT things platform brings us varied information together and brings the common language for devices and applications to communicate with one another. To decrease latency between the data generation and data processing stage due to the increase in smart applications and requirements, Cisco composed a new term called the Fog computing. Fog computing facilitates smart applications to carry out their action on network devices which can be switches, routers, or gateways, in place of sending data to Cloud datacentres [8]. To aid efficient data access, networking, computation, and storage and to scale the Cloud to the network edge, Fog computing becomes a new paradigm of distributed computing. Fog computing empowers a new variety of services at the edge and also offers a wide range of applications for IoT devices. It also backs heterogeneity, mobility, location awareness, low latency, huge scalability, and geo-distribution. In brief, the objectives of the Fog computing mode are to curtail the data volume and traffic to Cloud servers, improve the quality of service (QoS), and decrease latency [9]. The three-tier architecture is one of the fundamental and generally used architectures in Fog computing as illustrated in Fig. 1.1. The tiers are discussed as follows [10]

- Tier 1–Tier 1 contains IoT- devices, EU's smart hand-held devices (e.g., smart-watches, smartphones and, tablets), sensor nodes, etc. Often, these devices are called Terminal Nodes (TNs) and it is presumed that these TNs are rigged with Global Positioning System.
- Tier 2–Fog:In this layer, the Fog nodes consist of network devices like routers, gateways, switches, and Access Points (APs). Collaboratively, these Fog nodes get to share their computing facilities and

[*] Department of Computer Science & Engineering, NIT Srinagar, J&K, India (neelam_02phd17@nitsri.net).
[†] Department of Computer Science & Engineering, NIT Srinagar, J&K, India (ahsan@nitsri.net).

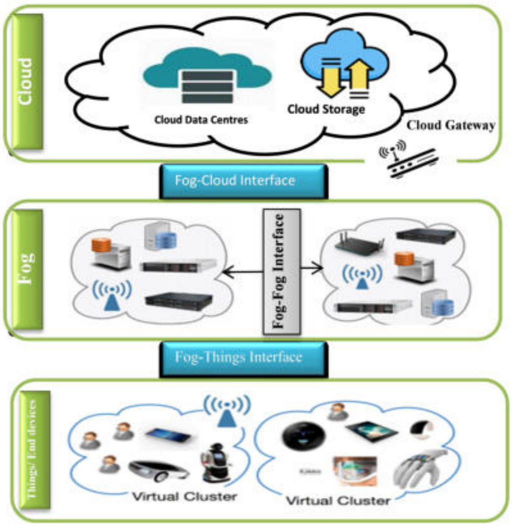


Fig. 1.1. *Three-tier Fog computing architecture [10].*

storage. The fog computing layer is another name given to Tier 2.

- Tier 3–Cloud: The upper layer, Tier 3 is the layer wherein the conventional Cloud servers and Cloud Data Centres (DC) reside. This tier holds adequate resources for storage and computation.

It's imminent that various security and privacy issues will endure sustaining because Fog is a nontrivial extension of the Cloud. While certain existent solutions in Cloud computing could solve many challenges of security and privacy in Fog computing as well, but because of mobility support which is a significant feature in Fog computing, it is likely to present many new security and privacy challenges. These introduced challenges may hamper the adaptability of Fog computing into the IoT. The domain of security and privacy challenges of Fog computing in IoT is yet in its infancy and is open for research.The main motivation behind IoT is to make human lives better and thriving, either by helping people make better choices and live better lives (less pressure, better well-being for impaired individuals, less repetitive jobs), or by making decisions within the framework that positively affect society, the environment, and the economy [1]. Plenty of security loopholes and conspicuous issues came into light after recent attacks on IoT gadgets and these hazards need to be addressed. Overall IoT Security expenditure is approximately expected to reach around $3.1 billion in 2021 estimated by Gartner. Irrespective of the consistent year-over-year increase in worldwide expenditure on IoT security, Gartner predicts that the largest obstruction to its growth will come from a lack of prioritization and implementation of security best practices and tools in IoT initiative planning. This will result in the decline of IoT security spending by 80 percent [2]. As securing IoT has always been in debates and news it becomes apparent that there is a requirement for more secure means of communication. Privacy and security risks were not concentrated to straightforwardly address the necessities of Fog computing. A few investigations were made concerning machine-to-machine communications [4] and smart grids [3]. As Fog devices commission at the Edge of networks on a bigger and more extensive scale, existing Cloud computing security solutions are not sufficient for Fog computing. The habitat of Fog devices is confronted with numerous threats which are not present in the well-governed Cloud [5] In Edge computing, some of the storage and computation tasks are shifted from Cloud data centers to the edge of the network, which might raise several issues related to security and privacy concerns. Specifically, privacy protection and data security are the most vital services in edge computing, which is our significant concern [6]. Individual information collected by IoT devices causes a privacy risk if not taken care of appropriately. So securing the Internet of Things is of utmost significance. Management of the devices should likewise require efficient authentication to abstain from hijacking and botnet proliferation. Therefore, for safeguarding the user and the business data that are collected by IoT devices,

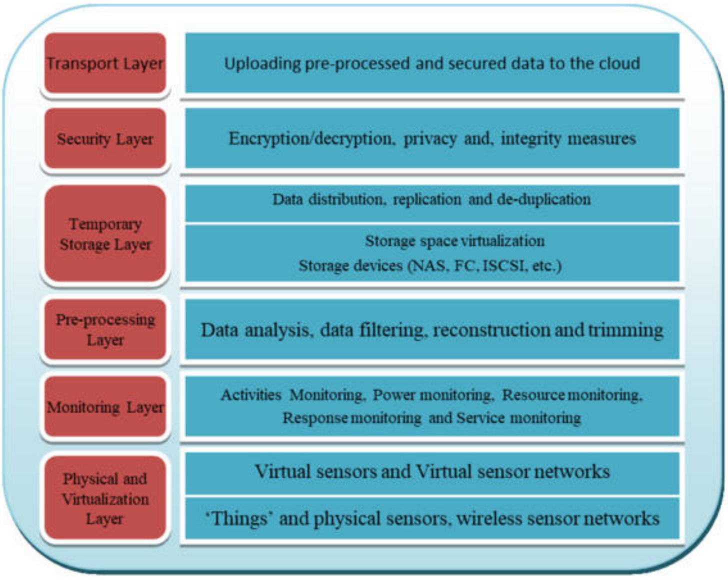

Transport Layer
Uploading pre-processed and secured data to the cloud
Security Layer
Encryption/decryption, privacy and, integrity measures
Temporary Storage Layer
Data distribution, replication and de-duplication
Storage space virtualization
Storage devices (NAS, FC, ISCSI, etc.)
Pre-processing Layer
Data analysis, data filtering, reconstruction and trimming
Monitoring Layer
Activities Monitoring, Power monitoring, Resource monitoring, Response monitoring and Service monitoring
Physical and Virtualization Layer
Virtual sensors and Virtual sensor networks
'Things' and physical sensors, wireless sensor networks

Fig. 2.1. *Smart Gateway-Layered architecture [18]*

innovations ought to be designed with security in mind. With layers of security deployed, overall a 'defense in depth' approach is required. The rest of this paper is divided into four sections. Section 2 provides an insight of various security issues and challenges in Fog and IoT environments. Section 3, then presents the Blockchain, a decentralized technology as one of the solutions, discusses work done in this field and highlights various challenges in Blockchain Technology as well. Section 4, discusses Cell Tree, a novel architecture for storage systems as another solution and highlights its advantages over Blockchain Technology. In lieu of a conclusion, Section 5 offers an outlook towards future development.

2. Related Work. Mohammad Aazam et. al. [18] identified Smart-Gateway-based communication, with Fog computing, the main motive for the design of this framework is for smart communication and aid in reducing the weight on Cloud. It also serves to mitigate the communication burden for the core network. For delay-sensitive applications in Fog computing, the above approach makes normal communication possible in real-time. Also, this technology will make it uncomplicated for the Cloud to create good services more aptly. This perception of IoT, Fog computing, and smart communication with Smart Gateway will bring a great deal of services. Figure 2.1 depicts their concept.

Maged Hamada Ibrahim et. al. [5] in their paper identified an efficient and secure framework that within the jurisdiction of Cloud service provider allows any Fog user in any node to mutually authenticate with any other Fog Server. Their design does not need a Fog user to be assimilated in any (Public Key Infrastructure) PKI. During the registration phase, the Fog user is needed to store a master secret key only once. The Fog server that is controlled by the Cloud service provider will mutually authenticate Fog users using this master key. The protocol comprises three stages: (a) Initialization phase, (b) Registration phase, and (c) Authentication phase.

If any of the Fog servers are compromised and depraved by an attacker, their scheme provides simple countermeasures. The master secret key with large enough bit-length of the client/user stays secure against a brute force even if all the Fog servers are compromised, and therefore, there is no need for the Fog user for any re-registration or re-initialization of a new master key. Also, once a Fog user registers, he is capable of mutual authentication with any Fog server that joins a Fog without the requirement for the user to re-register and there is no extra overhead on the user's side. However, this framework doesn't protect user's anonymity and that is a major drawback. The identity of the Fog user is flagrantly transmitted publicly, in the mutual authentication phase.

To enhance the effectiveness and competency of certificate revocation distribution in IoT surroundings, Arwa Alrawais, et. al. [15] describes a scheme in Fog computing. Their scheme comprises of Fog nodes, a Certification Authority (CA), IoT devices, and a back-end cloud. The proposed scheme offers an efficient certificate revocation information distribution approach. They used a bloom filter to create a shortlist that can adequately curtail the revocation list size with bearable overhead. Their design is a state-of-the-art technique in which all the renounced certificates are delivered immediately from the CAs to the Cloud, and after that to the Fog computing devices. The security of the certificate validation process is ensured by a quick update of the revocation information and it also wipes out the danger of obtaining a revoked certificate. Another way to add to security is to present a proof of reliability by allowing Fog nodes to sign each bloom filter. The signature serves as a definite proof that the vector originates from the Fog, as fabricating or modifying this signature is quite unfeasible. In their framework, the significance of employing the bloom filter is in the reduction of the computational overhead on resource-constrained IoT devices as the bloom filters utilize an effective hashing procedure for the verification of certification status. In this work, the authors also have put forward some possible suggestions for enhancing Privacy, Authentication, Access Control, Location Verification, and so on that still need attention.

In research conducted by Amandeep Singh Sohal et. al. [19], the authors illustrate a Cyber-security technique for diagnosing malignant edge-devices in a distributed Fog computing setting. For the early prognosis of the misbehaving edge devices and reliable edge devices, the proposed framework employs the two-state Markov model and categorizes the edge devices into four classes: Legitimate Device (LD), Hacked Device (HD), Under-attack Device (UD) and Sensitive Device (SD). The edge devices that the Intrusion Detection System (IDS) predicts to be hacked and posting wrong data continuously to the system are the HDs. These HDs although kept alive, are switched to Virtual Honeypot Device (VHD) to predict the path of the attacker successfully. Fig.2.2 illustrates the basic architecture of the Cyber-security framework under consideration which determines the malignant edge devices thereby building a more adaptive Fog computing security system. The main point in the proposed scheme is to allow for a legal edge device to revert from the VHD that may occur erroneously. Also, services to reinforce IDS, adaptive nature and false alarm controller have been included and properly tested. Designing an effective framework that efficiently deals with hacked devices transferred to VHD is a prospective research domain.

For the integrated Edge-Fog-Cloud network framework, Arij Ben Amor et. al. [20] put forth an anonymous and effective communication design. Their work contributes to the establishment of the Fog user-Fog server unidentified mutual authentication design in which the authentication takes place between Fog-server at the Fog layer and the Fog-user at the Edge by establishing a session key with each other without revealing the users' true identity. Whenever a Fog user moves within the same Fog from one Fog server to another, a light weighted authentication is guaranteed. Without the involvement of the Authentication server, the movement from one Fog server to another is done. In Fog-based Cloud computing, a new and secure authentication framework is presented. They conducted the formal validation and security analysis with the AVISPA tool that shows their results have enhanced privacy and security protection in comparison with some current schemes.

Pengfei Hu et. al. [12] in their survey paper discussed various attributes of Fog computing that includes Save bandwidth, Low latency and Support for mobility, Real-time interactions, Geographical distribution, and Decentralized data analytics, Heterogeneity, Privacy Protection, and Data Security, Interoperability and Low energy consumption. Several application cases like gaming and brain-machine interface, health care, augmented reality, IoT, and smart environments are enlisted to illustrate Fog computing applications. The primary technologies, like storage technologies and communication, computing, resource management, naming, privacy protection, and security were also outlined to inform how to reinforce its implementation and application in an elaborated pattern. Finally, some open issues and challenges which are worth farther investigation and research, including privacy and security, energy consumption, programming platform, are laid out. The security and privacy issues highlighted by this paper are explored in Table 2.1.

Saad Khan et. al. [11] in a study reviewed and analysed real-world Fog computing applications to analyse their potential security defects. To give a comprehensive survey, Fog related technologies like Cloudlets and Edge computing are also examined. Most of the Fog applications concentrate on functionality rather than considering security as part of their system, due to which many Fog platforms are being vulnerable. To

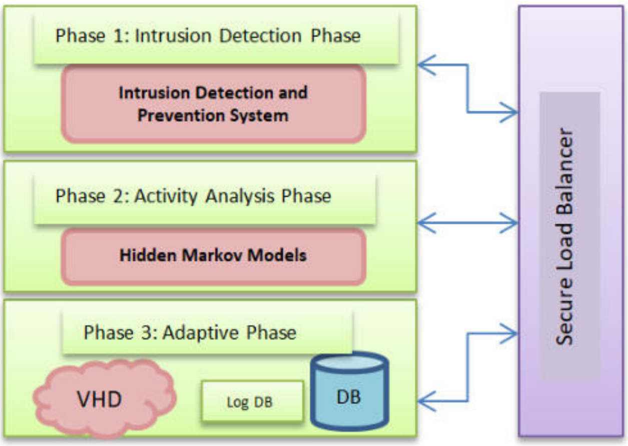

Phase 1: Intrusion Detection Phase
Intrusion Detection and Prevention System
Phase 2: Activity Analysis Phase
Hidden Markov Models
Phase 3: Adaptive Phase
VHD
Log DB
DB
Secure Load Balancer

Fig. 2.2. *Three Phase of Proposed Cyber-security Framework [19].*

formulate a systematic review, their study considers following twelve security categories: Advance Persistent Threats (APT), Data Breaches (DB), Access Control Issues (ACI), System and Application Vulnerabilities (SAV), Account Hijacking (AH), Denial of Service (DoS), Insecure APIs (IA), Data Loss (DL), Malicious Insider (MI), Insufficient Due Diligence (IDD), Shared Technology Issues (STI) and Abuse and Nefarious Use (ANU). This paper outlines the examination of how endorsed security solutions might be able to detect, prevent, and proactively shield against the threats highlighted by their survey. The objective of these security solutions is to safeguard the Integrity, Confidentiality, and Availability of the full Fog system and its users. Therefore, the adoption of a decision support system that is competent in saving the Fog platform from potential damage by recommending security measures to developers to prohibit the occurrence of susceptibilities pro-actively.

Eva Marín-Tordera et. al. [21] in their paper first reviewed the up-to-date technologies for Fog computing, paying more concentration to the contributions that inspect the performance edge devices play in developing a Fog node. They concentrated on the core services of a Fog node coupled with the challenges and opportunities towards their practical recognition soon. They present how a conceptual framework is developing towards a consolidated Fog node definition by plotting and comparing the ideas, lessons learned from their execution. After that, this paper also presented a logical view and an architectural approach for the first time about what a Fog node maybe? They categorize the job of Fog computing edge-devices into three fundamental types: 1) “dumb” devices as data producers/consumers, 2) “smart” edge devices with the ability to operate on their data, and 3) “truly smart” edge devices to execute distributed applications. Finally, this paper discusses open issues and challenges that include security and privacy in a Fog-based hierarchical scenario originating when the Fog node must offer a virtualized and abstracted view of its physical resources (i.e., sensing, networking, and computing) to upper layers.

Jun Zhou et. al. [16] identifies various privacy and security challenges in Cloud-Based IoT, recommend some methods to deal with the issues, and highlight future directions for the same. Table 2.2 concludes their paper.

Shahid Raza et. al. [22] has designed a scheme called SecureSense that appends security at the core of Cloud-connected IoT. Amidst a Cloud platform and an IoT device, SecureSense offers secure E2E data communication directly by using standardized Internet protocols. They enabled all three security methods of CoAP by incorporated the security protocol DTLS and IoT protocol CoAP into the SicsthSense Cloud platform and also outfitted these protocols in 6LoWPAN networks. The evaluation shows that the majority of the overall time is taken by the ECC functions and the computation of the Master Secret. The total time is much shorter in Pre-Shared Key (PSK) security mode since neither ECC functions are vital nor all handshake messages are required. These results also conclude that in comparison with the PSK mode, the extra and huge handshake

Table 2.1
Summary of Issues, Causes and Challenges in Fog.

Issue	Cause	Challenge
Man In The Middle Attack	Fog devices generally cannot utilize secure communication protocols because of their dearth of resources. Therefore an attacker can eavesdrop or impede the packets between Fog nodes.	The definite solution to Man-In-The-Middle Attack still an open challenge as it has been affirmed to be a silent attack against Fog computing.
Authentication	Entrusting on the Cloud central authentication servers is not a desirable decision, as Fog devices such as gateways may deal with many trust and authentication challenges that were not present in the Cloud case.	A holistic solution should be implemented to establish authentication and trust in the Fog.
Distributed Denial of Service (DDoS)	It is extremely hard for Fog nodes to deal with several requests simultaneously as they are resource-constrained. For a significant period, Fog nodes may become busy, by hurling a lot of insignificant service requests concurrently. With the result, resources that are used for hosting legal services become inaccessible. On the contrarily, Fog nodes can be utilized to eject a DDOS attack.	Compared to traditional DDoS, DDoS attacks using Fog devices will be more serious, the matter of DDoS requires to be addressed effectively in any subsequent Fog computing standardization.
Access Control	In Fog computing, scheming access control to traverse Client-Fog-Cloud and to reach the objectives and resource constraints at various levels is an issue of concern.	To create more powerful Access control mechanisms work needs to be done. The main aim of these techniques should be to help secure association and interoperability among heterogeneous resources in the Fog environment.
Fault Tolerance	There is an immense number of Fog nodes distributed widely geographically; when the service in an area is irregular, users should be able to turn to other adjacent nodes quickly by the alike mechanism.	When there is a breakdown in individual sensors, networks, applications, and service platforms, Fog computing should be capable of providing services normally.

messages for a certificate-based method do not fundamentally devote to the lengthy handshake time. Their outcomes fixed the performance criterion for all the three security methods of CoAP and present that for the Cloud-connected IoT, SecureSense is a feasible E2E communication security answer, in terms of time, storage, and energy overhead. They concluded that it is viable in battery-powered IoT gadgets (having just 32K of RAM) to use the strong government-grade certificate-based security, and the energy and timing overhead is agreeable for most IoT applications.

In another research paper, Antonio Escobar et. al. [23] presented possible ways for combining Cloud computing, Fog, and edge to advocate Internet of Things (IoT) solutions in achieving the challenging IoT needs. Their research concentrated on vertical distribution with clusters and without clusters, as well as sessions and incremental approaches. This paper also highlights that the critical part of IoT communication is Privacy and Communication challenges. It discusses communication with the Cloud using two paradigms viz End-to-Cloud and Gateway-to-Cloud. Drawbacks of both schemes make it crucial to research innovative ways like Transparent Gateways, Hardware Security, and End-to-End Encryption.

The core features discussed by Bidyut Mukherjee et. al. [24] in their paper are: 1) Intermittent Security using Session Resumption concept and 2) Flexible security build on the application resource-awareness. In their paper, they designed an End-to-End IoT Security Middleware. Amidst devices at the network edge and the core Cloud side of an application system lies this Middleware. This Middleware lies between the core Cloud side of an application system and devices at the network edge. This Middleware architecture based on an innovative security design offers affability for securing IoT-based application data, and to help in conditions of unstable network conditions within Cloud-Fog communication platforms by offering immediate re-connections. Their outcomes show the requirement for adaptability in the decision of an IoT security framework depending on resource constraints in bandwidth, computation, network reliability, memory, including the application for which the IoT system is being framed.

Mithun Mukherjee et. al. [9] discuss Privacy and security issues in Fog Computing which includes Trust, secure communication, authentication in Fog Computing, End-user Privacy, and malignant Attacks. This paper summarizes the up-to-date challenges in the domain of Security and Privacy threats for Fog Computing. Finally, they presented some open questions and challenges that can outline a blueprint for future research to resolve

Table 2.2
Summary of the issues in IoT discussed by [16]

Issue	Contribution	Future Direction
Fine-Grained Ciphertext Access Control	In Location Based Service (LBS) Fine-grained cipher text access control is a problem. It is seen that this issue can also be unfolded into several dimension scenarios and acquires several applications in outer space security.	A promising solution to this challenge is provided by designing a lightweight attribute-based encryption (ABE).
Location Privacy and Query Privacy	The query privacy would reveal their secret favorites, and moving route disclosure would disclose IoT users' living habits.	Scheming policy-hidden ABE exploiting the approach of a non-interactive proof system for bilinear groups would give us a favorable answer.
Secure Data Aggregation	One-way trapdoor permutation was solely used for secure data collection from a single user in the proposed efficient privacy-preserving technique. It is requisite to expand this proposed efficient privacy-preserving scheme to prevent privacy and security issues in further kinds of Cloud-based IoT. E.g., in smart grid IoT.	To attain secure data collection from multiple users in various types of Cloud-based IoT, a novel efficient privacy-preserving scheme needs to be designed.
Privacy-Preserving Outsourced Data Mining	It is vital to safeguard the user's identity, ensure the accuracy of an outsourced mining solution, privacy and location privacy, and guarantee that the solution can only be accessed by legal entities for example in vehicular IoT.	A challenging open problem is to develop a guaranteed outsourced data mining in the cipher text-domain.
Designing Lightweight Fully Homomorphic Encryption (FHE)	Public Key Fully Homomorphic Encryption (FHE) assuredly presents a substitute to established secure outsourced computation complying both multiplication and addition operations in the cipher text domain. A thorough search of the relevant literature yielded, that the vast volume of computational complexity still considerably disrupts developing lightweight FHE with its numerous application on resource-constrained users in Cloud-based IoT.	An effective privacy-preserving data aggregation without Public Key Homomorphic Encryption will help in the growth of Cloud-based IoT.

various issues in privacy and security in the Fog computing. Table 2.3 summarizes the Fog privacy and security issues that are open to research.

In this article, Yunguo Guan et. al. [26] have summarized the main challenges in solving data privacy and security challenges in Fog computing and identified various apprehensions about why the data security methods in Cloud computing cannot be legitimately adapted in Fog computing. Table 2.4 summarizes Privacy and Security needs in Fog computing as discussed in this article.

Ali Mohammad Saghiri et. al. [27] in their paper proposed a scheme for the Internet of Things based on Blockchain technology and cognitive systems. In their scheme, the status of things (from the things management layer) is detected by the cognitive engine of the cognitive process layer. Later, a convenient index of the smart contracts is triggered by the cognitive engine when it operates on actuators of the things. In the proposed framework, using peer-to-peer communication protocols several crypto currencies (coins) can be used for the payment process. For future research, the technologies of the Web of Things can be applied in the suggested scheme.

In another paper, Mohammad Alshehri et. al. [17] has suggested a new centralized Trust Management scheme for IoT. For the IoT Trust Management scheme, the most important characteristic of their scheme is the Super Node (SN), which is the main trust manager. They have explored the Trust Management Module, API module, and Repository and Communication Module. They have illustrated how it is managed, and how the design works effectively to incorporate trust in IoT communication. Khaled Salah et. al. [29] classifies IoT security issues in three types: Low-level security issues, Intermediate level, and High-level issues as mentioned in Fig.2.3.

Any two parties communicating with each other require authentication between them, to secure communication in IoT. The devices must be authenticated for particular access to services. There are a variety of authentication schemes for IoT. These schemes have distinct heterogeneous underlying architectures and environments which back IoT devices. Thus defining a standard authentication mechanism is a challenge. Also to provide access rights and information to the authorized ones, the authorization mechanism is required.

Syed Rameem Zahra and Mohammad Ahsan Chishti [128] in their paper tabulated the comparison of

Table 2.3
Open Research Challenges in Fog Privacy and Security

Issue	Open questions and Research challenges
Trust	In a FogNet, trust relations must be built by the Fog nodes with the devices using Fog network services. Additionally, Fog nodes that are deligated with data and processing requests by the IoT devices are required to create reliable interactions with the Fog nodes. In FogNet this two-way challenge makes the development of the trust model an important task.
Privacy Preservation	To support context-aware service location, the resources of End User's (EU's) devices are shared between other topographically neighboring devices, a large volume of data and other information about the EU need to be ensured in a highly secure manner. In this platform, Sensitive information such as identity and location of the Fog nodes can be easily revealed due to Man-In-The-Middle (MITM) attack. Therefore, to implement identity and location privacy for Fog computing is a formidable concern.
Authentication and Key Agreement	For data aggregated from resource-constrained devices, Fog nodes act as control and data aggregation points, therefore the issue of Authentication is one of the significant worries in Fog computing given the several level of gateways. Hence, Major challenges for Fog computing-based radio access Networks (F-RANs) are the authentication and key agreement protocols and they ought to be explored in the future. Also, to aid fine-grained access control, user-level key management and update mechanisms in a Fog storage framework is a very crucial task.
Intrusion Detection Systems	There is a need to deploy an edge Intrusion Detection System that can integrate the distinct detection components that will be dispersed inside the Fog network [25].
Dynamic Join and Leave Fog Node	A scheme should be designed to authenticate the EUs to the new Fog node and the privacy of the EUs must be preserved whenever a Fog node wants to exit the Fog layer. The errand is to frame a low complexity-based authentication between Fog node and EU is a crucial task in the expandable Fog network. The system should be able to find the users with their real identity once user misconduct is recognized by the Cloud service provider. Also to preserve the anonymity of the users is important.
Cross-Border Issue and Fog Forensic	To knock off cross-border legislation issues in Fog computing is an essential task

Table 2.4
Privacy and security requirements in Fog computing [26]

Data services	Different from Cloud computing	Privacy and security requirements
Storage	When processing is completed by the Fog layer, the data content will be altered and unrecognized to the data owner.	• Integrity verification • Public auditing • Minimum overhead • Dynamic support
Sharing	After the Fog layer processes the data, Access control of the data will be altered.	• Authorization revocation • Access efficiency • Fine-grained access control
Query	After the Fog layer processes the data, the keywords of the data will be altered.	• Secure searchability • Refined result • Dynamics support
Computation	The association among the data and computing functions will be altered after the Fog layer processes the data.	• Verifiability of outputs • Confidentiality of inputs • outputs and computing tasks

WSN's and IoT features. Their review also highlights various IoT applications and their security challenges like Smart city, Smart health, Smart building, Smart transport and Smart industry. They concluded that in order to address security issues in IoT environment, WSN's or other ad-hoc networks would not be 100% effective. Therefore more practical and efficient solutions are needed to address these security issues.

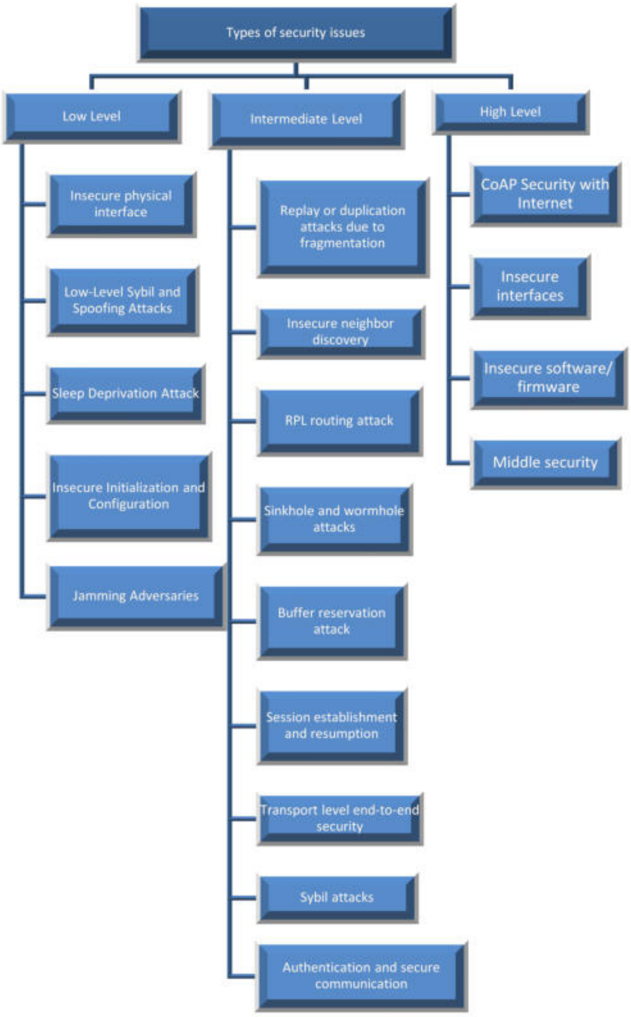

Types of security issues
Low Level
Intermediate Level
High Level
Insecure physical interface
Low-Level Sybil and Spoofing Attacks
Sleep Deprivation Attack
Insecure Initialization and Configuration
Jamming Adversaries
Replay or duplication attacks due to fragmentation
Insecure neighbor discovery
RPL routing attack
Sinkhole and wormhole attacks
Buffer reservation attack
Session establishment and resumption
Transport level end-to-end security
Sybil attacks
Authentication and secure communication
CoAP Security with Internet
Insecure interfaces
Insecure software/ firmware
Middle security

Fig. 2.3. *Three categories of security issues [29].*

2.1. Authentication Issues in IoT. This paper highlights several authentication issues in IoT and their proposed countermeasures by different researchers.

I Authentication and Secure Communication

Implications:

a Using Key Management Systems users and devices in IoT need to be authenticated.

b Any weak opening in security design or huge burden of communication security may disclose the network to many vulnerabilities [30][31][32].

c The overhead of Datagram Transport Layer Security (DTLS) needs to be reduced, due to constrained devices.

d To ensure security, cryptographic techniques are used. These mechanisms must take into account scarcity as well as the efficiency of other resources [33][34].

Proposed Solutions:

1 Compressed AH[42] and ESP[43]

2 SHA1 [92] algorithm takes less time and energy among different encryption techniques

3 Compressed IPsec for an end to end search by Raza [31, 44, 45]. Authors used AH and ESP or providing security using IPsec

4 Distinct versions of SHA1 and AES are implemented by [46] for encryption and authentication.

5 TPM(Trusted Platform module) using RSA [47, 48, 49]

6 Authentication with fuzzy extractor [50]

7 The proposed design by authors in Henze et. al. [52] allows for the configuration of IoT networks from a central location, thus safeguarding the IoT network from distrustful cloud services providers.

8 An authentication scheme that has secure packet forwarding, designed at offering privacy for location and identity on cloud-based IoT is given by Zhou et. al. [53].

9 To secure data communication between IoT devices, a platform in the SMARTIE project by Bohli et. al. [54].

II Transport level end-to-end security

Implications:

a Its main goal is to devise a secure scheme that ensures the correct destination node recieves data from the sender node reliably [35, 36].

b To establish message communication in a cryptographically secure form without disrupting privacy and maintaining the least overhead, a comprehensive authentication mechanism is required [37, **?**, 38].

Proposed Solutions:

1 Brachmann et. al. [35] suggested TLS-PSK , for end-to-end security, while accomplishing communication between HTTP and CoAP.

2 To allow negotiation of session keys, an extension of DTLS with nonce and PSK has been proposed.

3 For TLS, using the 6LOWPAN border router (6LBR) a designated authentication scheme is proposed by Granjal et. al. [36], which precludes the packets, operates on it to execute for the public key authentication computation, and then forwards packets .

4 For 6LoWPAN in tunnel and transport modes, the authors in [30] performed a preliminary assessment of the use of AH and ESP compression header security utilizing AES/CCM encryption at the hardware layer and an assumed application security profile.

5 An architecture coined BlinkToSCoAP for implementing end-to-end security in IoT is proposed in [37].

6 An approach implementing header compression for the 6LOWPAN protocol for decreasing DTLS overhead is suggested by Sinthan et al. [34] and Raza et al. [51].

7 Various header compression schemes have been suggested for implementing Transport Level end-to-end security [56].

8 An improved version of DTLS integrating header compression is suggested in Chavan et. al. [55] for securing IoT.

9 A lightweight design of Internet Key Exchange (IKE) designed to reform key management for 6LowPAN is suggested by Shahid et. al. [38].

III Insecure Neighbor Discovery

Implications:

a In IoT architecture, every device needs to be identified solely. To ensure this, message communication takes place that needs to be secure to assure that data transmitted in end-to-end communication (to a device) reaches the proper destination.

b Before data transmission, the neighbor discovery phase performs various steps along with router discovery and address resolution [39].

c Neighbor Discovery packets may have serious consequences without proper verification usage, along with denial-of-service [29].

Proposed Solutions:

1 Raza et al. [39] propose a security scheme with modules for key generation, authentication, data encryption, and secure neighbor discovery.

2 ECC is used to secure neighbor discovery [57].

IV Middleware Security

Implications:

a The IoT middleware developed to exchange communication between heterogeneous entities of the IoT architecture must be very secure for the provision of services [29] .

b To provide secure communication among heterogeneous entities distinct environments and interfaces using middleware are required to be implemented [40][41].

Proposed Solutions:

1 To secure distributed applications operating in an IoT environment, the VIRTUS middleware proposed by Conzon [40] implements encryption and authentication.

2 A semantic framework called Otsopack [58] works as middleware and uses TSC (Triple Space Computing) for interaction between applications and an open-ID based security method for secure data exchange.

3 Communication between heterogeneous IoT environments is proposed to be supported by a middleware server that has data filtering capability [41].

4 A standard framework with distinctive layers of security is suggested for M2M communication in IoT [60].

5 Open authentication and End-to-end security approaches are implemented in another middleware suggested by Ferreira et al. [59].

V Tampering and Malicious Code Injection

Implications:

a An attacker may physically modify an IoT device or communication link. This is called Tampering [104].

b An attacker can compromise a physical device and inject malicious code Injection onto it [105].

Proposed Solutions:

1 A mutual authentication protocol that is depends upon Physically Unclonable Fucntion (PUF). Depending upon the physical microstructure of device, the authentication takes place using a challenge response mechanism. Thus cloning the exact same structure by altering PUF is impossible which then eliminates tampering as well as malicious code injection [106].

VI Fake Node Injection and Side Channel Attack

Implications:

a In order to control flow of data between two nodes, an attacker can drop a malicious node between two authorized nodes of the network. This attack is called fake node injection [105].

b One of the attacks in side channel attack where in an attacker collects the encryption keys and later uses these keys to encrypt/decrypt data [104].

Proposed Solutions:

1 Aimed for WSN's Porambage et. al. [107] proposed "pervasive authentication protocol" (PAuthKey). PAuthKey obtained certificates from Cluster Head (CH) and then incorporates secure connection among end users and sensor nodes. This scheme successfully eliminates Fake Node Injection. Timing and power analysis attacks cause side channel attack.

2 Inbuilt verifiability of PAuthKey along with physical micro structure eliminates side channel attack [108].

3 A lightweight encryption algorithm along with masking technique can eliminate side channel attack [109].

VII Traffic Analysis, RFID Spoofing and RFID Unauthorized Access

Implications:

a In an attempt to obtain network information, an attacker sniffs the confidential data travelling between IoT devices [104].

b Information is imprinted on the RFID tag [105]. The attacker first gets access to this information by spoofing an RFID signal and then uses the original tag Id, sends its information depicting it as credible. This entire process is called RFID spoofing.

c Data Present on RFID nodes can be read, deleted or modified by the attacker due to absence of authentication schemes leading to RFID authorized access [104].

Proposed Solutions:

1 To protect IoT devices against traffic analysis Liu et. al. Liu et. al. [110] proposed effective and privacy preserving traffic obfuscation (EPIC) mechanism.

2 An on-board SRAM based Physically Unclonable Function (PUF) has been devised by Guin et. al. [111], which generates the device ID; a unique footprint for each IoT device. This ID will reduce the danger of spoofing and unauthorized access.

VIII Routing Information Attack, Selective Forwarding and Sink Hole Attack

Implications:

a In a routing information attack where an attacker creates inconvenience by altering, spoofing routing information, sending error messages, creating routing loops etc [104].

b Selective forwarding is a type of attack where an intruder drops some messages, alters some messages or simply selects some messages and forwards them to other nodes [112]. The result is that destination node receives the incomplete information.

c An attacker attracts other nodes in the network towards a compromised node (known as sinkhole node), so that network traffic flows towards it [105].

Proposed Solutions:

1 For Low power and Lossy Network's a Secure Routing Protocol (SRPL) uses hash chain authentication technique along with rank threshold concept has been proposed by Glissa et. al. [113]. This deals with routing attacks.

2 The above proposed scheme can be used for selective forwarding and sink hole attack. Pu et. al. [114] proposed CMD, a monitor based technique which uses RPL as routing protocol. This helps in detecting forwards misbehaviours.

3 Intrusion detection for SiNkhole attacks over 6Low-PAN for InterneT of ThIngs (INTI) proposed by Cervantes et. al. [115] discloses the identity of the malicious attacker node and isolates the detected sink hole node.

IX Man-in-the-Middle Attack and Replay Attack

Implications:

a In Man-in-the-Middle-Attack (MiTM) an attacker monitors or eavesdrops communication taking place between two IoT devices and gets access to the confidential information [104].

b In reply attack, the network is kept busy by capturing a signed packet and forwarding it over and over again to the destination [112].

Proposed Solutions:

1 MQTT and MQTT_SN suggested by Singh et. al. [116] prevents MiTM attack by ensuring secure device-to- device (D2D) communication. To incorporate Elliptic Curve Cryptography (ECC), MQTT uses Key-Policy (KP) and MQTT_SN uses cipher-text policy (CP) Attribute Based Encryption (ABE).

2 Park et. al. [117] suggested a scheme that prevents MiTM by authenticating inter-device communication. Session keys are generated and distributed by each sensor.

3 Based on Identity Based Cryptography (IBC) Ashibani et. al. [118] has proposed a signcryption scheme. This framework provides confidentiality, authentication and integrity simultaneously and thus prevents replay attacks.

X Data inconsistency, Unauthorized access and Data breach

Implications:

a In IoT, data inconsistency is referred to as the state of data when an attacker attacks on integrity of data during transmission or stored in a database [119].

b In unauthorized access, malicious users can get access to confidential or sensitive data or gain ownership rights [119].

c Leakage of confidential information in an unlawful or unauthorized pattern is called data breach [119].

Proposed Solutions:

1 To secure data transmissions within IoT devices Song et. al. [120] proposed a Chaos-based privacy preserving cryptographic technique along with Message Authentication Code (MAC). This guarantees data integrity.

2 Data stored in remote semi-trusted data storages need to be sure of data integrity, Machado et. al. [121] proposed a three-level split Blockchain based framework.

3 A Blockchain based framework along with ABE has been proposed by Rahulamathavan et. al. [122]. It supports non-repudiation, data integrity, preserves privacy of transaction data, inflicts access control and hence furnishes an end-to-end privacy preserving IoT system.

4 Zheng et al. [123] devices a privacy preserving efficient medical data sharing framework along with ABE.

5 Gope and Sikdar [124] proposed a lightweight privacy preserving two factor authentication schemes for preventing data breach.

6 To reduce the risk of privacy leakage Gai et. al. [125] proposed Dynamic Privacy Protection (DPP) model.

7 The authors in [126] have proposed a protocol called Improved Secure Directed Diffusion (ISDD) for ensuring end-to-end security of data in IoT environments.

2.2. IoT Security Loopholes. Table 2.5 briefly summarizes various loopholes in security of Fog and IoT devices inferred from this literature review which need to be addressed. In-depth analysis of these issues with effective and efficient security techniques need to be developed in order to make IoT a secure platform.

3. Blockchain Technology. Blockchain was devised by Bitcoin [69]. Blockchain extensively offers trustworthy and authorized identity registration, goods, assets, ownership tracking, and product monitoring [67]. Blockchain is the open ledger of all the Bitcoin exchanges that have ever taken place right back to the absolute first transaction. Subsequently the Blockchain is a consistently advancing technology that is continually developing as new blocks are being included. The transactions which are being included onto the Blockchain are being handled by computers associated with the network. These computers are frequently alluded to as nodes. These nodes are situated across the world thus being a decentralized technology. Each block in the Blockchain is appended to the chain in the sequential manner and anything that occurs on the network occurs in general. Blockchain is devised with durability and longevity in mind. It isn't constrained by any single substance. All the hubs that are a part of the network are a part of that community. Consequently it has no single point of failure. There are four key distinguishing features of Blockchain technology listed as follows [96].

1 Use of Smart Contract:Blockchain Technology having no single point of failure and using Smart Contracts, will be feasible to move things of significant worth, for example, property vehicles and so much, more safely. It will forestall any kind of scams that currently exist. This makes it very secure. In this way, smart contracts which when deployed on the Blockchain breathe life into it when purchasing selling, selling-purchasing every one of these things of significant worth. By the pre-eminence of cryptography upon the Blockchain everything will be digitally scanned and digitally signed. It will be a safe and effective method of moving things of profound worth.

2 Peer-to-Peer system: Another distinctive property of the Blockchain technology is that it cuts any kind of agent in the course of action. At the center of this technology is a peer-to-peer framework which is ensured security by cryptography, henceforth it has the ability to pull out attorneys, estate agents, account assistants and different experts from the center of the procedure e.g., when selling a house you can encourage that transaction legitimately through the Blockchain, consequently saving your cash.

3 Speed: Third distinguishing factor of Blockchain technology is that it allows for deals, transactions and all kinds of stuff completed very quickly e.g., the property deeds could be traded very quickly. It soothes out these regularly tedious assignments.

4 Capacity: Due to Peer-to-peer technology capacity of the whole network can be increased.

Blockchain does not only have impact on financial world but also contains other fields. These include digital rights, betting, debt management, escrow transfers, microfinance lenders, equity markets, private markets, the Remittance industry, and crowd funding platforms, derivative markets, land record deeds, E-commerce sites, global payment systems, P2P lending services, healthcare services, ownership records, election, casting a ballot

Table 2.5
Loopholes in Fog and IoT security

Loophole	Description	Reference
Cyber Attack	Due to huge quantity of data throughput and the probability of being able to obtain delicate information from both IoT devices and cloud, the Fog platform is a tempting target for Cyber-criminals.	Saad et. al. [11]
Denial-of-Service (DoS) attack	Denial-of-Service (DoS) attack is one of the malicious attacks that can be attempted. As a majority of the devices linked to the network are not mutually authenticated, therefore hurling a DoS attack turns out to be straight forward.	Mithun et.al. [9] Pengfei Hu et. al. [12]
Mobility of Fog nodes	One more important issue is that since Fog nodes leave or join the Fog layer time and again; Fog nodes are very dynamic. The well-studied Security and privacy techniques in Cloud Computing are not precisely applied due to the mobility of nodes.	Mithun et.al. [9]
Authentication	The End User's (EU) randomly moves around over the network and the Fog nodes joins and leaves a Fog layer frequently. Thus, the process of mutually authenticating EU's and Fog node is quite a challenge. To carry out data and processing requests by Fog nodes, EU's need to establish trusted relations with the Fog node. In IoT Authentication witnesses considerable challenges in efficiency and scalability. Traditional authentication methods remain insufficient, and there exists a demand for an efficient, secure, user-friendly and scalable scheme to cater to the needs of resource-constrained IoT devices.	Maged Hamada Ibrahim et. al. [5],Alrawais et. al. [15]
Efficient Intrusion Detection techniques	Currently, Intrusion Detection techniques are commonly utilized to alleviate attacks such as DoS attacks, scanning attacks, insider attacks or Man-in-the-middle attack. By employing IDS methods at each level of Fog Computing, many challenges emerge, such as false alarm control, real-time notification, correct response, and alarm parallelization.	Anwar et. al. [13]
Access Control	In Fog computing, another challenge is Access Control. How the access control policies are designed to traverse client-Fog-Cloud to serve the aims and resource constraints at different levels? The open issues need to be addressed to develop more robust Access Control policies that aim to strengthen interoperability and secure collaboration among the heterogeneous resources in Fog.	Pengfei Hu et. al. [12]
Public Key Infrastructure (PKI)	Managing the Public Key Infrastructure (PKI) which is needed to support secure communications is a significant challenge.	Stojmenovic et. al. [14], Law et. al. [28]
Location Based Service	IoT users living patterns would be exposed by moving route exposure and the query privacy would reveal their private favourites, query privacy and location privacy for Cloud-based IoT users in LBS (Location Based Service) has to be well preserved alongside privacy, trust, and authentication.	Zhou et. al. [16]
Mischievous Nodes	Minimal work has been done until now on the authority of trust or security improvement in the IoT environment, particularly concerning handling with mischievous nodes that are presently legal members of an IoT community.	Mohammad et. al. [17]
Resource Limitations	IoT devices have resource limitations. This requires protocols to be energy-efficient and lightweight in spite of requiring tedious calculations along with the advancement of energy harvesting methods.	Kamalinejad et. al. [64]
Heterogeneous Devices	Multi-layer security technology is required to be incorporated for heterogeneous devices that range from small low power devices with sensors to high-end servers.	Khaled et. al. [29]
Conversion Mechanism	The protocols incorporated at various levels need to interoperate by providing conversion mechanisms, for standardizing a global security mechanism for IoT.	Khaled et. al. [29]
Single Point of failure	The IoT environment is more vulnerable to single points of failures because of heterogeneous networks, protocols, and architectures.	Khaled et. al. [29]
Standard Verification Protocol	A standard verification protocol is an important element for tackling IoT security. Vulnerabilities should be exploited before deployment because after deployment it gets difficult to detect and alleviate. Along with physical malfunctioning, the implementation of a security algorithm in the hardware, packet processing, and routing mechanism also needs to be verified.	Khaled et. al. [29]
Scalable and Trusted Management, and Software Updates	Scalable and trusted management and software updates to millions of IoT devices are one of the open challenges.	Khaled et. al. [29]
User's Anonymity	In most of the research conducted on Fog, user's anonymity has not been protected. Identity of user is transmitted over public channel.	Maged Hamada [5]
Hacked devices	Designing an effective framework that efficiently deals with hacked devices transferred to virtual honeypot devices (VHD) is a prospective research domain.	Amandeep et. al. [19]
Privacy and Communication challenge	The critical part of IoT communication is Privacy and Communication challenge. It is crucial to research innovative ways like Transparent Gateways, Hardware Security, and End-to-End Encryption.	Antonio et. al. [23]
Revision of existing solutions	With the tremendous growth of IoT devices and with new attacks emerging due to changes in economic incentives or discovery of new bugs, the existing solutions may need revision.	Acharya et. al. [83]

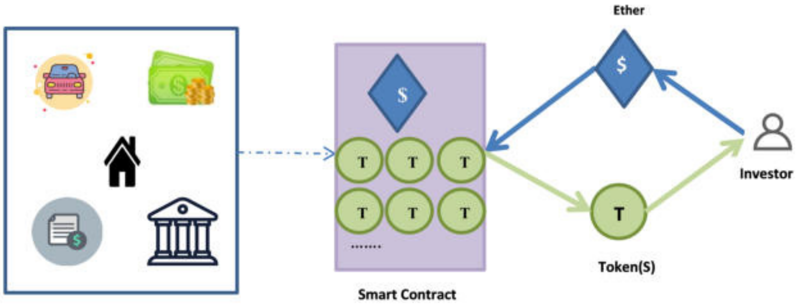


FIG. 3.1. *Ethereum Blockchain.*

and Intellectual property rights.

The Blockchain ensures trustworthy decentralized management, administration, and tracking at each phase in the supply chain and life span of an IoT device. Blockchain assures data integrity and authentication by ensuring that data being transmitted is cryptographically secured and signed by the legitimate sender. Blockchain smart contracts can ensure Authentication, Authentication, and Privacy. Every IoT device once installed and connected to the Blockchain network would have his unique GUID and symmetric key pair; therefore in Blockchain distribution and key management are wiped out completely. This will result in the use of lightweight security protocols. These lightweight protocols would fit and organize the need for the compute and memory resources of IoT devices. Blockchain guarantees tamper-proof storage of authorized transactions. Blockchain is used within IoT to store sensor data, manage device configuration and enable micro-payments [68]. Blockchain technology excludes the requirement for 3rd party verification [71].

3.1. Understanding the strength of Blockchain. The address space of Blockchain has 160 bit as compared to 128 bit in IPV6 [29] . Address length is 20 bytes or a public key generated ECDSA hash of 160 bit. Ethereum Blockchain was created with decentralized applications in mind [97]. Ethereum is launched in 2015 by Vitalik Bulterin. It is the world's programmable Blockchain. Ethereum laid the beginning of a new era of internet where i) payments and money are inbuilt ii) clients can own their data and apps cannot steal from you or spy on you iii) an open financial system becomes accessible to everyone iv) no person or company is the controller and is built on neutral, open access infrastructure. The 'ether' is the crypto currency that powers the Ethereum. Fig.3.1 illustrates the working of Ethereum Blockchain.

The Ethereum transaction has a couple of parameters as shown in the Fig. 3.2. The basic set of steps in realizing the strength of Ethereum Blockchain are as follows:

I User sends the Ethereum transaction ***Ti*** from a wallet (say metamask).

II The user has a private key ***Pr*** which is 32 bytes long and is randomized, 64-hex character string. Private keys can be generated at the user end by using a safe randomizer.

III ***Pr*** is sent through a function which is called ***ECDSA*** (Elliptic curve digital signature Algorithm). This function produces public key ***Pk*** which is 64 bytes long. ECDSA has property that ***Pk*** can be created from ***Pr*** but ***Pr*** cannot be created from ***Pk*** (without trying all combinations). This is strength of Blockchain.

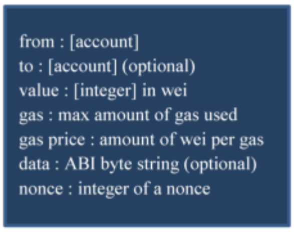
IV Create hash of the ***Pk*** by using the ***Keccak-hash(Pk)*** and take last 20 bytes of that i.e., ***B96.....255***. This would be the Ethereum Account (from field of the transaction).

V ***Ti*** is signed by ***Pr***. The output will be signed transaction ***Tis***.

VI ***Tis*** are run through the ***ECRECOVER*** function. The output will be ***Pk*** and the Ethereum Account (from field of the transaction).

This will make sure that the Ethereum transaction is authentic. It can be made sure that the account that was used in the 'from' field of the transaction is the same account that has the underlying private key that was used to sign the transaction and create the transaction signature. This is why in Blockchain every participating node can easily verify if the transaction is correct. The entire procedure is illustrated in Fig. 3.3.

All the blocks are chained together using cryptographic hashing [98]. Each block contains the hash of the



from : [account]
to : [account] (optional)
value : [integer] in wei
gas : max amount of gas used
gas price : amount of wei per gas
data : ABI byte string (optional)
nonce : integer of a nonce

Fig. 3.2. *Ethereum transaction.*

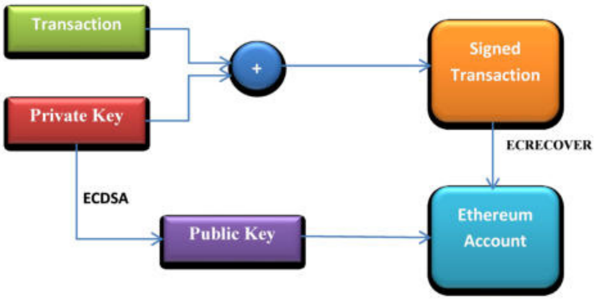


Fig. 3.3. *Understanding strength of Ethereum Blockchain.*

previous block, to ensure the correct sequence of transactions in the Blockchain, as shown in Fig. 3.4 Blocks have hashes of the previous block and this ensures transaction integrity. Any alterations to the transaction(s) in a block will alter all the blocks thereafter. In the event that a hacker attempts to alter any transaction, not only does he need to modify the transaction in the block, but all other blocks in the Blockchain. Moreover, hacker additionally needs to apply the change to every single node on the network, which is a computationally costly job to do. Each Blockchain has its own genesis block, which is the initial block of every Blockchain. The Bitcoin network has its own genesis block, and likewise Ethereum has its own genesis block.

Every participating node in the network has the copy of same information. Full nodes are the computers that store the Blockchain. Fig. 3.5 shows the full nodes in the Blockchain network containing the Blockchain.

Mining. Mining process creates new blocks on the chain using miners. A miner has following tasks:

- Combine the hash of the previous block and its transactions to derive a new hash
- The new hash is stored into the current block

To guarantee that all the miners have an equal opportunity to mine a block, Blockchain network will add a difficulty target in each block. So, result of the hash must meet the difficulty target to mine the block. In order to achieve this, miners inject a number called nonce into the block. Now miners will compete with each other in order to meet the difficulty target by guessing the value of nonce. So, the job of miners basically is to find the value of nonce.

This process of identifying the nonce is called Proof-of-Work (PoW). When the block is mined successfully change is accepted by all the nodes. The key idea behind PoW is that finding a nonce is difficult but verifying a nonce is easy once it is found. When mining is successful, the respective miner earns mining fees as reward.

3.2. Blockchain and IoT related work. Mayra Samaniego et. al. [68] discusses that a hosting location for deployment of Blockchain is a key challenge. Hosting on resource-constrained IoT devices is not advisable due to: a) Absence of computational resources b) Absence of sufficient bandwidth c) Need to preserve power. The paper performs experiments for both cloud and fog as a hosting platform. The findings of their evaluation are that Fog outperforms cloud. The fog has low latency while the cloud has high latency. Dr. B. V. Ramana Reddy [93] discusses four components of Blockchain as in Fig. 3.6.

Oscar Novo [70] in his paper brings forth some of the following benefits of access control in IoT as illustrated

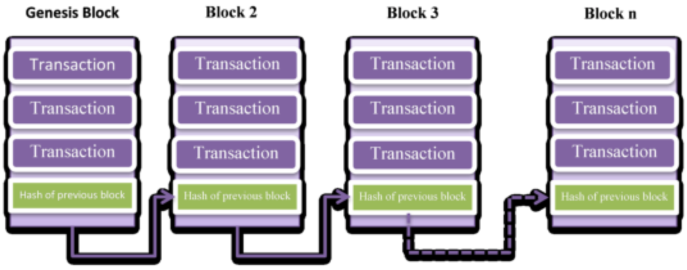


Fig. 3.4. *Hashing to chain the blocks in a Blockchain.*

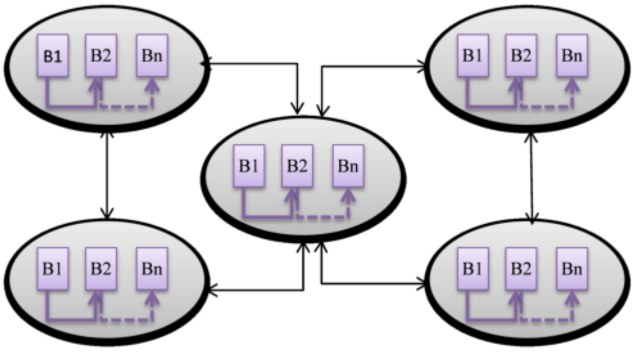


Fig. 3.5. *Nodes in the Blockchain network.*

in Table 3.1. The main focus of this research is on spawning a single smart contract that describes the policy rules of the administrative framework. Also, larger units of IoT devices that are not capable to execute Blockchain technology are included in their framework. The solution this paper provides experiences the burden of wait that the Blockchain network suffers in order to release the access control information. This waiting sceptically influences the performance of the proposed framework. The performance of the management hub is acceptable. The capability of IoT devices with respect to scalability is overall acceptable. In the proposed solution various threats as mentioned in Table 3.2 were identified using the STRIDE [72] model. Their research also suggests a solution for these security issues by introducing certification Authority. The CA will sign the management hub nodes and through this process the IoT devices could check the validity of the management hub.

A. Ouaddah et. al. [73] in their paper presented FairAccess as a new decentralized pseudonymous and privacy-preserving authorization management system that uses the steadiness of Blockchain technology to handle access control on account of constrained devices. This FairAccess is a crypto currency Blockchain based access control framework. For every resource-requester pair, their proposed model creates a distinct smart contract for the access control policy.

Nallapaneni Manoj Kumara et. al. [74] in their paper explains the probable security and privacy issues in consideration of the component intercommunication in IoT. The issue in centralized data Management Servers (CDMS) is that more arrangements and possibilities exist for revealing the sensitive parts of the data to the outside world through the false authentication, device spoofing. This paper mentions three major components of IoT as shown in Fig. 3.7 i) Things with Networked sensors and Actuators (TNSA) ii) Raw Information and Processed Data Storage (RI-PDS) iii) Analytical and Computing Engines (ACE). This paper also mentions various challenges faced by the integration of Blockchain technology with IoT that includes Limitation with storage facility, Lacking of skills required in the field, Lack of workforce, Legal issues, Variation in Computing Capabilities, Processing time and Scalability.

The authors in [66] discussed various issues of identity in IoT which included ownership and identity

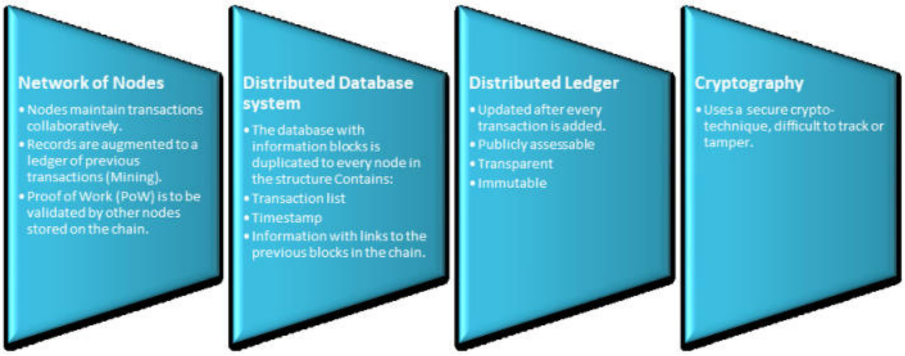

Network of Nodes
• Nodes maintain transactions collaboratively.
• Records are augmented to a ledger of previous transactions (Mining).
• Proof of Work (PoW) is to be validated by other nodes stored on the chain.
Distributed Database system
• The database with information blocks is duplicated to every node in the structure Contains:
• Transaction list
• Timestamp
• Information with links to the previous blocks in the chain.
Distributed Ledger
• Updated after every transaction is added.
• Publicly assessable
• Transparent
• Immutable
Cryptography
• Uses a secure crypto-technique, difficult to track or tamper.

Fig. 3.6. *Four Pillars of Blockchain Technology.*

Table 3.1
Advantages of access control in IoT

Advantages	Description
Mobility	To manage IoT devices every administrative domain in the network has freedom of its own. Access control is enforced by rules in Blockchain.
Accessibility	The system allows the accessibility of control rules at any time. Also, all the information regarding access control is distributed; therefore any event of some server failures doesn't prevent access to such vital information.
Concurrency	Multiple managers are concurrently granted access or ability to update the access control protocols.
Lightweight	No modification required to adapt their solution by IoT devices. Through the Blockchain network, communication amongst all the managers and IoT nodes occurs, thereby facilitating cross-platform communication.
Scalability	Supports multiple IoT devices interconnected by the various constrained networks to a single Blockchain.
Transparency	The system conceals the IoT device locations and also in what manner any resource is accessed.

relationships, authorization and authentication, governing of data and privacy. Slock.it [99] developed slock that is a smart lock technology which enables the Blockchain technology to control physical objects like (cars, bikes or houses). TransActive grid developed a technology that is a collaboration of hardware and software. This technology enables users to buy and sell solar energy securely from each other [100].

Based on Blockchain technology, Filament has built an open technology stack that enables devices to communicate, discover, and interact with each other in a completely distributed and autonomous way [100]. Bahga et. al. [61] introduced a decentralized peer-to-peer framework called BPIIOT for Industrial IoT using Blockchain Technology. Without the requirement of a trusted intermediary, the BPIIOT framework permits peers in a trust less network to connect. BPIIOT has an adequately wider scope than Slock.it and is capable of developing various peer-to-peers and distributed manufacturing applications. This paper mentions some benefits of Blockchain for IIoT as listed in Table 3.3.

To support the sharing of services among IoT devices and autonomous workflow, the authors in [62] [63] described smart contracts of the block that can facilitate the above thing. Seyoung Huh et. al. [75] proposed Blockchain to control and configure IoT devices. RSA public-key cryptography is used to manage keys; Ethereum stores the Public keys and on individual devices, the private keys are stored. This paper discusses in their evaluation that it has 12 sec transaction time and for time-sensitive domains, such topology may be difficult. Also to save entire Blockchain a proxy or a huge repository is needed. Utilizing a proxy can be simple but may jeopardize security as a third party is engaged. Another solution would be to utilize a large repository which would be very costly and still not feasible to cater to small IoT devices.

Sayed Hadi Hashemi et. al. [77] defines a complete architecture of three layers having Blockchain at the storage layer. Their work also defines data sharing protocol, data management. This paper also discusses different mechanisms and their impact that includes direct access to Blockchain, Client access, Publisher subscriber access.

Matthias Mettler et. al. [78] discusses the role of Blockchain in healthcare. Blockchain technology offers opportunities for usage in managing public health, drug counterfeiting, medical research that is user-oriented

Table 3.2
Threats identified in [70]

Threat	Effect
Spoofing	A noxious management hub could impersonate a management hub
Tamper	Access control information that is directed to the IoT devices can be altered
Repudiate	A device can claim that they have not taken any activity.
DoS	Debase the data directed to an IoT device or unveil unapproved data of the IoT device

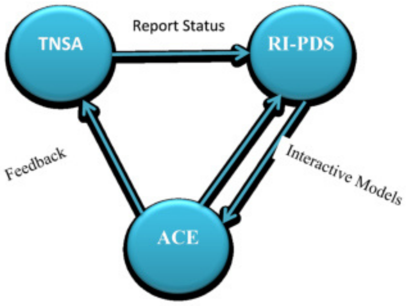


Fig. 3.7. *Three major components of IoT [74].*

and based on private patient data. It will also help advances in the latest digital health and business model initiatives. Asaph Azaria et. al. [79] discussed smart contracts in IoT. Their work describes the application of Blockchain to manage and access medical records, manage the relationship between different parties of Blockchain. They put forth the idea of having a decentralized record management system to manage EMRs (Electronic medical records) by employing Blockchain technology. When handling sensitive information crucial considerations are Confidentiality, accountability, authentication and data sharing. MedRec manages all of them.

Jie Zhang et. al. [80], discuss the use of Blockchain technique in PSN-based (Pervasive Social networking) healthcare. An updated version of the 'IEEE 802.15.6 Display Authentication Association' protocol is developed to initialize secure links.

Zyskind et. al. [81] discussed Blockchain implementation in a cryptographic approach, to protect personal data. This paper proposes a platform that enables Blockchain technology as an access-control moderator having an off-Blockchain storage solution.

Nabil Rifi et al. [76] put forth a Blockchain technology based publisher- subscriber technique and a data access protocol by adopting smart contracts. This paper used the model based on smart contracts that allow authentication, maintaining regulations, and communication among various nodes and groups of the system. The last component of the system uses off-chain database IPFS [82].

Authors in papers [84] [85] have proposed a solution to reduce the content stored on Blockchain, using alternate consensus mechanism [86][87][88] and using different graph topologies [89][90].

Leemon Baird et. al. [89] proposed Hedera a distributed ledger platform as well as an organization that resolves the aspects that restrict public DLT acceptance by the mainstream. A novel platform for providing consensus in a distributed setup is realized through a data structure called hashgraph and consensus algorithm. This paper also highlights the differences between Blockchain and hash graph. The hashgraph performs being quick, fair, ACID-compliant, lo-cost, efficient, time-stamped, Byzantine and DoS immune.

Serguei Popov [90] discussed the tangle that consists of a directed acyclic graph (DAG) for saving transactions is the next evolutionary stride that succeeds in the Blockchain.

Xiaoqi Li et. al. [95] in their paper presented several security issues of Blockchain technology. In their research they discussed every possible risk and vulnerability in Blockchain and also analyzed its probable causes and consequences. Their work also highlighted some real world attacks on Blockchain, and focused on the exploited vulnerabilities that lead to these types of attacks.

Table 3.3
Benefits of Blockchain for IIoT

Benefits	Description
Decentralized and Trust less	Each transaction in Blockchain is verified and validated through the consensus between all the transacting peers since it is decentralized; the peers need not to trust each other.
Resilient	Since Blockchain architecture is not centralized, the single point of failure doesn't exit. Blockchain cannot be altered or deleted, so Blockchain is an immutable ledger.
Scalable	As many new peers (or miners) keep on joining the chain, computing capacities of the network scales up.
Secure and Auditable	The security of every transaction on a Blockchain network is guaranteed by secure cryptographic methods and auditability as every transaction is transparent to everyone on the network.
Autonomous	No trusted third party is involved as all the IoT devices perform transactions autonomously with each device owning its Blockchain account.

According to [102], blockchain technology is now evolving as an influential framework for COVID-19 management. Blockchain was a focus to Chinese government since October 2019. After the 2020 corona virus breakout, the Chinese hospitals have been using Blockchain technologies in several applications like electronic heath records to insurance claims. Popular pharmaceutical companies have cooperated with software corporation SAP SE, of Walldorf, Germany, on Blockchain solutions for tracking of supply chain and identification of fake drug identification. As tests are conducted for COVID-19 vaccines and treatments, Blockchain can be used to endorse the trials. Furthermore, Blockchain can be adapted to learning about the monitoring disease outbreak patterns.

The rapid and uncontrolled outbreak of COVID-19 [103], points out the collapse of existing healthcare surveillance systems to manage public health emergencies on time. Blockchain is emerging as a tool to aid with several facets of containing the outbreak. The US department of Health and Human services office of the National coordinator for Health Information Technology broadcasted a nationwide interoperability guidelines requisitioning ubiquitous, secure infrastructure, authenticates all the participating entities, identity verification and persistent representation of authorization to access electronic health information. Blockchain technology could meet these expectations as it is a decentralized architecture, with main aspects like, data provenance, immutable audit trials and robustness. While keeping data privacy and security regulation intact, multiple nodes in permissioned Blockchain share and report important information instantly. Blockchain assists with prevention and control of disease by offering ways to improve many public health activities. The use of Blockchain can assist in prevention of pandemic by enabling early observation of epidemic, faster tracking of drug trials and impact management of outbreak and treatment.

The increasing demand of electronic medical records (EMRs) at this time of COVID-19 panedemic is evident. With security, privacy and transparency benefits of Blockchain, health records on it are eagerly pursued but extremely difficult to do. This time of crisis like COVID-19 will encourage research directions in this field in order to implement this at national and global levels [129].

Jayasree et. al. [127] in their review paper categorizes security attacks in IoT into four broad domains: (a) Physical Attacks (b) Network Attacks (c) Software Attacks (d) Data Attacks; and lists various countermeasures in these categories. Their paper also highlights the evolution of Blockchain technology and its benefits when incorporated with IoT and IIoT. Their survey discussed various security issues in IoT and IIoT w.r.t Blockchain technology and traditional solutions that need to be analyzed in-depth.

3.3. Challenges in Blockchain Technology. Table 3.4 briefly summarizes various issues in Blockchain Technology in IoT inferred from this literature review which needs to be addressed. In-depth analysis of these challenges with effective and efficient techniques needs to be developed in order to make Blockchain and IoT emerge as a promising platform for future.

4. Cell Tree. Cell Tree is a novel architecture for distributed storage systems that facilitates the storage of data in highly programmable and largely independent cells that are "assimilated" into a tree structure. Each cell has its policies and it allows data to change over time. Each cell is governed by its selected crew. The architectural design goal is to allow the evolution of the cell tree organically and designed to adapt itself across several applications. Anasuya Acharya et. al. [83] proposed a distributed data repository. The key philosophy

Table 3.4
Challenges in Blockchain Technology

Challenge/Issue	Description	Reference
Storage	All nodes connected in IoT store a replicated ledger on a Blockchain. Therefore, if stored on the Blockchain itself, storage would prove to be quiet inefficient.	Nabil et. al. [76]
Time consuming	The process of mining and validating blocks is time consuming as the count of nodes and the transaction count continues to grow.	Nabil et. al. [76]
Vulnerable	Blockchain provides robust approaches for securing IoT, but the approach is still vulnerable.	Jiang et. al. [65]
Secure Mining	Miners hashing power for the consensus mechanism can be jeopardized, therefore permitting the attacker to host the Blockchain.	Khaled et. al. [29]
Private Keys	Blockchain accounts can be compromised because of the limited randomness of private keys.	Khaled et. al. [29]
Race attack	Race attacks should be avoided which can cause double-spending during some transactions.	Khaled et. al. [29]
Scalability	Blockchain architectures face scalability challenges. All the full nodes are required to store the whole chain in order to fully validate any new blocks as the underlying Blockchain framework is ever-expanding.	Acharya et. al. [83]
Illegal content	The data stored may include illegal content which may result in legal complications for the Blockchain.	RomanMatzutt et. al. [91]
Irreversible effects	Due to the immutable (unable to change) nature of Blockchain, any implementation bugs can create irreversible effects.	Acharya et. al. [83]
Expensive POW	The consensus protocol employed by Blockchains called Proof-of-Work (PoW) ecologically turned out to be expensive. Therefore, this is another Blockchain limitation.	Acharya et. al. [83]
Availability and Consistency	In distributed architecture of data systems, there exists a trade-off between availability and consistency. Blockchain remains available and partition tolerant at the cost of its consistency. In Ethereum, the Blockchain resulted to be considerably faster than the Bitcoin. One major consequence of a quicker block time is its diminished security; therefore for newly mined blocks, multiple confirmations are required by many Blockchain applications to prevent transactions from double-spending.	Bahga et. al. [61]
Software Vulnerabilities in Smart Contracts	Smart contracts may suffer from software vulnerabilities that can be manipulated by hackers.	Bahga et. al. [61]
Risk of Attack	On a Blockchain, smart contracts serve as policy agreements between transacting groups that are not legally enforced to the outside network, any attacks can pose risks to the organisations, block miners and also to the entire Blockchain network.	Siegel et. al. [94]
Lack of Awareness	The lack of knowledge and awareness about the Blockchain technology in domains other than the financial sector, affects its widespread adoption.	Bahga et. al. [61]
Lawful Enforcement	Some regulations for decentralized systems like Blockchain and the need for lawful enforcement of smart contracts are needed in order to mitigate any conflicts among transacting groups.	Bahga et. al. [61]
Blockchain Adoption in Indian healthcare System	Since Blockchain is a decentralized architecture, its servers are across geographical regions. In order to keep transaction data encrypted high computing power is required that consumes lot of electricity which is deficit in India. As of today a uniform IT system for healthcare has not been implemented in India. One of the challenges in adoption of Blockchain is limited insight into the product lifecycle. India does not have prerequisite assets in place to go ahead with digital health system.	Garg [130]

is to allow for different solutions in the single system to coexist so as to facilitate the evolution of the system with time over several applications. The complete framework is agnostic to how every module is implemented, as several sub-problems are assigned to modules. In their design, an important feature is to allow for different users of the system to focus on different parts of the structure and remain burden less by the whole system data. Another thing in their proposed architecture is having a multi-level confirmation for new connecting blocks so that clients that trust the nodes in the lower levels of the hierarchy receive a quick confirmation of any block added to the system, and the ones that don't trust would wait for a confirmation from higher levels. The cell tree framework strives to be modular, cellular and evolving. The basic building block of the cell tree

Table 4.1
Comparative analysis of Cell Tree and Blockchain technology

	Blockchain	Cell Tree
Nodes	A conventional Blockchain can be viewed as a single node Cell Tree, with a programmed cell and a large crew.	A Cell Tree has many parallels with Blockchain in lieu of Blocks.
Multi-Level confirmation	In Blockchain framework, the entire chain is bound to accept any system update that takes place.	In a Cell Tree, based on its local policies, a Crew that operates a node updates its cell independently and then assimilates it into the tree.
Reserved Hash Pointers	Blockchain uses chain topology. In this, recently linked nodes bear 'hash pointers' to previous nodes. The reversed direction of hash pointers reflects the fact that in a conventional Blockchain, a block gets confirmed when future blocks attach to it.	Cell Tree uses tree topology. In this, the 'hash pointers' refer from old parent node to its new child nodes. In Cell Tree, a Cell (or an update to a cell) is confirmed by already existing nodes.
Distributed Ownership	In a Blockchain, for each fork, consensus throughout the entire network will be required, if multiple forks have to be essentially retained, making it infeasible as the numbers of forks grow.	In Cell Tree architecture, every single node is claimed and controlled by its own designated crew. The scalability of the framework greatly improves by associating the ownership of nodes to comparably small crews that function in parallel. Regardless of whether a node's crew behaves nefarious, any amends introduced to the node's cell get incorporated into the Cell Tree only after they get verified and acknowledged by the crews supervising it.
Dynamic Nodes	Blockchain focussed on the immutability or 'persistence' assurance of the block.	Cell Tree introduces 'consistency' to guarantee that each cell has been developed into its current form in consistency with the protocols defined by the cell.
Excising Malignant Cells	In Blockchain, the blocks that belong to a chain cannot be removed as per Blockchain design. This may lead to socio-legal issues whenever any unlawful data is facilitated over a Blockchain.	A Cell Tree allows to make deactivation of the malignant cells viable to, with a minimal effect on the complete tree.

Table 4.2
Features of Cell Tree

Features	Description
Mirroring, Pruning, and Grafting	Subtrees are allowed to be pruned from or to be grafted on a Cell Tree. In terms of mirroring, grafting is possible in multiple locations with little impact on other nodes.
Excising Cells	If the crews of all the nodes supervising it agree, the contents of a cell can be detected or may be changed without any respect to its program.
Computed Reads	A function of a cells' content can be accessed.
Secret Cells	Through confidential sharing or protected multi-party computation policies, the crew members of a node can offer to legitimate clients access to the contents of the cell or to the functions (crew members need not be aware of cell contents)
Computing or Multiple Cells	Framing of concurrent algorithms to work independently on each cell.
Saplings	A Cell Tree having its root incorporated into some other parent Cell Tree is termed as a sapling.
Higher Arity Trees	By permitting a particular crew to handle a subtree rather than a single node, Higher Arity nodes may be effortlessly simulated.
Incentivization	Different parts of a Cell Tree may employ different incentivization mechanisms.

is a cell that holds the data and an (even smaller) nucleus. Nucleus has code that specifies how a cell can evolve. A cell is hosted as a node of a binary tree. Each node is operated by a crew. The crew of each node is also authoritative to monitor nodes in comparatively small subtree that is rooted there. Several algorithms have been implemented by the crew members that define how cells evolve, policies for monitoring how some other cells evolve, and means to incorporate changes and distribute assimilation information across the tree [83]. Table 4.1 presents a comparative analysis of Cell tree and Blockchain. Cell Tree has several other features highlighted by the paper [83] that can be used to further strengthen the security aspect of IoT as listed in Table 4.2.

5. Conclusion. Currently, the Fog platform is primarily an active point for cyber-crimes due to a lack of centralized control and poorly secured edge nodes. Also, since majority of devices networked together are not being authenticated mutually, attacks become inevitable. Keeping in view the very dynamic nature of Fog nodes that keep on joining or leaving the Fog layer very frequently the current stringent security and privacy policies adapted in Cloud Computing environment are not directly applicable due to node mobility. In this paper we tried to highlight various risks and vulnerabilities in Fog environment and reviews countermeasures in this field. With resource-constrained IoT devices, authentication faces several challenges such as scalability and efficiency. In this review, we tried to highlight several authentication challenges in IoT and outlined existing solutions in this domain. This paper further tabulates security loopholes in Fog and IoT environments. We focused on Blockchain as one of the solutions for authentication in IoT, analysed its strength, reviewed existing research in this field, researched about its adoption in COVID-19 fight and summarized challenges faced by Blockchain technology in IoT environment. Finally, we proposed Cell Tree a novel architecture for storage systems as another solution for eliminating some of security issues in IoT, summarized its advantages over Blockchain technology and tabulates its features which can be implemented to improve the efficiency of IoT security.

REFERENCES

[1] D. Bastos, M. Shackleton and F. El-Moussa, *Internet of Things: A survey of Technologies and security Risks in Smart Home and City Environments* in IEEE Conference on Living in the Internet of Things: Cybersecurity of the IoT, London, UK, 28-29 March 2018.

[2] Gartner, *[2] Worldwide IoT Security Spending Forecast*, https://www.gartner.com/newsroom/id/3291817. (Last accessed on 15 may 2020.

[3] W. Wang and Z. Lu, *Survey Cybersecurity in the smart grid: Survey and challenges* in International Journal of Computer Networks, vol. 57, pp: 1344-1371, April 2013.

[4] X. Liang, X. S. Shen and X. Lin,, *"Grs: The green, reliability, and security of emerging machine to machine communications* in IEEE Communications Magazine, vol. 49, no. 4, pp. 28-35, 2011.

[5] Maged Hamada Ibrahim,*Maged Hamada Ibrahim, Octopus: An Edge-Fog Mutual Authentication Scheme*in International Journal of Network Security, vol.18, no.6, pp: 1089-1101, Nov. 2016.

[6] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, *Edge computing: Vision and challenges*in IEEE Internet Things Journal, vol. 3, no. 5, pp: 637-646, Oct. 2016.

[7] Atzori, L., Iera, A., Morabito, G, *The internet of things: A survey*in Journal of Computer Networks, vol. 54, no. 15, pp: 2787–2805, 2010.

[8] O. Osanaiye, S. Chen, Z. Yan, R. Lu, K. K. R. Choo, and M. Dlodlo,*From Cloud to Fog Computing: A Review and a Conceptual Live VM Migration Framework*in Special Section on Recent Advances in Cloud Radio Access Networks, vol. 5, pp: 8284–8300, 2017.

[9] Mithun Mukherjee, Rakesh Matam, Shu Leandros Maglaras, Mohamed Amine Ferrag, Nikumani Choudhury and Vikas Kumar,*Security and Privacy in Fog Computing: Challenges* in Intelligent Systems for the Internet of Things Journal, vol. 5, pp: 19293 – 19304, 2017.

[10] S. Sarkar, S. Chatterjee, and S. Misra,*Assessment of the Suitability of Fog Computing in the Context of Internet of Things*in IEEE Transactions on Cloud Computing, vol. 6, no.1, pp: 46 – 59, 2018.

[11] Saad Khan, Simon Parkinson and Yongrui Qin, *Fog computing security: a review of current applications and security solutions* in Journal of Cloud Computing: Advances, Systems, and Applications, vol. 6, no.19, 2017.

[12] Pengfei Hu, Sahraoui Dhelim, Huansheng Ning and Tie Qiu,*Survey on Fog computing: architecture, key technologies, applications, and open issues* in Journal of Network and Computer Applications, vol.98, pp: 27–42, 2017.

[13] S. Anwar, Jasni Mohamad Zain, Mohamad Fadli Zolkipli, Zakira Inayat, Suleman Khan, Bokolo Anthony and Victor Chang , *From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions* in Security and Privacy in Cloud Computing Environments, vol. 10, no. 2, pp. 1-24, Mar. 2017.

[14] I. Stojmenovic and S. Wen, *The Fog computing paradigm: Scenarios and security issues* in Proceedings of Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1-8, Sep. 2014.

[15] Arwa Alrawais, Abdulrahman Alhothaily, Chunqiang Hu and Xiuzhen Cheng, *Fog Computing for the Internet of Things: Security and Privacy Issues* in IEEE Internet Computing, vol. 21, no. 2, March 2017.

[16] Jun Zhou, Zhenfu Cao, Xiaolei Dong and Athanasios V. Vasilakos, *Security and Privacy for Cloud-Based IoT: Challenges, Countermeasures, and Future Directions* in IEEE Communications Magazine, vol. 55, no. 1, January 2017.

[17] Mohammad Alshehri and Farookh Khadeer Hussain, *A Centralized Trust Management Mechanism for the Internet of Things (CTM-IoT)* in Proceedings of the 12th International Conference on Broad-Band Wireless Computing, Communication and Applications, Barcelona, Spain, 8–10 November 2017.

[18] Mohammad Aazam, Eui-Nam Huh, *Fog Computing and Smart Gateway Based Communication for Cloud of Things* International Conference on Future Internet of Things and Cloud, 27-29 August 2014.

[19] Amandeep Singh Sohal, Rajinder Sandhu, Sandeep K. Sood, and Victor Chang, *A Cybersecurity framework to identify*

malicious edge device in Fog computing and Cloud-of-things environments in Article in Computers & Security, 2017.

[20] Arij Ben Amor, Mohamed Abid, Aref Meddeb, *A Privacy-Preserving Authentication Scheme in an Edge-Fog Environment* in IEEE/ACS 14th International Conference on Computer Systems and Applications, 2017.

[21] Eva Marín-Tordera, Xavi Masip-Bruin, Jordi García-Almiñana, Admela Jukan, Guang-Jie Ren and Jiafeng Zhu,*Do we all really know what a Fog node is? Current trends towards an open definition* in Journal of Computer Communications, vol. 109, pp: 117–130, September 2017.

[22] Shahid Raza, Tómas Helgason, Panos Papadimitratos and Thiemo Voigt,*SecureSense: End-to-End secure communication architecture for the Cloud-connected Internet of Things* in Future Generation Computer Systems, vol. 77, pp: 40–51, 2017.

[23] Antonio Escobar and Matthias Eberl,*Cloud, Fog, and Edge: Cooperation for the Future?* in IEEE Second International Conference on Fog and Mobile Edge Computing (FMEC), 2017.

[24] Bidyut Mukherjee, Roshan Lal Neupane and Prasad Calyam,*End-to-End IoT Security Middleware for Cloud-Fog Communication* in IEEE 4th International Conference on Cyber Security and Cloud Computing, 2017.

[25] Tiago Cruz, Luis Rosa, Jorge Proença, Leandros Maglaras, Matthieu Aubigny, Leonid Lev, Jianmin Jiang, and Paulo Simões,*A Cyber Security Detection Framework for Supervisory Control and Data Acquisition Systems* in IEEE Transactions on Industrial Informatics, vol. 12, no. 6, pp. 2236-2246, December 2016.

[26] Yunguo Guan, Jun Shao, Guiyi Wei, and Mande Xie,*Data Security and Privacy in Fog Computing* in IEEE Early Access Articles, vol. pp. no. 99, pp: 1-6,2018.

[27] Ali Mohammad Saghiri, Monireh Vahdati, and Kamran Gholizadeh,*A Framework for Cognitive Internet of Things based on Blockchain* in 4th International Conference on Web Research (ICWR), 2018.

[28] Y. W. Law, M. Palaniswami, G. Kounga, and A. Lo,,*WAKE: Key Management Scheme for Wide-Area Measurement Systems in Smart Grid,* in IEEE Commun. Mag., vol. 51, no. 1, pp: 34-41, Jan. 2013.

[29] Khaled Salah, Minhaj Ahmad Khan ,*IoT Security: Review, Blockchain Solutions and Open Challenges* in Future Generation Computer Systems 82, pp: 395-411, November 26, 2017.

[30] J. Granjal, E. Monteiro, J.S. Silva ,*Network-layer security for the Internet of Things using TinyOS and BLIP* in International Journal of Communication Systems, 27 (10) , pp: 1938–1963, 2014

[31] Shahid Raza, Simon Duquennoy, Tony Chung, Dogan Yazar, Thiemo Voigt, and Utz Roedig,*Securing Communication in 6LoWPAN with Compressed IPsec* in International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS), pp. 1–8, 2011.

[32] J. Granjal, E. Monteiro, J.S. Silva,*Enabling network-layer security on IPv6 wireless sensor networks* in 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, pp: 1–6, 2010.

[33] P.N. Mahalle, B. Anggorojati, N.R. Prasad, R. Prasad,*Identity authentication and capability-based access control (ICAC) for the internet of things* in J. Cyber Security. Mobility 1 (4), pp: 309–348, 2013.

[34] D.U. Sinthan, M.-S. Balamurugan ,*Identity authentication and capability-based access control (IACAC) for the Internet of Things* in JCyber Secur. Mob.1 (4), pp: 309–348, 2013.

[35] M. Brachmann, O. Garcia-Morchon, M. Kirsche ,*Security for practical CoAP applications: Issues and solution approaches* in10th GI/ITG KuVS Fachgespraech Sensornetze (FGSN 2011), 2011.

[36] J. Granjal, E. Monteiro, J.S. Silva ,*End-to-end transport-layer security for internet-integrated sensing applications with mutual and delegated ecc public-key authentication* inIFIP Networking Conference, pp.1–9, 2013.

[37] G. Peretti, V. Lakkundi, M. Zorzi ,*BlinkToSCoAP: An end-to-end security framework for the Internet of Things* in 7th International Conference on Communication Systems and Networks (COMSNETS), pp. 1–6, 2015.

[38] S. Raza, T. Voigt, V. Jutvik ,*Lightweight IKEv2: a key management solution for both the compressed IPsec and the IEEE 802.15. 4 security* in Proceedings of the IETF Workshop on Smart Object Security, vol. 23, 2012.

[39] R. Riaz, K.-H. Kim, H.F. Ahmed ,*Security analysis survey and framework design for IP connected LoWPANs* in International Symposium on Autonomous Decentralized Systems, pp. 1–6, 2009.

[40] D. Conzon, T. Bolognesi, P. Brizzi, A. Lotito, R. Tomasi, M.A. Spirito ,*The VIRTUS middleware: An XMPP based architecture for secure IoT communications* in 21st International Conference on Computer Communications and Networks, ICCCN, 2012, pp. 1–6.

[41] C.H. Liu, B. Yang, T. Liu ,*Efficient naming, addressing and profile services in Internet-of-Things sensory environments* in Ad Hoc Netw. 18 (Suppl. C), pp: 85–101,2014.

[42] S. Kent ,*RFC 4302 - IP authentication header* 2005. https://tools.ietf.org/html/rfc4302.

[43] S. Kent ,*RFC 4303 - IP Encapsulating Security Payload (ESP)* 2005. https://tools.ietf.org/html/rfc4303.

[44] S. Raza, T. Chung, S. Duquennoy, D. Yazar, T. Voigt, U. Roedig ,*Securing Internet of Things with Lightweight IPsec* in SICS, Lancaster University, UK, 2011. URL http://soda.swedishict.se/4052/2/reportRevised.pdf.

[45] S. Raza, S. Duquennoy, J. Hglund, U. Roedig, T. Voigt ,*Secure Communication for the Internet of Things–a comparison of link-layer security and IPsec for 6LoWPAN* in Secur. Commun. Netw. 7 (12), pp: 2654–2668, 2014.

[46] J.W. Hui, P. Thubert ,*Compression Format for IPv6 Datagrams in 6LoWPAN Networks draft-IETF-6lowpan-hc-13* 2010. https://tools.ietf.org/html/draft-ietf-6lowpan hc-13.

[47] T. Kothmayr, C. Schmitt, W. Hu, M. Brnig, G. Carle ,*A DTLS based end-to-end security architecture for the Internet of Things with two-way authentication* in 37th Annual IEEE Conference on Local Computer Networks - Workshops, pp. 956–963, 2012.

[48] T. Kothmayr, C. Schmitt, W. Hu, M. Brnig, G. Carle ,*DTLS based security and two-way authentication for the Internet of Things* in Ad Hoc Netw. 11 (8) (2013) 2710–2723.

[49] S.L. Kinney ,*Trusted Platform Module Basics: Using TPM in Embedded Systems* in Newnes, Newton, MA, USA, 2006.

[50] X. Huang, Y. Xiang, E. Bertino, J. Zhou, L. Xu ,*Robust multi-factor authentication for fragile communications* in IEEE

Trans. Dependable Secure Comput. 11 (6), pp: 568–581, 2014.

[51] S. Raza, S. Duquennoy, T. Chung, D. Yazar, T. Voigt, U. Roedig,*Securing Communication in 6LoWPAN with Compressed IPsec* in International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011, pp. 1–8.

[52] M. Henze, B. Wolters, R. Matzutt, T. Zimmermann, K. Wehrle ,*Distributed configuration, authorization and management in the cloud-based internet of things* in IEEE Trustcom/BigDataSE/ICESS, pp: 185–192, 2017.

[53] J. Zhou, Z. Cao, X. Dong, A.V. Vasilakos ,*Security and privacy for cloud-based IoT: Challenges* in IEEE Commun. Mag. 55 (1), pp: 26–33, 2017.

[54] J.M. Bohli, A. Skarmeta, M.V. Moreno, D. Garca, P. Langendörfer,*SMARTIE project: Secure IoT data management for smart cities* in International Conference on Recent Advances in Internet of Things (RIoT), pp: 1–6, 2015.

[55] A.A. Chavan, M.K. Nighot ,*Secure CoAP using enhanced DTLS for the Internet of Things* in Internat. J. Innovative Res. Comput. Commun. Eng. 2 (12), 7601–7608, 2014.

[56] S. Raza, D. Trabalza, T. Voigt, "6LoWPAN compressed DTLS for CoAP in IEEE 8th International Conference on Distributed Computing in Sensor Systems, pp: 287–289, 2012.

[57] R. Harkanson, Y. Kim ,*Applications of elliptic curve cryptography: A light introduction to elliptic curves and a survey of their applications* in Proceedings of the 12th Annual Conference on Cyber and Information Security Research, CISRC '17, ACM, New York, NY, USA, pp: 6:1–6:7, 2017.

[58] A. Gmez-Goiri, P. Ordua, J. Diego, D.L. de Ipiña ,*Otsopack: Lightweight semantic framework for interoperable ambient intelligence applications* in Comput. Hum. Behav. 30 (Suppl. C), pp: 460–467, 2014.

[59] H.G.C. Ferreira, R.T. de Sousa, F.E.G. de Deus, E.D. Canedo ,*Proposal of a secure, deployable and transparent middleware for the Internet of Things* in 9th Iberian Conference on Information Systems and Technologies, CISTI, pp: 1–4, 2014.

[60] J. Granjal, R. Silva, E. Monteiro, J.S. Silva, F. Boavida ,*Why is IPSec a viable option for wireless sensor networks* in 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, pp: 802–807, 2008.

[61] A.Bahga, V.K.Madisetti ,*Blockchain Platform for Industrial Internet of Things* in Tech.Rep.2016. URL http://file.scirp.org/pdf/JSEA_2016102814012798.pdf.

[62] K. Christidis, M. Devetsikiotis ,*Blockchains and smart contracts for the InternetofThings* in IEEEAccess4, pp: 2292–2303, 2016.

[63] V. Pureswaran, P. Brody ,*Device Democracy - Saving the future of the Internet of Things* i IBM, 2014. http://www 01.ibm.com/common/ssi/cgibin/ssialias?htmlfid=GBE03620USEN.

[64] P. Kamalinejad, C. Mahapatra, Z. Sheng, S. Mirabbasi, V.C.M. Leung, Y.L. Guan ,*Wireless energy harvesting for the Internet of Things* in IEEE Commun. Mag. 53 (6), 2015, pp: 102–108.

[65] X.Li, P.Jiang, T.Chen, X.Luo, Q.Wen,*A Survey on the Security of Blockchain Systems* in Future Gener. Comput. Syst. ,2017.

[66] I. Friese, J. Heuer, N. Kong ,*Challenges from the Identities of Things: Introduction of the Identities of Things discussion group within Kantara initiative* in IEEE World Forum on Internet of Things (WF-IoT), pp: 1–4, 2014.

[67] P. Otte, M. de Vos, J. Pouwelse ,*TrustChain: A Sybil-resistant scalable Blockchain* in Future Gener. Comput. Syst., 2017.

[68] Mayra Samaniego, Uurtsaikh Jamsrandorj and Ralph Deters ,*Blockchain As a Service for IoT* in IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp: 433-436, 2016.

[69] Nakamoto, Satoshi ,*Bitcoin: A peer-to-peer electronic cash system*, 2008.

[70] Oscar Novo ,*Blockchain meets IoT: An Architecture for scalable Access Management in IoT* in Journal of Internet of Things Class Files, vol. 14, no. 8, March 2018.

[71] Nallapaneni Manoj Kumar, Archana Dash, Neeraj Kumar Singh ,*Internet of Things (IoT): An Opportunity for Energy-Food-Water Nexus* in 1st IEEE International Conference on Power Energy, Environment & Intelligent Control (PEEIC2018), GL Bajaj, Greater Noida, India, 13th and 14th April 2018.

[72] S. Hernan, S. Lambert, T. Ostwald, and A. Shostack,*Uncover security design flaws using the STRIDE approach* in MSDN Magazine, Nov. 2006.

[73] A. Ouaddah, A. A. Elkalam, and A. A. Ouahman ,*Towards a Novel Privacy-Preserving Access Control Model Based on Blockchain Technology in IoT* in Cham: Springer International Publishing, pp: 523–533, 2017.

[74] Nallapaneni Manoj Kumara and Pradeep Kumar Mallick ,*Blockchain technology for security issues and challenges in IoT* in International Conference on Computational Intelligence and Data Science (ICCIDS 2018), 132, pp: 1815–1823, 2018.

[75] Seyoung Huh, Sangrae Cho and Soohyung Kim ,*Managing IoT devices using Blockchain platform* in ICACT2017 February 19 -22, 2017.

[76] Nabil Rifi, Elie Rachkidi, Nazim Agoulmine and Nada Chendeb Taher ,*Towards using Blockchain Technology for IoT data access protection* in IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB), 12-15 Sept. 2017.

[77] Sayed Hadi Hashemi, Faraz Faghri, Paul Rauschy and Roy H Campbell ,*World of Empowered IoT Users* in IEEE First International Conference on Internet-of Things Design and Implementation (IoTDI), 2016.

[78] Matthias Mettler, M.A. HSG Boydak ,*Blockchain Technology in Healthcare The Revolution Starts Here* in IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), 2016.

[79] Asaph Azaria, Ariel Ekblaw, Thiago Vieira and Andrew Lippman ,*MedRec: Using Blockchain for Medical Data Access and Permission Management* in International Conference on Open and Big Data (OBD), 2016.

[80] Jie Zhang, Nian Xue, and Xin Huang ,*A Secure System For Pervasive Social Network-Based Healthcare* in IEEE Access Volume: 4, pp: 9239 – 9250, 2016.

[81] Zyskind, Guy, and Oz Nathan,*Decentralizing privacy: Using Blockchain to protect personal data* in In Security and Privacy Workshops (SPW), IEEE, pp: 180-184, 2015.

[82] *IPFS - Content Addressed, Versioned, P2P File System*, https://ipfs.io/docs/ (Last accessed on 15 may 2020)

[83] Anasuya Acharya, Manoj Prabhakaran and Akash Trehan,*Cell Tree: A New Paradigm for Distributed Data Repositories* May 17,2019.

[84] Adam Back, Matt Corallo, Luke Dashjr, Mark Friedenbach, Gregory Maxwell, Andrew Miller, Andrew Poelstra, Jorge Timón, and Pieter Wuille ,*Enabling Blockchain innovations with pegged sidechains* https://blockstream.com/sidechains. pdf, 2014.

[85] Joseph Poon and Thaddeus Dryja ,*The bitcoin lightning network: Scalable off-chain instant payments* 14, 01, 2016. https://lightning.network/ lightning-network-paper.pdf

[86] *Cardano* 2015. https://www.cardano.org

[87] Yossi Gilad, Rotem Hemo, Silvio Micali, Georgios Vlachos, and Nickolai Zeldovich ,*Algorand: Scaling byzantine agreements for cryptocurrencies* In Proceedings of the 26th Symposium on Operating Systems Principles, pp: 51–68. ACM, 2017.

[88] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov ,*Ouroboros: A provably secure proof-of-stake Blockchain protocol.* In CRYPTO, Springer International Publishing, pp: 357 388, 2017.

[89] Leemon Baird, Mance Harmon, and Paul Madsen ,*Hedera: A governing council and public hashgraph network* 2017. https://www.hederahashgraph.com/whitepaper

[90] Serguei Popov ,*The tangle* 2017 http://iotatoken.com/ IOTA_Whitepaper.pdf

[91] RomanMatzutt, JensHiller, MartinHenze, JanHenrik Ziegeldorf, Dirk Müllmann, Oliver Hohlfeld, and Klaus Wehrle,*A quantitative analysis of the impact of arbitrary Blockchain content on bitcoin*in 2018.

[92] D.Eastlake, P.E.Jones,*RFC3174-US Secure Hash Algorithm1 (SHA1)* 2001. URLhttps://tools.ietf.org/html/rfc3174

[93] Dr. B. V. Ramana Reddy,*Blockchain: A Game changer for securing IoT Data* in Volume 8, Issue 5, pp: 580-588, MAY 2019.

[94] Siegel, D.,*Understanding the DAO Hack for Journalists* https://medium.com/@pullnews/understanding-the-dao-hack-for-journalists-2312dd43e99(Last accessed on 15 may 2020)

[95] Xiaoqi Lia, Peng Jianga, Ting Chenb, Xiapu Luoa, and Qiaoyan Wenc,*A Survey on the Security of Blockchain Systems* 6 Mar 2018. https://arxiv.org/pdf/1802.06993.pdf

[96] *Block chain developer course* https://vomtom.at (Last accessed on 15 may 2020)

[97] *WWW.Ethereum.org*(Last accessed on 15 may 2020)

[98] *Understanding How Blockchain Works*
https://blog.ndcconferences.com/understanding-Blockchain/ (Last accessed on 15 may 2020)

[99] *Slock.it* https://slock.it (Last accessed on 15 may 2020)

[100] *TransactiveGrid* http://transactivegrid.net (Last accessed on 15 may 2020)

[101] *Filament (2016) Foundations for the Next Economic Revolution Distributed Exchange and the Internet of Things* https://filament.com/assets/downloads/Filament

[102] *Blockchain adoption could help in COVID-19 fight* https://www.bioworld.com/articles/435042-blockchain-adoption-could-help-in-covid-19-fight (Last accessed on 15 may 2020)

[103] *Blockchain and Corona virus: could it prevent future pandemics* https://www.finextra.com/blogposting/18570/blockchain-and-corona-virus-could-it-prevent-future-pandemics (Last accessed on 15 may 2020)

[104] Andrea, I., Chrysostomou, C., Hadjichristofi, G.,*Internet of things: security vulnerabilities and challenges*in 2015 IEEE Symposium on Computers and Communication (ISCC), pp. 180–187, 2015.

[105] Ahemd, M.M., Shah, M.A., Wahid, A,*Iot security: a layered approach for attacks and defenses* in 2017 International Conference on Communication Technologies (ComTech), pp. 104–110, 2017.

[106] Aman, M.N., Chua, K.C., Sikdar, B.,*A light-weight mutual authentication protocol for iot systems* in GLOBECOM 2017 - 2017 IEEE Global Communications Conference, pp. 1–6, 2017.

[107] Porambage, P., Schmitt, C., Kumar, P., Gurtov, A., Ylianttila, M.,*Pauthkey: a pervasive authentication protocol and key establishment scheme for wireless sensor networks in distributed iot applications* in Int. J. Distributed Sens. Netw. 10 (7), 357430, 2014. https://doi.org/10.1155/2014/357430

[108] Aman, M.N., Chua, K.C., Sikdar, B.,*A light-weight mutual authentication protocol for iot systems* in GLOBECOM 2017 - 2017 IEEE Global Communications Conference, pp: 1–6, 2017.

[109] Choi, J., Kim, Y.,*An improved lea block encryption algorithm to prevent side-channel attack in the iot system* in 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–4, 2016.

[110] Liu, J., Zhang, C., Fang, Y.,*Epic: a differential privacy framework to defend smart homes against internet traffic analysis*in IEEE Internet Things J. 5 (2), pp: 1206–1217, 2018.

[111] Guin, U., Singh, A., Alam, M., Caedo, J., Skjellum, A.,*A secure low-cost edge device authentication scheme for the internet of things* in 31st International Conference on VLSI Design and 17th International Conference on Embedded Systems (VLSID), pp: 85–90, 2018.

[112] Varga, P., Plosz, S., Soos, G., Hegedus, C.,*Security threats and issues in automation iot* in 2017 IEEE 13th International Workshop on Factory Communication Systems (WFCS), pp:. 1–6, 2017.

[113] Glissa, G., Rachedi, A., Meddeb, A.,*A secure routing protocol based on rpl for internet of things* in IEEE Global Communications Conference (GLOBECOM), pp. 1–7, 2016.

[114] Pu, C., Hajjar, S.,*Mitigating forwarding misbehaviors in rpl-based low power and lossy networks* in 2018 15th IEEE Annual

Consumer Communications Networking Conference (CCNC), pp: 1–6, 2018.

[115] Cervantes, C., Poplade, D., Nogueira, M., Santos, A.,*Detection of sinkhole attacks for supporting secure routing on 6lowpan for internet of things* in 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), pp: 606–611, 2015.

[116] Singh, M., Rajan, M.A., Shivraj, V.L., Balamuralidhar, P.,*Secure mqtt for internet of things (iot)*in: 5th International Conference on Communication Systems and Network Technologies, pp. 746–751, 2015.

[117] Park, N., Kang, N.,*Mutual authentication scheme in secure internet of things technology for comfortable lifestyle*in Sensors 16 (1), 2015.

[118] Ashibani, Y., Mahmoud, Q.H.,*An efficient and secure scheme for smart home communication using identity-based signcryption.*in 2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC), pp: 1–7, 2017.

[119] Chan, M.,*Why Cloud Computing Is the Foundation of the Internet of Things* 2017. https://www.thorntech.com/2017/02/cloud-computing-foundation-internetthings/.

[120] Song, T., Li, R., Mei, B., Yu, J., Xing, X., Cheng, X.,*A privacy preserving communication protocol for IoT applications in smart homes* in IEEE Internet Things J. 4 (6), 1844–1852, 2017.

[121] Machado, C., Frhlich, A.A.M.,*Iot data integrity verification for cyber-physical systems using blockchain* in 2018 IEEE 21st International Symposium on Real-Time Distributed Computing (ISORC), pp. 83–90, 2018.

[122] Rahulamathavan, Y., Phan, R.C., Rajarajan, M., Misra, S., Kondoz, A.,*Privacy-preserving blockchain based iot ecosystem using attribute-based encryption*in IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pp. 1–6, 2017.

[123] Zheng, D., Wu, A., Zhang, Y., Zhao, Q., *Efficient and privacy-preserving medical data sharing in internet of things with limited computing power* in IEEE Access 6, 28019–28027, 2018.

[124] Gope, P., Sikdar, B.,*Lightweight and privacy-preserving two-factor authentication scheme for iot devices.* in IEEE Internet Things J.. 2018.

[125] Gai, K., Choo, K.R., Qiu, M., Zhu, L.,*Privacy-preserving content-oriented wireless communication in internet-of-things.*in IEEE Internet Things J. 5 (4), 3059–3067, 2018. [126]

[126] Sengupta, J., Ruj, S., Bit, S.D.,*End to end secure anonymous communication for secure directed diffusion in iot.* in Proceedings of the 20th International Conference on Distributed Computing and Networking, ICDCN '19, pp. 445–450, 2019. https://doi. org/10.1145/3288599.3295577.

[127] Jayasree Sengupta, Sushmita Ruj and Sipra Das Bit,*A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT.*in Journal of Network and Computer Applications, 30 oct 2019.

[128] Syed Rameem Zahra and Mohammad Ahsan Chisht,*Assessing the Services, Security Threats, Challenges and Solutions in the Internet of Things.*in Scalable Computing: Practice and Experience, Vol 20, Number 3, pp 457-484, 30 Aug 2019.

[129] *Five ways in which Blockchain technology can aid a recover* https://www.livemint.com/opinion/columns/five-ways-in-which-blockchain-technology-can-aid-a-recovery-11589479234967.html (Last accessed on 15 may 2020)

[130] Rahul K. Garg, *Is Blockchain in Indian healthcare worth the challenges* https://health.economictimes.indiatimes.com/news/health-it/is-blockchain-in-indian-healthcare-worth-the-challenges/64095898 (Last accessed on 15 may 2020).

Edited by: Anand Nayyar
Received: Jun 12, 2020
Accepted: Jul 17, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 543–553, DOI 10.12694:/scpe.v21i3.1783

A MACHINE TRANSLATION SYSTEM FROM HINDI TO SANSKRIT LANGUAGE USING RULE BASED APPROACH

NEHA BHADWAL[*] PRATEEK AGRAWAL[†] AND VISHU MADAAN[‡]

Abstract. Machine Translation is an area of Natural Language Processing which can replace the laborious task of manual translation. Sanskrit language is among the ancient Indo-Aryan languages. There are numerous works of art and literature in Sanskrit. It has also been a medium for creating treatise of philosophical work as well as works on logic, astronomy and mathematics. On the other hand, Hindi is the most prominent language of India. Moreover,it is among the most widely spoken languages across the world. This paper is an effort to bridge the language barrier between Hindi and Sanskrit language such that any text in Hindi can be translated to Sanskrit. The technique used for achieving the aforesaid objective is rule-based machine translation. The salient linguistic features of the two languages are used to perform the translation. The results are produced in the form of two confusion matrices wherein a total of 50 random sentences and 100 tokens (Hindi words or phrases) were taken for system evaluation. The semantic evaluation of 100 tokens produce an accuracy of 94% while the pragmatic analysis of 50 sentences produce an accuracy of around 86%. Hence, the proposed system can be used to understand the whole translation process and can further be employed as a tool for learning as well as teaching. Further, this application can be embedded in local communication based assisting Internet of Things (IoT) devices like Alexa or Google Assistant.

Key words: Rule based approach, Natural Language Translation, Parts of speech tagging, Sanskrit Translation, Hindi Translation

AMS subject classifications. 68T50

1. Introduction. Machine translation is defined as the branch of artificial intelligence which covers the task of converting a source language to any other target language from the set of natural languages. The meaning of the text should be preserved and output should be fluent as well as correct. It is one of the applications of natural language processing (NLP), which is the study of interaction between human languages and the computers. It is important for the computers to read, analyze, understand and derive meaningful information from the human natural language. This needs to be done in an accurate and optimized way. Apart from machine translation, other applications of NLP include sentiment analysis, speech recognition, auto summarizing and topic segmentation.

1.1. NLP phases. Major phases involved in NLP, once the system receives the input, are as follows:

1.1.1. Morphological processing. It is the study of recognizing how a base word is modified to form words having similar meanings and syntactical structures.

1.1.2. Lexical analysis. It is the process of dividing the whole text into smaller units called lexicons. The lexicon of a language can be a paragraph, a sentence or a word.

1.1.3. Syntactical analysis (Parsing). It is the process which analyses the arrangement of the words which shows their relationship and checks the sentence for grammar.

1.1.4. Semantic analysis. It is the process of extracting and checking the dictionary meaning of each word. The text is checked for meaningfulness.

[*] School of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab, India (bhadwalneha21@gmail.com)

[†] School of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab, India and Institute of ITEC, Univeristy of Klagenfurt, Austria (prateek061186@gmail.com) : corresponding author

[‡] Department of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab (India) (vishumadaan123@gmail.com)

Table 1.1
A Brief Comparison of Hindi and Sanskrit Grammar

Basis	**Hindi**	**Sanskrit**
Alphabets	45 characters (varnas)	46 characters (varnas)
Total Vowels	12 vowels (swaras)	13 vowels (swaras)
Total Consonants	33 consonants (vyanjanas)	33 consonants (vyanjanas)
Number	2; singular plural	3; singular, dual, plural
Gender	2; masculine, feminine	3; masculine, feminine, neuter
Person	3; first, second, third	3; first, second, third
Ordering of Sentence	S-V-O (Subject-Verb-Object)	Order free language
Total Tenses	3; Present, Past, Future	6; Aorist, Perfect, Present, Imperfect, First future, Second future
Verb Mood	3; Indicative, Imperative, Subjective	4; Imperative, Conditional Benedictive, Potential

Table 1.2
Vibhakti-Kaaraka relationship

Vibhakti	**Karaka**	**Meaning**
Nominative	Kartaa	Performer/ Subject
Accusative	Karama	Object
Instrumental	Karana	Instrument
Dative	Sampradana	For whom the action is performed
Ablative	Apadana	From where (place) the action is performed
Genitive	Sambandha	Denotes possession
Locative	Adhikarana	Location
Vocative	Sambodhana	Used to address someone

1.1.5. Discourse integration. It is the study of the relationship between any two sentences in a text. The meaning of one sentence may depend upon the preceding sentence and also brings about the meaning of next sentence.

1.1.6. Pragmatic analysis. It is the process which derives such aspects of the natural language that require real world knowledge. The text is re-interpreted to convey what it actually meant.

The benefits of using a machine translator instead of doing it manually are numerous. It is time-saving and optimizes effort and cost. It can also provide confidentiality and multi-lingual support which may not be possible in case of manual translation. However, problems may arise if context consideration is not taken into account. Similarly, languages may be ambiguous or have different structures. Also, machine translation is still not fully accurate.

1.2. Linguistic Features of Sanskrit. The Sanskrit language is one of the ancient Indo-Aryan languages. Vedic Sanskrit has been used to write many of the ancient documents. The classical Sanskrit is described in a famous grammar called Astadhyayi (Eight Chapters) composed by Panini. Sanskrit is written in Devanagari script as well as in various regional scripts. There are numerous works of drama and poetry in Sanskrit. It has also been a medium for creating treatise of philosophical work as well as works on logic, astronomy and mathematics. Grammatically, it is similar to Indo-European languages as it is an inflected language. Sanskrit language comprises of 46 characters (varnas) out of those, there are 33 consonants (vyanjanas) and 13 vowels (swaras). Sanskrit language comprises of a total of eight cases. There are six tenses (kaala). In addition, there are four moods (arthaa). These four moods and six tenses are together known as a total of ten Lakaaras of the Sanskrit grammar. In addition, Sanskrit word order is free which means most of the sentences can be read and written in free-word-order.

1.3. Comparison of Hindi and Sanskrit language. Hindi is one of the official languages of India [5]. It is the fourth most widely spoken language in the world after Mandarin, English and Spanish [22]. Hindi is directly derived from Sanskrit language. It is regarded as the Apabhramsha (corrupted version) of Prakrit which is the Apabhramsha of Sanskrit. Though both the languages have same root, there are grammatical differences between Hindi and Sanskrit.

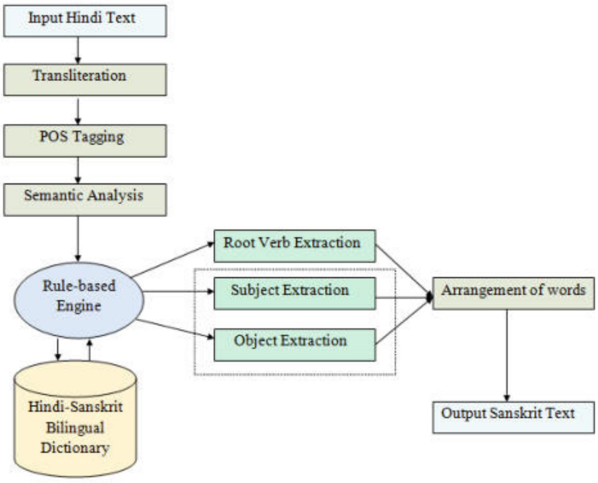


FIG. 1.1. *Architectural Work-flow of Hindi to Sanskrit Translation System (HST)*

Hindi has two types of vowels, short (Hrasva svaras) and long (Deergha svaras) vowels while Sanskrit has three types of vowels, short (Hrasva svaras), long (Deergha svaras) and elongated (Pluta svaras) vowels. In Sanskrit, the last syllable of the word is pronounced completely unless it is marked with a halanta (symbol). For Hindi, word order matters while for Sanskrit it does not matter. There are three basic parts of speech in Sanskrit namely, Shabda (nouns, pronouns, adjectives), dhaatu (verbs) and avyaya (indeclinables). Table 1.1 presents a comparison of Hindi and Sanskrit grammar.

1.3.1. Nouns and Gender. In Sanskrit, every noun has 24 forms which is a combination of a case (vibhakti) and a number (vachana). Also, each noun has a gender. However, adjectives in Sanskrit do not have fixed gender. It should comply with the noun it describes in gender, case and number.

1.3.2. Vibhaktis. A word can have eight possible vibahktis. Six of these, which are related to actions, are called kaarakas. Possessive and denominative are not related to any action. Table 1.2 depicts the association between vibhakti and kaaraka.

1.3.3. Sandhi. When a certain set of letters come together either between or within words, there occurs certain euphonic changes. Depending on the type of letters involved in coalescence, Sandhi can be categorized as Swara Sandhi, Vyanjana Sandhi and Visarga Sandhi.

The paper is organized as follows. The main results are in 4, the proposed algorithm is in 3, and the conclusions follow in 5. The previous related work is in 2.

2. Previous Work. A method is presented in [14] to perform syntax analysis of Hindi sentences via a parsing technique based on probability. This technique makes use of CYK (Cocke–Younger–Kasami) parsing algorithm. This morphological analyzer tool was used to identify whether the Hindi sentence is right or not when studied semantically by creating parse tables and making use of morphological information with around 80% accuracy in the syntax analysis stage around 89% in the semantic analysis phase (given the syntax analysis phase generates positive results). A rule-based lexicon parser is proposed by [23] for parsing tokens of Hindi sentence. The Hindi sentence was analyzed both syntactically and semantically after it was tokenized. This analysis was performed with the help of Hindi Wordnet by applying multiple tags. This parser was claimed to produce an accuracy of around 89% when evaluated for different sentences. English-Sanskrit-Hindi translation pair divergences are presented by [8]. The authors study English-Sanskrit and Hindi Sanskrit language pairs to identify divergences specific to each language involved. Verb frames for Hindi language were developed by [4]. A linguistic tool was developed to analyze and understand Hindi verbs. A classification of verbs was performed

Algorithm 2 HST algorithm

```
procedure HST(W, $V_w$, P, C, F, T, G, N, $P_v, S_v, P_n, S_n, P_1, P_2, D$)
    W: Set of all Hindi words given in the sentence
    Vw: Set of all verb form words in Hindi
    P: Set of all prepositions in Hindi
    C: Set of all cases in Hindi
    F: Set of all noun word forms in Hindi
    T: Set of all tenses in Sanskrit
    G: Set of all genders in Sanskrit
    N: Set of all numbers in Sanskrit
    Pv: Set of prefixes for Sanskrit verb form words
    Sv: Set of suffixes for Sanskrit verb form words
    Pn: Set of prefixes for Sanskrit noun form words
    Sn: Set of suffixes for Sanskrit noun form words
    P1: Set of translated Sanskrit noun words
    P2: Set of translated Sanskrit verb words
    D: Set of Sanskrit words in sentence after translation
    Set $S_{in}, S_{out}$ as empty
    FOR EACH $w_i$ in W
        IF$w_{i+1}$ ! =  NULL AND  $w_{i+1}$ = = ' है' | ' था'  THEN
          IF $w_i$.contains(' कर' | ' बन' | ' लग') THEN
                $S_{in} \leftarrow (w_{i-1}$ U $w_i$ U $w_{i+1})$
          ELSE
                $S_{in} \leftarrow (w_i$ U $w_{i+1})$
          ENDIF
            Compare reverse $[S_{in}]$ with predefined ending phrase list E[n] from corpus
          Store tense 't', number 'r' and gender 'g' corresponding to E[n]
          IF substrings $s_1, s_2, s_3 \ldots s_m \in S_{in}$match with E[n] THEN
            $S_{out} \leftarrow MAX$ [length $(s_1)$, length $(s_2) \ldots$ length$(s_m)$]
          ENDIF
          $S_v \leftarrow (S_{in} - S_{out})$
          IF $S_v == V_w$ THEN
            Store Sanskrit verb, $V_s$ corresponding to $V_w$ present in corpus
          ENDIF
          IF (t & r & g) = = (T[n] & N[n] & G[n]) THEN
            Store prefix, $P_v$ and suffix, $S_v$ corresponding to (T[n] & N[n] & G[n]) from corpus
          ENDIF
        ENDIF
        $P_2 \leftarrow P_2 \cup (P_v \cup V_s \cup S_v)$
    ENDFOR
    FOR EACH $w_i$ in (W-$S_{in}$)
        IF $w_{i+1}$ ! =  NULL AND  $w_{i+2}$ ! =  NULL AND  $w_{i+1}$ = = (' और' | ' या' | ' अथवा') THEN
          $w_i \leftarrow (w_i \cup$ ' ने')
          $w_{i+2} \leftarrow (w_{i+2} \cup$ ' ने')
        ELSE
          $w_i \leftarrow (w_i \cup$ ' ने')
          $w_{i+2} \leftarrow (w_{i+2} \cup$ ' को')
        ENDIF
        IF $w_{i+1}$ ! =  NULL AND  $w_{i+1}$ = = P THEN
          IF $w_{i+1}$ = = ' ने' THEN
            $c \leftarrow nominative$
          ELSE IF $w_{i+1}$ = = ' को' THEN
            $c \leftarrow accusative$
          ELSE IF $w_{i+1}$ = = ' के लिए' THEN
            $c \leftarrow dative$
          ELSE IF $w_{i+1}$ = = (' का' | ' के  ' | ' की')THEN
            $c \leftarrow genitive$
          ELSE IF $w_{i+1}$ = = (' में' | ' पर') THEN
            $c \leftarrow locative$
          ELSE IF $w_{i+1}$ = = ' से' THEN
            IF $w_{i+2}$.contains (' कर') THEN
                  $c \leftarrow ablative$
                  i = i+1
                ELSE
                  $c \leftarrow instrumental$
            ENDIF
            i = i+2
        ENDIF
        Store word form $w_f$ as last character of $w_i$
        $w_f \leftarrow$ last $[w_i]$
        IF (c & $w_f$ & r) = = (C[n] & F[n] & N[n]) THEN
            Store prefix $P_N$ and suffix $S_N$ corresponding to (C[n] & F[n] & N[n]) from the corpus
        ENDIF
        $P_1 \leftarrow (P_1 \cup P_N \cup w_i \cup S_N)$
    ENDFOR
    Combine two outputs to give final output D
    $D \leftarrow (P_1 \cup P_2)$
    Display D
```

and verb frames were created using the kaaraka relationships of the verbs. The criteria of classification was the syntactic and semantic differences between the verbs.

Natural language processing technique is applied by [11] to parse Hindi words and to extract the root words after performing stemming on each individual word. A tool is implemented by [26] to paraphrase the Hindi sentences by using active-passive voice rules and synonym-antonym replacement methods. An algorithm for the transliteration from English to Sanskrit text is presented providing 100% accuracy [12]. The process performs the mapping by making use of Hindi Unicode characters.

Also, a review of various types of MTS is presented by [24]. In their survey, the researchers highlighted and categorized mainly used approached for machine translation system as; rule-based, corpus-based and hybrid machine translation. In addition, [18] provided a view of example based machine translation system (EBMT). It was then compared with RBMT and SMT systems. The prominent characteristics of Sanskrit grammar and a comparison of English language and Sanskrit language were presented. The paper has shown the divergence between English and Sanskrit language with the help of illustrative examples. An overview of a FOS (Free and Open Source) rule-based MTS named Apertium is given by [7].A comparison of rule-based MT and Statistical MT was performed keeping in account the origin of the languages [29]. A five-way comparison between RBMT and SMT followed by English and Indian origin language was done.

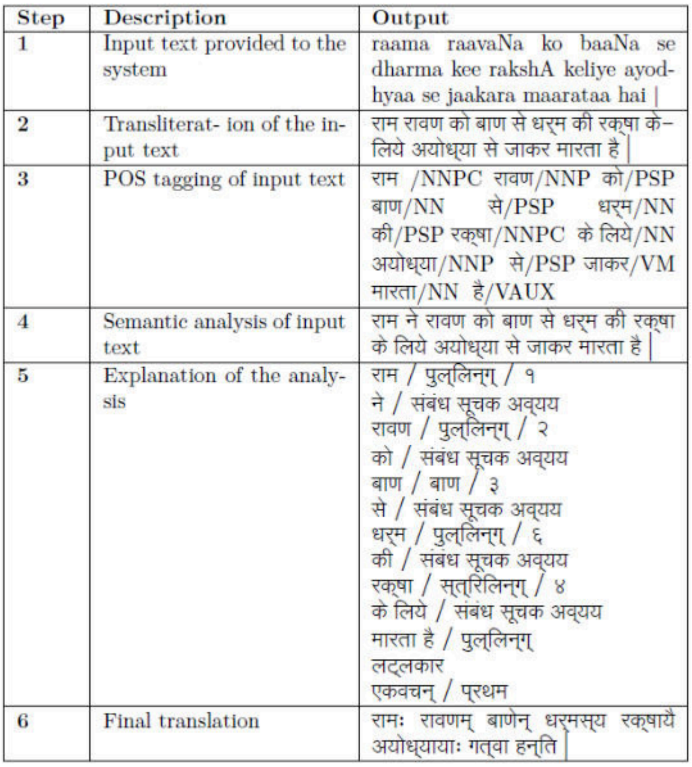
An evaluation method is given by [15] which they applied to the three major MT approaches namely, rule-base, phrase-based and neural MT. They have used a case study as the basis of their research work, which performs the translation from English to German language. A number of research efforts were reviewed by [28] under the example-based machine translation paradigm and attempted to categorize them into various classes. The features of various EBMT systems were discussed. The limitations and advantages of an EBMT system were highlighted.

Statistical methods were discussed by [6] along with pre-processing and post-processing of words to improve the performance of English to Arabic language MT such as morphological tokenization, syntactic reordering, orthographic normalization, morphological de-tokenization, orthographic de-tokenization and orthographic enrichment etc. An MTS was designed by [10] for English to Finnish translation, which uses rule-based approach. An empirical study was performed by [21] of the top five state-of-the-art machine translation systems from English to the languages which were considered to be having lesser resources namely, Lao(la), Myanmar(mm) and Thai(th) in both the directions. Various methods of MT were emphasised such as string-to-tree, tree-to-string, phrase-based, hierarchical phrase-based, operational sequence model statistical MTS.

A machine translation system based on rule-based approach was proposed by [17], which worked from English to Sanskrit language. The proposed approach, makes use of a set of transfer rules which are handwritten and convert the lexicons (could be paragraphs, sentences or phrases) from English to Sanskrit language. On a similar note, a process engine named EtranS which accepts sentence in English and produces its equivalent sentence in Sanskrit after translation, was developed by [3]. The approach followed was rule-based and they discussed firstly some syntactical features of the Sanskrit language followed by a comparison between Sanskrit grammar and Context Free Grammar. Another rule-based approach to manifest knowledge representation of a machine translation procedure from Sanskrit to English language was presented by [9]. Various MTS which involve Sanskrit as either a source, target or a key support language are discussed by [13]. They also presented different techniques used by researchers for machine translation, such as, Corpus based, Rule based and Direct translation. The principal objective of this paper was to find out the Sanskrit language suitability, morphology and apply most suitable MT techniques. Furthermore, an RDR POS tagger was developed by [20] to tag the words based on part of speech of the given sentence. A rule-based MT approach from Hindi to English was proposed by [16]. In his Ph.D. thesis report, [1] proposed a machine translation system for Sanskrit to Hindi language.

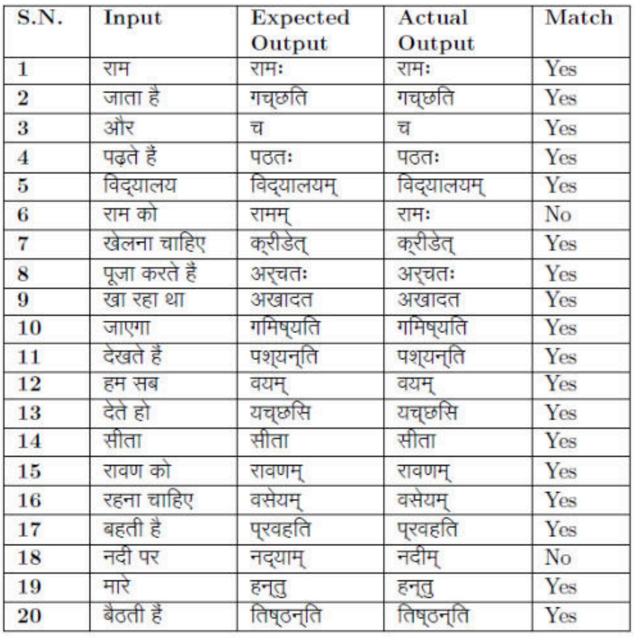
[2] provide an overview of Machine Learning giving an insight to why it is the future of computing industry. Ant Colony Optimization algorithm and its variants are reviewed based on various categories of problems they are applied to by [27]. [19] highlights various state-of-the-art Virtual Reality (VR) & Augmented Reality (AR) technologies that will prove to be beneficial for tourism and hospitality industry. A cumulative analysis of multiple text document classification algorithms has been discussed by [25].

Table 2.1
Steps followed by HTS

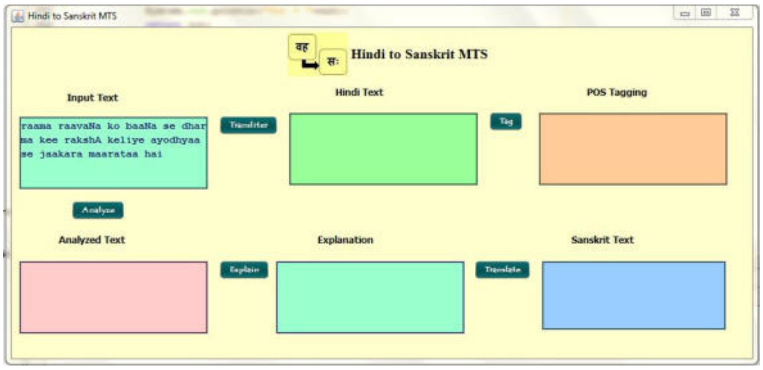


Step	Description	Output
1	Input text provided to the system	raama raavaNa ko baaNa se dharma kee rakshA keliye ayodhyaa se jaakara maarataa hai |
2	Transliterat- ion of the input text	राम रावण को बाण से धर्‌म की रक्‌षा के–लिये अयोध्या से जाकर मारता है |
3	POS tagging of input text	राम /NNPC रावण/NNP को/PSP बाण/NN से/PSP धर्‌म/NN की/PSP रक्‌षा/NNPC के लिये/NN अयोध्या/NNP से/PSP जाकर/VM मारता/NN है/VAUX
4	Semantic analysis of input text	राम ने रावण को बाण से धर्‌म की रक्‌षा के लिये अयोध्या से जाकर मारता है |
5	Explanation of the analysis	राम / पुल्‌लिन्‌ग् / १ ने / संबंध सूचक अव्‌यय रावण / पुल्‌लिन्‌ग् / २ को / संबंध सूचक अव्‌यय बाण / बाण / ३ से / संबंध सूचक अव्‌यय धर्‌म / पुल्‌लिन्‌ग् / ६ की / संबंध सूचक अव्‌यय रक्‌षा / स्‌त्‌रिलिन्‌ग् / ४ के लिये / संबंध सूचक अव्‌यय मारता है / पुल्‌लिन्‌ग् लट्‌लकार एकवचन् / प्‌रथम
6	Final translation	रामः रावणम् बाणेन् धर्‌मस्‌य रक्‌षायै अयोध्‌यायाः गत्‌वा हन्‌ति |

Table 2.2
Sample for Semantic Analysis of tokens (Hindi words & phrases)



S.N.	Input	Expected Output	Actual Output	Match
1	राम	रामः	रामः	Yes
2	जाता है	गच्‌छति	गच्‌छति	Yes
3	और	च	च	Yes
4	पढ़ते हैं	पठतः	पठतः	Yes
5	विद्‌यालय	विद्‌यालयम्	विद्‌यालयम्	Yes
6	राम को	रामम्	रामः	No
7	खेलना चाहिए	क्‌रीडेत्	क्‌रीडेत्	Yes
8	पूजा करते हैं	अर्‌चतः	अर्‌चतः	Yes
9	खा रहा था	अखादत	अखादत	Yes
10	जाएगा	गमिष्‌यति	गमिष्‌यति	Yes
11	देखते हैं	पश्‌यन्‌ति	पश्‌यन्‌ति	Yes
12	हम सब	वयम्	वयम्	Yes
13	देते हो	यच्‌छसि	यच्‌छसि	Yes
14	सीता	सीता	सीता	Yes
15	रावण को	रावणम्	रावणम्	Yes
16	रहना चाहिए	वसेयम्	वसेयम्	Yes
17	बहती है	प्‌रवहति	प्‌रवहति	Yes
18	नदी पर	नद्‌याम्	नदीम्	No
19	मारे	हन्‌तु	हन्‌तु	Yes
20	बैठती हैं	तिष्‌ठन्‌ति	तिष्‌ठन्‌ति	Yes



Input Text

raama raavaNa ko baaNa se dhar
ma kee rakshA keliye ayodhyaa
se jaakara maarataa hai

Hindi Text

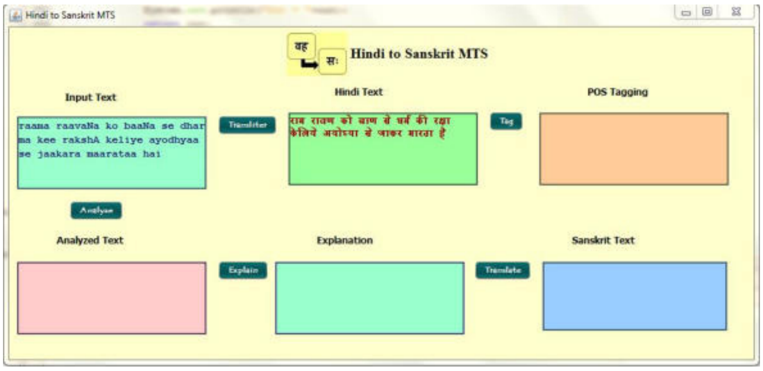
POS Tagging

Analyzed Text

Explanation

Sanskrit Text

FIG. 2.1. *Providing input text to the system*



Input Text

raama raavaNa ko baaNa se dhar
ma kee rakshA keliye ayodhyaa
se jaakara maarataa hai

Hindi Text

राम रावण को बाण से धर्म की रक्षा
केलिये अयोध्या से जाकर मारता है

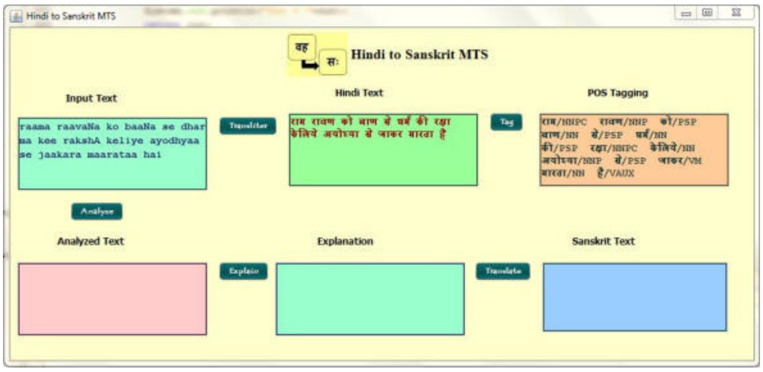
POS Tagging

Analyzed Text

Explanation

Sanskrit Text

FIG. 2.2. *Transliteration of input text*



Input Text

raama raavaNa ko baaNa se dhar
ma kee rakshA keliye ayodhyaa
se jaakara maarataa hai

Hindi Text

राम रावण को बाण से धर्म की रक्षा
केलिये अयोध्या से जाकर मारता है

POS Tagging

राम/NNPC रावण/NNP को/PSP
बाण/NN से/PSP धर्म/NN
की/PSP रक्षा/NNPC केलिये/NN
अयोध्या/NNP से/PSP जाकर/VM
मारता/NN है/VAUX

Analyzed Text

Explanation

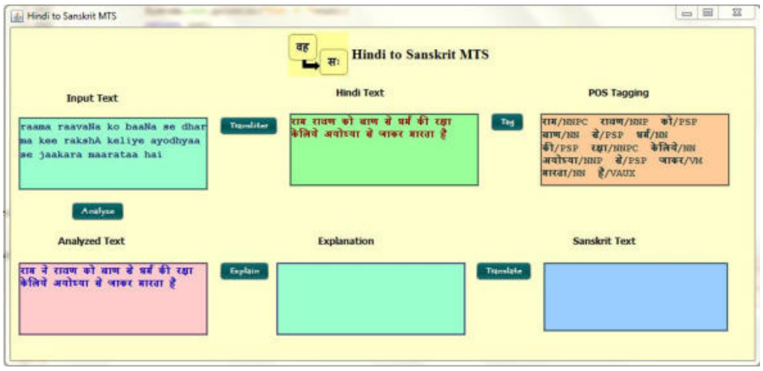
Sanskrit Text

FIG. 2.3. *POS tagging of input text*

TABLE 2.3
Confusion matrix for Semantic Analysis

	Output tokens			
		C	**I**	**Total**
	C	74	06	80
Input tokens	**I**	00	20	20
	Total	74	26	100

***** **C** - Correct **I** - Incorrect



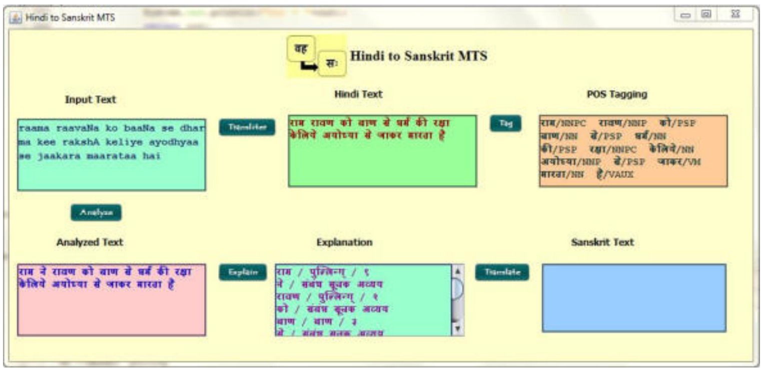
raama raavaNa ko baaNa se dhar
ma kee rakshA keliye ayodhyaa
se jaakara maarataa hai

राम रावण को बाण से धर्म की रक्षा
केलिये अयोध्या से जाकर मारता है

राम/NNPC रावण/NNP को/PSP
बाण/NN से/PSP धर्म/NN
की/PSP रक्षा/NNPC केलिये/NN
अयोध्या/NNP से/PSP जाकर/VM
मारता/NN है/VAUX

राम ने रावण को बाण से धर्म की रक्षा
केलिये अयोध्या से जाकर मारता है

FIG. 2.4. *Semantic analysis of input text*



raama raavaNa ko baaNa se dhar
ma kee rakshA keliye ayodhyaa
se jaakara maarataa hai

राम रावण को बाण से धर्म की रक्षा
केलिये अयोध्या से जाकर मारता है

राम/NNPC रावण/NNP को/PSP
बाण/NN से/PSP धर्म/NN
की/PSP रक्षा/NNPC केलिये/NN
अयोध्या/NNP से/PSP जाकर/VM
मारता/NN है/VAUX

राम ने रावण को बाण से धर्म की रक्षा
केलिये अयोध्या से जाकर मारता है

राम / पुल्लिन्ग् / १
ने / संबंध सूचक अव्यय
रावण / पुल्लिन्ग् / १
को / संबंध सूचक अव्यय
बाण / बाण / ३
से / संबंध सूचक अव्यय

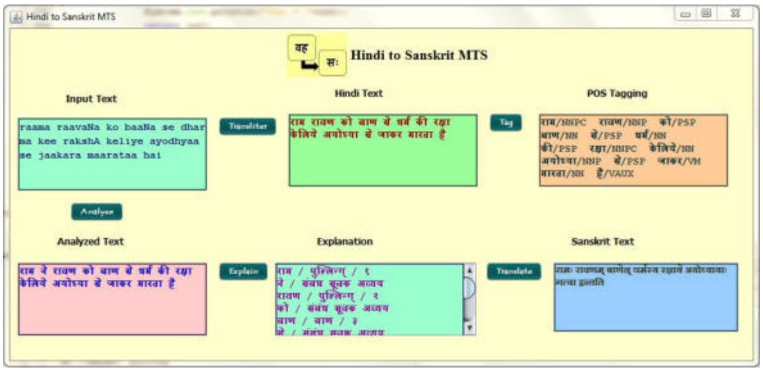
FIG. 2.5. *Explanation of the analysis*

3. Proposed Methodology. The Hindi to Sanskrit Translation system (HST) is a rule-based model. The prototype model that we have developed is a machine translation system from Hindi to Sanskrit language and it makes use of the rule-based approach. A rule-based machine translation system (RBMT) makes use of dictionaries (bilingual or multilingual) and grammars of both the source and target languages. This linguistic information contains the semantic, syntactical and morphological patterns of each of the languages. An RBMT system can either be a dictionary based MT, a transfer based MT or an Interlingua. Our system is a direct system (dictionary based MT) which maps input (Hindi text) to output (Sanskrit text) with basic rules. We have developed a Java application for implementation purposes. The system is fast and easy to implement. Moreover, the accuracy and performance can be further improved by the addition of more complex rules to cover the overall features of the language.

The proposed model takes in as input a Hindi text, processes it and produces the corresponding Sanskrit text as output as shown in figure 2.6. The processing phase is divided into multiple modules. These modules are described in the following subsections and algorithm 2 presents the steps involved:

3.1. Transliteration module. Transliteration refers to the process of converting a set of characters that are in the source language to a set of characters that are in the target language which have similar pronunciation. In other words, the phonetic similarity of the two languages is taken into account for transliteration. We have used a set of rules to perform this mapping from Latin to Devanagari script giving a 100% accuracy based on [12].

3.2. POS tagging module. POS (Part-Of-Speech) tagging is referred to as the task of assigning every word in a text to a particular part of speech such as, noun, pronoun, verb, adjective, adverb, etc. This tagging depends upon the actual meaning and the context of the word in the text. We have used the RDR (Ripple Down Rule based) POS tagger which uses an error-driven approach to construct tagging rules in the form of



Input Text

raama raavaNa ko baaNa se dharma kee rakshA keliye ayodhyaa se jaakara maarataa hai

Hindi Text

राम रावण को बाण से धर्म की रक्षा केलिये अयोध्या से जाकर मारता है

POS Tagging

राम/NNPC रावण/NNP को/PSP बाण/NN से/PSP धर्म/NN की/PSP रक्षा/NNPC केलिये/NN अयोध्या/NNP से/PSP जाकर/VM मारता/NN है/VAUX

Analyzed Text

राम ने रावण को बाण से धर्म की रक्षा केलिये अयोध्या से जाकर मारता है

Explanation

राम / पुल्लिन्ग् / १
ने / संबंध सूचक अव्यय
रावण / पुल्लिन्ग् / २
को / संबंध सूचक अव्यय
बाण / बाण / ३
से / संबंध सूचक अव्यय

Sanskrit Text

रामः रावणम् बाणेन् धर्मस्य रक्षार्थे अयोध्यायाः गत्वा हन्ति

Fig. 2.6. *Final translation*

Table 2.4
Sample for Pragmatic Analysis of Hindi sentences

Sentence	Actual Input	Ideal Input	Expected Translation	Actual Translation	Match
S-1	सीता पुस्तक पढ़ती है		सीता पुस्तक पढ़ती है		सीता पुस्तकम् पठति		सीता पुस्तकम् पठति		Yes
S-2	राम को खेलना चाहिए		राम को खेलना चाहिए		रामः क्रीडेत्		रामः क्रीडेत्		Yes
S-3	राम और श्याम पूजा करते हैं		राम और श्याम पूजा करते हैं		रामः श्यामः च अर्चतः		रामः श्यामः च अर्चतः		Yes
S-4	राम रावण को मारे		राम रावण को मारे		रामः रावणम् हन्ततु		रामः रावणम् हन्ततु		Yes
S-5	गीता नदी पर जाती है		गीता नदी पर जाती है		गीता नद्याम् गच्छति		गीता नदीम् गच्छति		No
S-6	राम फल खा रहा था		राम फल खा रहा था		रामः फलम् अखादत		रामः फलम् अखादत		Yes
S-7	राम विद्यालय जाएगा		राम विद्यालय जाएगा		रामः विद्यालयम् गमिष्यति		रामः गमिष्यति		Yes
S-8	राम, श्याम और सीता विद्यालय जाते हैं		राम, श्याम और सीता विद्यालय जाते हैं		रामः, श्यामः सीता च विद्यालयम् गच्छन्ति		रामः, श्यामः सीता च विद्यालयम् गच्छन्ति		Yes
S-9*	तुम दोनो जाना चाहिए		तुम दोनो को जाना चाहिए		युवाम् गच्छेतम्		युवाम् गच्छेतम्		Yes
S-10*	राम और श्याम जाता है		राम और श्याम जाते हैं		रामः श्यामः च गच्छतः		रामः गच्छति		No

* - Semantically incorrect sentence

a binary tree [20]. The tagger produces an accuracy of 95.77 for Hindi language which is a key factor for our translation system.

3.3. Rule-base engine. We have created a database to store rules for translation from Hindi to Sanskrit text. The database consists of multiple rules which are being used to identify and map verbs, nouns, etc. This is done by identifying the number (vachana), gender, case and person (purusha). The conversion from source to target language takes place by comparing, mapping and identifying the corresponding values from the corpus.

3.4. Root Verb extraction module. This module outputs the equivalent Sanskrit verb for the input Hindi verb. The root verb depends upon person, gender and number of the noun to which the verb corresponds to. The algorithm first tries to extract the person, number and case for the verb. Then, it extracts the root verb, compares and finds an appropriate mapping from Hindi to Sanskrit.

Table 2.5
Confusion matrix for Pragmatic Analysis

		Output tokens		
		C	**I**	**Total**
	C	30	05	35
Input tokens	**I**	02	13	15
	Total	32	18	50

*** C** - Correct **I** - Incorrect

4. Implementation and Results. We have implemented our Hindi to Sanskrit translation (HST) system in Windows using Java. A GUI (Graphical User Interface) has been developed wherein the user is prompted to enter the Hindi sentence as input from the keyboard (in Latin script) and the end result, in the form of Sanskrit sentence, is displayed on the interface. The input text is first transliterated to phonetically similar Devanagari script, Ripple Down Rule-based Parts Of Speech (RDR-POS) tagging is performed to identify different parts of speech, input sentence is then analyzed semantically and finally, the translation to Sanskrit is performed. It also shows the result after transliteration and POS tagging (using RDR POS tagger) of the input Hindi sentence before translation to Sanskrit. Various steps followed by the system are explained and shown in figure 2.1, 2.2, 2.3, 2.4, 2.5, and 2.6. The description and output of every step is presented concisely in table 2.1.

Semantic analysis is applied to tokens which eventually combine to form sentences. Out of 100 tokens, 80 are correct for which a sample is displayed in the form of table 2.2. The 20 incorrect tokens are not recognized by the system. Since no output is produced for them they are not displayed in the table. Table 2.3 depicts the confusion matrix created for the analysis purpose.

Accuracy = (74+20)/100 = 94/100 = 0.940

Error Rate = (6+0)/100 = 6/100 = 0.060

Pragmatic analysis is performed on 50 sentences, out of which 35 are correct and 15 are incorrect. The sample for analysis is displayed in table 2.4. Table 2.5 depicts the confusion matrix created for the analysis purpose.

Accuracy = (30+13)/50 = 43/50 = 0.860

Error Rate = (5+2)/50 = 7/50 = 0.140

5. Conclusions and Future Work. The work proposed above strives to translate different kinds of possible Hindi sentences to equivalent Sanskrit sentences. All of these modules provide a deep and thorough understanding of the interaction between the two languages and their translation process. The rule-based approach can explain the detailed comparison of the two languages of interest and the logical process of how we reach a particular result. These features make this translation system an excellent tool for study involving languages and their interaction. Apart from being a learning tool, it can be used as a teaching pedagogy tool. This will be beneficial for the students and teachers who are involved in the process of learning and teaching Sanskrit language.

In future, we can try to translate interrogative and more complex type of sentences. We can also attempt to use other approaches to machine translation apart from the rule based approach and compare the results. The system can be enhanced to support voice translation taking speech (in source language) as input and producing speech (in target language) as output. We can try to minimize the processing time and the memory requirements of the translator so that the use of the computer resources is optimized.

REFERENCES

[1] P. Agrawal, *A Machine Translation System for Sanskrit to Hindi language*, PhD thesis, IKG-Punjab Technical University, 2018.

[2] J. Alzubi, A. Nayyar, and A. Kumar, *Machine learning from theory to algorithms: An overview*, Journal of Physics: Conference Series, 1142 (2018), p. 012012, https://doi.org/10.1088/1742-6596/1142/1/012012.

[3] P. Bahadur, A. Jain, and D. S. Chauhan, *Architecture of english to sanskrit machine translation*, 2015 SAI Intelligent Systems Conference (IntelliSys), (2015), https://doi.org/10.1109/intellisys.2015.7361204.

[4] R. Begum, S. Husain, L. Bai, and D. M. Sharma, *Developing verb frames for hindi*, in LREC, 2008.

[5] G. o. I. Department of Official Language, *Constitutional Provisions: Official Language Related Part-17 of The Constitution Of India*, 2015.

[6] S. Ebrahim, D. Hegazy, M. G. M. Mostafa, and S. R. El-Beltagy, *English-arabic statistical machine translation: State of the art*, Computational Linguistics and Intelligent Text Processing Lecture Notes in Computer Science, (2015), p. 520–533, `https://doi.org/10.1007/978-3-319-18111-0_39`.

[7] M.L. Forcada, M. Ginesti-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J.A. Pérez-Ortiz, F. Sánchez-Martinez, G. Ramí rez-Sá nchez, F.M. Tyers, *Apertium: a free/open-source platform for rule-based machine translation*, Machine Translation 25 (2) (2011), p. 127–144, `https://doi.org/10.1007/s10590-011-9090-0`.

[8] P. Goyal and R. M. K. Sinha, *Translation divergence in english-sanskrit-hindi language pairs*, Lecture Notes in Computer Science Sanskrit Computational Linguistics, (2008), p. 134–143, `https://doi.org/10.1007/978-3-540-93885-9_11`.

[9] V. K. Gupta, N. Tapaswi, and S. Jain, *Knowledge representation of grammatical constructs of sanskrit language using rule based sanskrit language to english language machine translation*, 2013 International Conference on Advances in Technology and Engineering (ICATE), (2013), `https://doi.org/10.1109/icadte.2013.6524744`.

[10] A. Hurskainen and J. Tiedemann, *Rule-based machine translation from english to finnish*, Proceedings of the Second Conference on Machine Translation, (2017), `https://doi.org/10.18653/v1/w17-4731`.

[11] L. Jain and P. Agrawal, *Text independent root word identification in hindi language using natural language processing*, International Journal of Advanced Intelligence Paradigms, 7 (2015), p. 240, `https://doi.org/10.1504/ijaip.2015.073705`.

[12] L. Jain and P. Agrawal, *English to sanskrit transliteration: an effective approach to design natural language translation tool*, International Journal of Advanced Research in Computer Science (IJARCS, 8 (2017), pp. 1–10, `https://doi.org/https://doi.org/10.26483/ijarcs.v8i1.2860`.

[13] J. K. and J. R., *Sanskrit machine translation systems: A comparative analysis*, International Journal of Computer Applications, 136 (2016), p. 1–4, `https://doi.org/10.5120/ijca2016908290`.

[14] A. Kumar, Saurabh, and M. Raza, *Syntax and semantic analysis of devanagari hindi*, International Journal of Recent Scientific Research, 8 (2017).

[15] V. Macketanz, E. Avramidis, A. Burchardt, J. Helcl, and A. Srivastava, *Machine translation: Phrase-based, rule-based and neural approaches with linguistic evaluation*, Cybernetics and Information Technologies, 17 (2017), p. 28–43, `https://doi.org/10.1515/cait-2017-0014`.

[16] S. Mall and U. C. Jaiswal, *Developing a system for machine translation from hindi language to english language*, 2013 4th International Conference on Computer and Communication Technology (ICCCT), (2013), `https://doi.org/10.1109/iccct.2013.6749607`.

[17] V. Mishra and R. Mishra, *English to sanskrit machine translation system: a rule-based approach*, International Journal of Advanced Intelligence Paradigms, 4 (2012), p. 168, `https://doi.org/10.1504/ijaip.2012.048144`.

[18] V. Mishra and R. B. Mishra, *Study of example based english to sanskrit machine translation*, Polibits, 37 (2008), p. 43–54, `https://doi.org/10.17562/pb-37-5`.

[19] A. Nayyar, B. Mahapatra, D. N. Le, and G. Suseendran, *Virtual reality (vr) & augmented reality (ar) technologies for tourism and hospitality industry*, International Journal of Engineering & Technology, 7 (2018).

[20] D. Q. Nguyen, D. Q. Nguyen, D. D. Pham, and S. B. Pham, *A robust transformation-based learning approach using ripple down rules for part-of-speech tagging*, AI Communications, 29 (2016), p. 409–422, `https://doi.org/10.3233/aic-150698`.

[21] W. P. Pa, Y. K. Thu, A. Finch, and E. Sumita, *A study of statistical machine translation methods for under resourced languages*, Procedia Computer Science, 81 (2016), p. 250–257, `https://doi.org/10.1016/j.procs.2016.04.057`.

[22] M. Parkvall, *Världens 100 största språk 2007 (the world's 100 largest languages in 2007)*, 2007.

[23] S. Ramteke, K. Ramteke, and D. R., *Lexicon parser for syntactic and semantic analysis of devanagari sentence using hindi wordnet*, International Journal of Advanced Research in Computer and Communication Engineering, 3 (2014).

[24] S. Saini and V. Sahula, *A survey of machine translation techniques and systems for indian languages*, 2015 IEEE International Conference on Computational Intelligence and Communication Technology, (2015), `https://doi.org/10.1109/cict.2015.123`.

[25] S. S. Sehra and A. Nayyar, *Paper on algorithms used for text classification*, 2013.

[26] N. Sethi, P. Agrawal, V. Madaan, and S. K. Singh, *A novel approach to paraphrase hindi sentences using natural language processing*, Indian Journal of Science and Technology, 9 (2016), `https://doi.org/10.17485/ijst/2016/v9i28/98374`.

[27] R. Singh and A. Nayyar, *Ant colony optimization — computational swarm intelligence technique*, (2016).

[28] H. Somers, *Review article: Example-based machine translation*, Machine Translation, 14 (1999), pp. 113–157, `https://doi.org/10.1023/A:1008109312730`, `https://doi.org/10.1023/A:1008109312730`.

[29] S. Sreelekha, P. Bhattacharyya, and D. Malathi, *Statistical vs. rule-based machine translation: A comparative study on indian languages*, Advances in Intelligent Systems and Computing International Conference on Intelligent Computing and Applications, (2017), p. 663–675, `https://doi.org/10.1007/978-981-10-5520-1_59`.

Edited by: Anand Nayyar
Received: Jun 21, 2020
Accepted: Jul 27, 2020

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2020 SCPE. Volume 21, Issues 3, pp. 555–568, DOI 10.12694:/scpe.v21i3.1749

NEEF: A NOVEL ENERGY EFFICIENT FUZZY LOGIC BASED CLUSTERING PROTOCOL FOR WIRELESS SENSOR NETWORK

ANSHU KUMAR DWIVEDI*AND A.K. SHARMA

Abstract. The uttermost requirement of the wireless sensor network is prolonged lifetime. Unequal energy degeneration in clustered sensor nodes lead to the premature death of sensor nodes resulting in a lessened lifetime. Most of the proposed protocols primarily choose cluster head on the basis of a random number, which is somewhat discriminating as some nodes which are eligible candidates for cluster head role may be skipped because of this randomness. To rule out this issue, we propose a deterministic novel energy efficient fuzzy logic based clustering protocol (NEEF) which considers primary and secondary factors in fuzzy logic system while selecting cluster heads. After selection of cluster heads, non-cluster head nodes use fuzzy logic for prudent selection of their cluster head for cluster formation. NEEF is simulated and compared with two recent state of the art protocols, namely SCHFTL and DFCR under two scenarios. Simulation results unveil better performance by balancing the load and improvement in terms of stability period, packets forwarded to the base station, improved average energy and extended lifetime.

Key words: Energy Efficiency, Wireless Sensor Network, Clustering, Fuzzy Logic, Cluster Head

AMS subject classifications. 68M11, 94D05

1. Introduction. In recent past decades, wireless sensor network (WSN) has emerged as a vital part of our daily life. With the drastic progression of microelectronics technology which consumes low power in electronic circuitry, WSN is applied in diverse real-time applications like commercial monitoring, healthcare sensing, surrounding monitoring, battlefield surveillance etc. [1]. WSN contains sensor nodes (SN) which can experience, accumulate and compute information from the environment and also maintain it for a protracted time frame. A WSN is a blend of four subsystems altogether with a sensing module, a transceiver module, a processing module, and a power supply module (battery) [2]. These SNs use battery for energy delivery, limited memory for collecting information from the vicinity in which these are deployed and also microprocessor for processing the data and later transferring it to the base station (BS). So, it has been a premier issue for researchers to devise a mechanism to use the energy of SN effectively. WSN can be classified as heterogeneous or homogeneous networks [3]. In a homogeneous network, all nodes possess equal capacities in terms of processing, memory, radio range and energy, whereas in a heterogeneous network, it may be different. Clustering is one of the better solutions for sparing power of SNs and extending network lifetime [4]. Clustering is a method wherein all the SNs are grouped according to some criteria and each group is headed by one of the nodes called a cluster head (CH) [5]. The CH compresses the data supplied by their cluster members (CMs) via statistics fusion to reduce the redundancy and improving the power dissipation rate of the network.

Various clustering algorithms were proposed in the last decades like LEACH [4], PEGASIS [6], HEED [7], SEP [8], LEASE [9], EDFCM [10], SPEZ [11] etc. In WSN, there are two types of information gathering schemes: Hierarchical and Non-Hierarchical. In a hierarchical scheme, SNs communicate the records to the BS via CH in one hop while, in a non-hierarchical scheme, SNs send the records in single and/or multi-hop to the BS via a CH resulting in conserving more energy. Most of the clustering protocols rotate the CH role so that the energy dissipation can be balanced in the network. However, regardless of the dynamic rotation of the CH role, the energy imbalance takes place due to communication distance between SN and the CH. The location of the BS additionally influences the lifetime of the network as a longer distance will use more energy for communication. If the chosen CHs are nearer to the BS then it will dissipate less energy and if the CHs are at distant place then it will drastically deplete energy level. Some researchers have proposed multi-hop

*Madan Mohan Malviya University of Technology, Gorakhpur, India (anshucse.dwivedi@gmail.com)

communication to the BS but it also depletes energy of CH nearer to the BS.

In this paper, we have propound a novel energy efficient fuzzy logic (NEEF) based clustering protocol for WSN that makes use of designed fuzzy system for energy efficacy while selecting the CHs thereby protracting network lifetime. The main contributions are highlighted as follows:

- The influential parameters that affect the battery level of SNs are identified from related work and segregated into primary and secondary factors.
- The primary factors considered are remnant energy level and communication cost to be borne by CH, whereas secondary factors are the density of SN and its aloofnes to the BS.
- To emphasise the election of best suitable node for the CH role, weights are assigned to Fuzzy fitness output value (FF1 and FF2) after experimental evaluation through simulation.
- To balance the load of CH, non-CH nodes choose their CH on the basis of chance obtained from designed fuzzy system which considers the load of CH node and distance from non-CH node to the considered CH.

Subsequent part of this paper is organised as follows: Section 2 discusses literature survey. System model with network and energy dissipation model is discussed in section 3. A description about NEEF protocol is presented in section 4. Simulation experiment and result analysis is done in section 5 and section 6 provides concluding remarks. .

2. Relevant Work. This section discusses some pertinent clustering algorithms in WSN. Maximal clustering algorithms use rounds to describe the lifetime of WSN. Each round consists of CH selection, formation of cluster and the data collection. More the number of rounds, the longer will be the lifetime of WSN. LEACH [4] is a pioneering protocol in clustering algorithms. The goal of LEACH is to choose a node as CH in such a manner that every node gets an opportunity to become a CH. The reason is that a CH node dissipates higher energy than non-CHs nodes, therefore, a node will not dissipate power by turning into the CH repeatedly. It uses randomness in selecting CH, which may converge to no CH in a round. Gupta et al. implemented fuzzy logic for clustering of SN in WSN [12]. This work is an improvement over LEACH. The inputs for fuzzy systems are node degree, centrality and residual energy. However, it uses centralised approach by making use of BS for clustering. Centralised approaches are not easily scalable because of dependency on BS. The CH election mechanism relying on the Fuzzy reasoning (CHEF) was proposed by Kim et al. [13]. This protocol determines probability of node to act as CH. It utilises the transmission range to the BS and the remnant node energy as fuzzy elements for the CH selection, unlike LEACH. The contrast between LEACH and CHEF shows more effective cluster formation in CHEF than on LEACH.

LEACH-FL is an enhanced LEACH variant with Fuzzy Logic [14]. It differs from LEACH in term of the factors used, viz. distance of node from the sink, type of battery used, and the density of a node. Selection of CH is a centralised process, similar to LEACH, handled by the BS, which computes the probability of a SN to become a CH. The authors of LEACH-FL have shown through experiments that the suggested protocol has reduced the energy dissipation rate. The lifetime of network using LEACH-FL protocol exceeds the lifespan of the network while considering LEACH. Lee and Chen have [15] propound a fuzzy logic-based clustering strategy in which a CH node is elected on basis of outstanding energy of a node and the expected outstanding energy. SEP-FL [16] is an enhanced variant of SEP [8], centred on the choice of CH by adjusting the remaining energy probabilities for each node. It offers a larger duration of stabilisation and a reduced duration of disturbance thereby improving node lifetime. The method is based on each node's distance from the BS and remnant energy level. EAUCF [17] is a fuzzy based unequal clustering approach. It proposes to lessen the energy depletion of CHs in pairs as they are either near to the BS or possibly have limited battery power left. EAUCF has a stronger output in terms of first node death (FND), quarter node death (QND) and relatively lower energy depletion in contrast to LEACH, CHEF [13] and EEUC [18].

MOFCA [19] is another technique of clustering in mobile sensor networks. Based on range to node and residual energy, CHs are determined. The radius of the CH is very important in relation to opportunity, which means that if a CH is closer to the BS and has more energy, it can gather and communicate more information. An enhanced variant of EAUCF [20] is FBUC [21] or Fuzzy Based Unequal Clustering. In addition, FBUC utilises a probabilistic limit function instead of a predefined limit number as compared to EAUCF [17] and provides a fuzzy input variable called degree of node which is used to select the CH during cluster radius contest.

In different scenarios, FBUC exhibit lower power dissipation and longer network life than its counterpart in terms of first node death (FND) and last node death (LND). For a WSN to enhance lifetime, a CH selection scheme focused on fuzzy logic and particle swarm optimisation is suggested in [20]. DUCF [22] is another technique that uses fuzzy logic for clustering. Energy, proximity to the BS, and node temperature are regarded to be fuzzy inputs, and all nodes are selected in each round. This technique considers the most suitable node as a CH reducing energy consumption. FBECS [23], on the other hand, considers energy of SN, distance from the BS and density as an input to the FIS for selection of CH. It considers zonal network structure and corresponding probability during the clustering process. FBECS is capable of extending lifetime and stability time. DFLC [24] is another protocol based on the fuzzy logic that is carried out on nodes within the network in a distributed manner. DFLC considers network like a tree where it's nodes can be BS, cluster member and CH. DFLC is compared to ACAWT [25],LEACH, and CHEF [13]. Experimental results demonstrate that DFLC exhibit better performance than other algorithms in terms of chosen performance metrics. ECPF [26] also makes use of the fuzzy logic. Three procedures have been used to extend the life expectancy of the network. For fuzzy logic based calculation, ECPF utilises node degree and node centrality as fuzzy inputs for generating output for CH election.

SCHFTL [27], is based on fuzzy logic system, in which the sensor node uses different parameter at different levels. The first level parameters are remaining energy and centrality, the second level parameters used are communication quality and distance from the BS and the third level are total energy and DOS attack. With the help of this parameter, super cluster head is selected out of the chosen CH. This protocol avoids the data overload, data loss and data retransmission, thereby increases the network life span. DFCR [28] routing protocol is propound that applies unequal clustering mechanism to solve the hotspot problem of WSN by minimising the size of cluster, that are closest to the BS. E-CAFL [29] is another routing protocol proposed to enhance CAFL [30] protocol by allowing node density. It uses three parameters, viz. distance from sink, remaining energy and density of node as input for FIS for estimation of rank for selecting the CH.

In the above mentioned protocols, emphasis is not given to the parameters that majorly affect the energy level of SNs. We have determined the influential parameters and used them in fuzzy logic for best SNs selection in network. In most of the protocols mentioned in related work, they do not consider efficient cluster formation mechanism. We have fuzzy fitness value obtained during CH selection with other influential parameters while forming the clusters in our porposed work.

3. Preliminaries.

This section presents the assumptions made in the network model for the proposed work in line with the energy dissipation model.

3.1. Network Model.

Major presumptions that are made for the network are:

- SNs are randomly arranged in target area.
- After the deployment, all the nodes are stationary.
- BS doesn't have energy constraint.
- Deployed nodes are homogeneous in terms of resources.
- Every node has only one CH.
- The communication link is symmetric.
- The distance between two SN is determined by RSSI (Received Signal Strength Index).
- Initially, SN are unacquainted about their location.

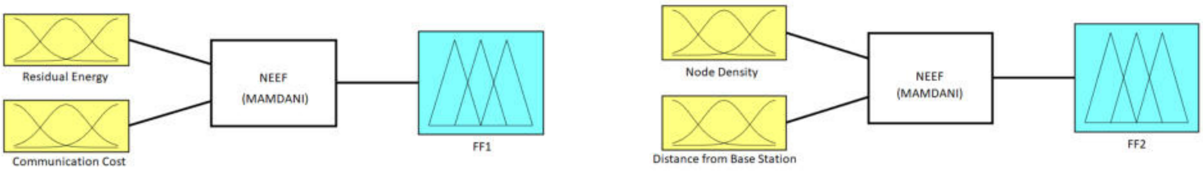
3.2. Energy dissipation model.

For the analysis of proposed work, the energy model adopted in [4] is employed. The energy of the network may be depleted in sensing, aggregation, amplification, transmission and reception. For transmitting and receiving s bits over d distance, energy dissipations are given by

$$E_{Tx}(s,d) = \begin{cases} sE_{elec} + s\epsilon_{fs}d^2, & d < d_o \\ sE_{elec} + s\epsilon_{mp}d^4, & d > d_o \end{cases} \tag{3.1}$$

where do is a threshold which determines either free space or multipath model adopted and it can be calculated by $d_o = \sqrt{\epsilon_{fs}/\epsilon_{mp}}$.

$$E_{Rx}(s) = E_{Rx-elec}(s) = sE_{elec} \tag{3.2}$$



(a) FF1 *(b) FF2*

Fig. 4.1. *Designed FIS*

In amplification of signal, energy dissipated is calculated by

$$E_{amp} = \begin{cases} \epsilon_{fs} d^2, & d < d_o \\ \epsilon_{mp} d^4, & d > d_o \end{cases} \tag{3.3}$$

The communication cost to be borne by a CH in a round is calculated by

$$E_{CH} = ns(E_{elec} + \epsilon_{fs} d_{BS} + E_{DA}) \tag{3.4}$$

The total energy exhausted by a member of a cluster is calculated by

$$E_{CM} = s(E_{elec} + \epsilon_{fs} d_{CH}) \tag{3.5}$$

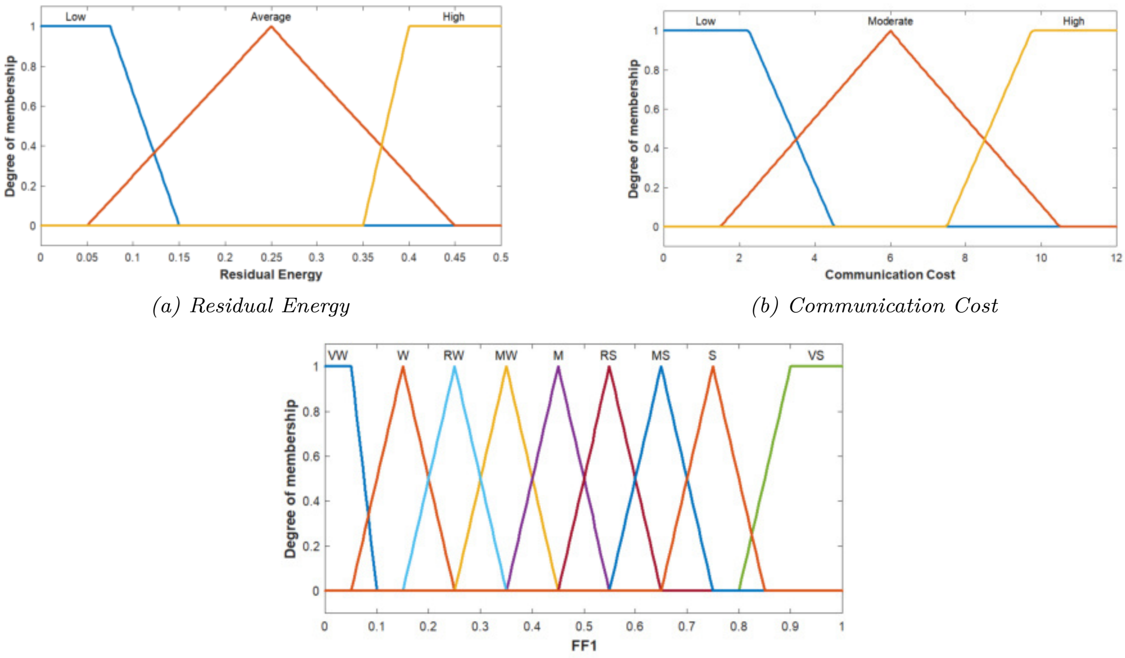
wherein d_{CH} is the distance to its CH.

4. Proposed NEEF protocol. The proposed NEEF protocol consists of four stages in a round i.e. *pre-deployment stage, CH selection stage, cluster formation stage and data dissemination stage.*

4.1. Pre-deployment stage. Before the deployment of SNs in the field, network administrator is required to allocate unique ID to the SNs. The information about the BS is also fed into the SN so that it can determine the BS during the operation of the network. For initial setup of the network and determining the neighbourhood, a TDMA slot is fed into each SN so that collision free broadcast can take place.

4.2. CH selelction stage. Once the deployment of SN is complete, it's time to select the optimal candidates to play the role of CH. Since the role of CH is very crucial, fuzzy logic is applied to determine the optimal candidate. Fuzzy logic is mostly applied to solve the uncertainties in any system. For efficient selection of CH in WSN, there are several overlapping factors like remnant energy, distance between node to the BS, density in neighbourhood, communication cost, etc. Thus, Fuzzy logic is appropriate to solve the optimal CH selection problem as it can blend various factors dealing with uncertainties and provide better results. Since, the nodes are unaware about the location of the BS, a Hello_PKT(BSID) is broadcast by the BS so that every SN can estimate the aloofness from the BS. The SNs will make a broadcast as per the TDMA slot provided in pre-deployment phase. Once all the SNs are aware of the required parameters (distance to the BS, remnant energy, communication cost and density around node), the computation for CH candidature begins at each SN. Two Fuzzy inference system (FIS) have been designed for computing the Fuzzy fitness values (FF1 and FF2) of SNs as shown in Fig. 4.1.

For FF1, residual energy and communication cost are chosen as the input variables. The linguistic variables (LV) chosen are Low(Lw), Average(Ag), High(Hg) and Low(Lw), Moderate(Md), High(Hg) respectively. LV for output variable are Very Weak (VW), Weak (W), Rather Weak (RW), Medium Weak (MW), Medium (Mm), Rather Strong (RStr), Medium Strong (MStr), Strong (VStr) and Very Strong (VStr). The membership function (MF) for different LV which are derived for input and output variables are shown in Fig. 4.2. The FIS processes these input variables on the basis of LV and establishes a functional relationship between input and output LV on the basis of set of IF-THEN mapping rules. These rules which are used for calculating the FF1 is depicted in Table 4.1. These IF-THEN rules are evaluated using Mamdani inference method [31] which we have also depicted in Fig. 4.1 and Fig. 4.5. The reason for using this method is its simplicity and ability to easily interpret and draw conclusion on the basis of given IF-THEN rules. For defuzzification, we have used



(a) Residual Energy

(b) Communication Cost

(c) Fuzzy Fitness Value (FF1)

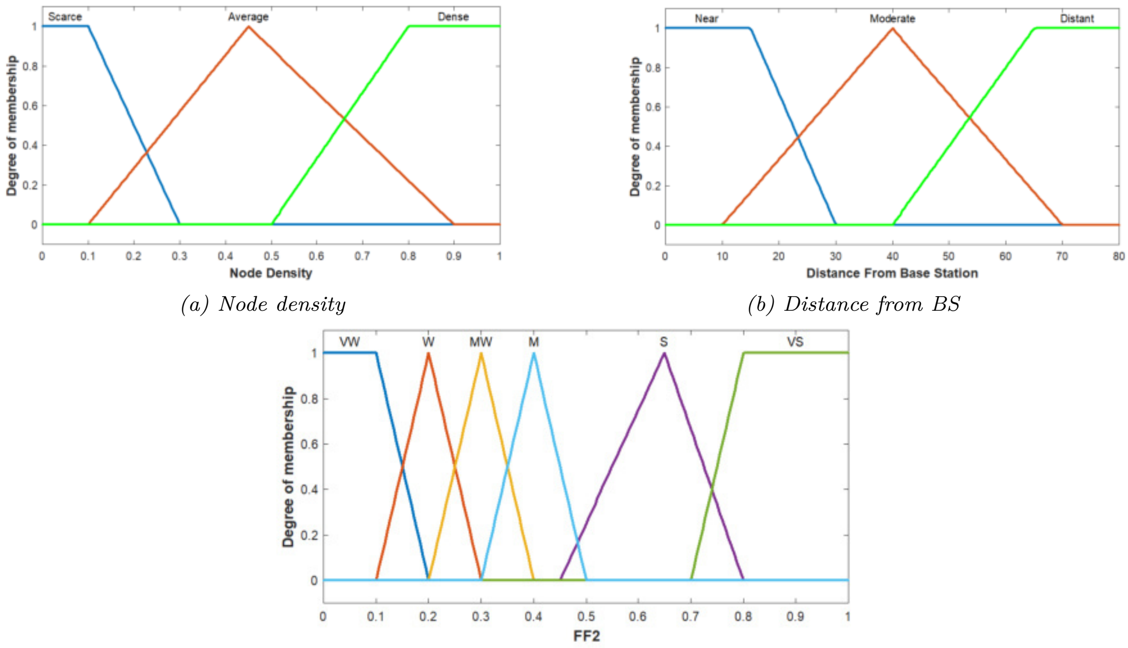
Fig. 4.2. *Membership Functions*

Table 4.1
Fuzzy Rules for FF1

Rule No.	Residual Energy	Communication cost	FF1
1	Lw	Hg	VW
2	Lw	Md	W
3	Lw	Lw	RW
4	Ag	Hg	MW
5	Ag	Md	Mm
6	Ag	Lw	RStr
7	Hg	Hg	MStr
8	Hg	Md	Str
9	Hg	Lw	VStr

COA method which is commonly used [30, 32]. For calculating the Fuzzy fitness value FF2, two parameters are considered: node density and distance from BS as shown in Fig. 4.1.

These two parameters have some significance during selection of CH candidate. Node density provides the estimation of neighbouring nodes which can reduce the intra-communication cost as more cluster members will lead to more dissipation of energy as well as coordination overhead. The objective is to distribute the load of CH role at par. Distance from BS determines the communication overhead which the CH has to borne for finally forwarding the data. If the number of neighbouring nodes is more and the BS is at distant place then more number of packets is to be forwarded to the BS which will deplete the energy of CH quickly. The LV for Node density is Scarce (Scr), Average (Ag) and Dense (Ds). Similarly, Near (Nr), Moderate (Md) and Distant (Dt) are the LV opted for Distance from the BS. For output variable FF2, the LV are Very Weak (VW), Weak(W), Medium Weak (MW), Medium(Mm), Strong (Str) and Very Strong(Str). Triangular and trapezoidal membership functions are chosen for interior values and boundary values respectively as shown in Fig. 4.3. The fuzzy IF-Then rules for mapping the input to output variables are depicted in Table 4.2.



(a) Node density

(b) Distance from BS

(c) Fuzzy Fitness Value (FF2)

Fig. 4.3. *Membership Functions*

Table 4.2
Fuzzy Rules for FF2

Rule No.	Node density	Distance from BS	FF2
1	Scr	Nr	MStr
2	Scr	Md	MW
3	Scr	Dt	MW
4	Ag	Nr	Str
5	Ag	Md	Mm
6	Ag	Dt	W
7	Ds	Nr	VStr
8	Ds	Md	Mtr
9	Ds	Dt	VW

After the computation of FF1 and FF2, every SN calculates its probability of being CH by

$$SN(k).prob = \alpha \times SN(k).FF1 + \beta \times SN(k).FF2 \qquad s.t.(\alpha + \beta) = 1 \tag{4.1}$$

where α and β are arbitrary constants.

We have considered the value of α and β as 0.7 and 0.3 respectively as we got better results with these values as shown in Fig. 4.4. While carrying out experimental analysis of values assigned to α and β , we calculated the average CH to network energy ratio for 500^{th} round with varying values of α and β. The reason for weightage of α more than β is that FF1 considers communication cost as well the remnant energy level of SN which are more influential in choosing efficient CH. After each node computes the probabilistic value for its CH candidature, each node broadcast its probability. The SNs with highest probability are selected as CH candidates. Only $p\%$ CHs are elected in each round. The process of selecting CH is defined in Algorithm 1. We have designed a FIS for computing chance of each CH as shown in Fig. 4.5.

There are three input variables: number of member nodes, communication distance to CH node and fuzzy fitness value FF1 of CH node. The output variable is Chance of CH which determines the probability of CH to be chosen by non-CH node. The LV for input and output variables are depicted in Table 4.3. Membership

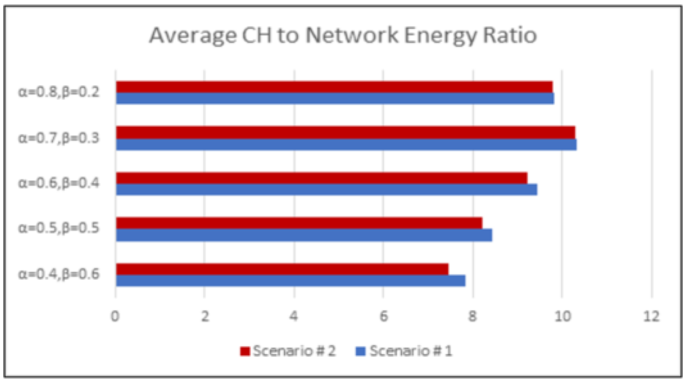


Fig. 4.4. *Experimental evaluation of α and β for average CH to network energy ratio*

Algorithm 1 Selection of CH

```
Tn ← Total nodes
k ← ID of SN
cluster_count ← 0
SN(k).Energy← current SN energy level
SN(k).ND ← Neighbouring nodes in communication range
SN(k).CC← communication cost if chosen as CH
SN(k).Type← N
Normal node
SN(i).DBS← Distance of SN to BS
for each node SN(k) do
    SN(k).FF1← Fuzzy(SN (k).Energy,SN (k). CC ) // Fitness Value1
    SN(k).FF2← Fuzzy(SN (k).ND,SN (k). DBS ) // Fitness Value2
    SN(k).Prob← $\alpha \times$ SN (k).FF1 + $\beta$ SN (k). FF2 ) // $\alpha$ and $\beta$ are arbitrary constants
end for
for each node SN(k) do
    Broadcast SN(k).Prob
    if SN(i).type == ”N” && SN(k).Energy>0 then
        if SN(i).Prob > rest of the nodes in Tn&& cluster count <p% then
            SN(i).Type ← ”C” //SN is now CH
            Count_CH++ //Increment the count of CHs
        end if
    end if
end for
```

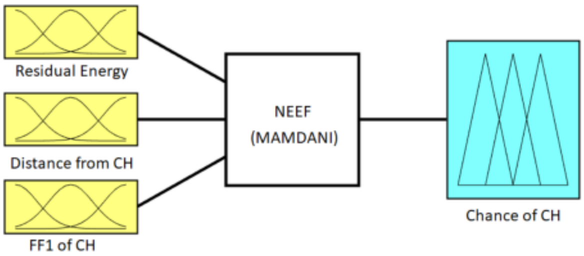
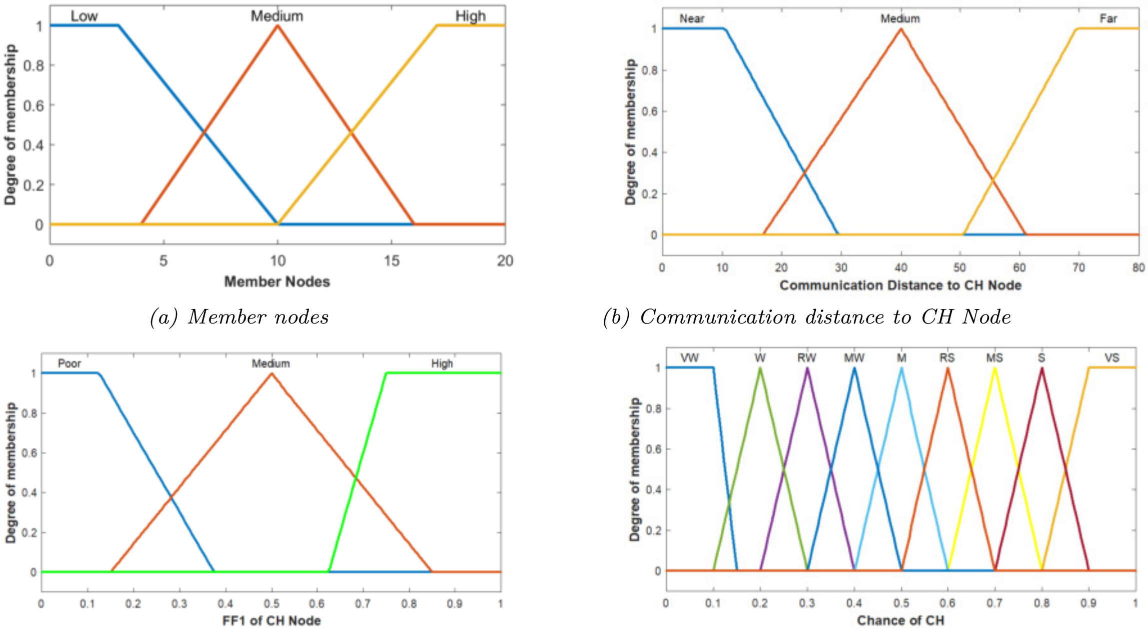


Fig. 4.5. *FIS for computing Chance of CH*

Table 4.3
Fuzzifier linguistic variables

Variable Name	Linguistic Variable
Member nodes	Low(Lw),Medium(Mm),High(Hg)
Communication distance to CH node	Far(Fr),Medium(Mm),Near(Nr)
Fuzzy Fitness(FF) value 1	Poor(Pr),Medium(Mm),High(Hg)
Chance of CH	Very Strong(VStr), Strong(Str), Medium Strong(MStr), Medium(Mm), Medium Weak(MW), Rather Weak(RW), Weak(W),Very Weak(VW)



(a) Member nodes

(b) Communication distance to CH Node

(c) Fuzzy Fitness Value (FF1)

(d) Chance of CH

Fig. 4.6. *Membership Functions*

functions of input and output variables are depicted in Fig. 4.6. IF-THEN rules for mapping input to output variables are depicted in Table 4.4. Each non-CH node computes the chance of each CH candidates and the CH candidate with maximum chance is chosen as optimum CH by the node. The non-CH node sends join request to the optimum CH node and receives acknowledgement from CH node with TDMA slot for data collection during the round. The cluster formation process is illustrated in Algorithm 2.

4.3. Data dissemination stage. Once the selection of CHs is completed and clusters are formed, a data dissemination stage comes into play. Data is generated from the target area on periodic basis. After sensing the target area, SN forwards the data to the CH as per the TDMA slot for collision free communication. Once the CH collects data from all the cluster members, it aggregates the data and transmits it to the BS for further processing. In this way, one round is concluded in NEEF protocol.

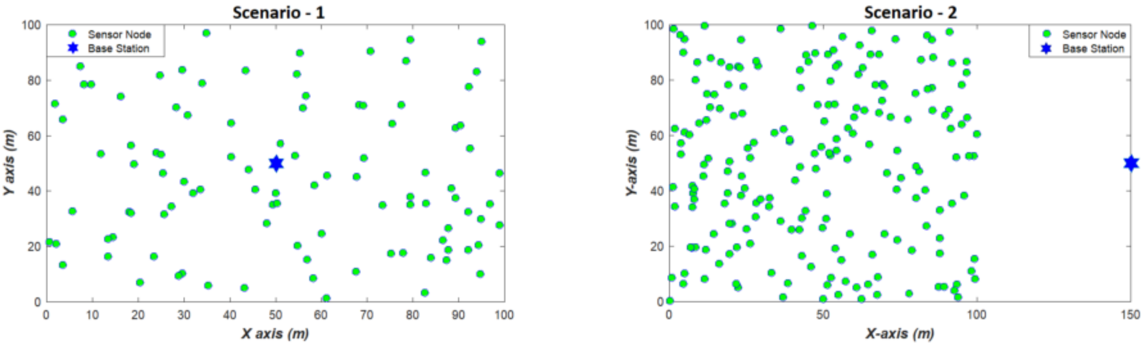
5. Simulation experiments and result analysis. NEEF is simulated and evaluated along with SCHFTL [27] and DFCR [28] protocols using MATLAB. The field size is considered to be 100 x 100 m^2 with randomly scattered SNs. The operation of the network is split into rounds. For every round, CH selection, cluster formation and data dissemination take place. The parameters for simulation are described in Table 5.1which are kept similar to SCHFLT [27] and DFCR [28]. The performance metrics chosen for evaluation of proposed work are alive nodes per round, packets to the BS, average energy of network, stability period (FND),

TABLE 4.4
Fuzzy rules for computing Chance of CH

Rule No.	R_Energy	DBS	Density	Rank
1	Lw	Fr	Pr	VW
2	Lw	Fr	Mm	Vw
3	Lw	Fr	Hg	W
4	Lw	Mm	Pr	W
5	Lw	Mm	Mm	W
6	Lw	Mm	Hg	RW
7	Lw	Nr	Pr	RW
8	Lw	Nr	Mm	MW
9	Mm	Nr	Hg	MW
10	Mm	Fr	Pr	RW
11	Mm	Fr	Mm	RW
12	Mm	Fr	Hg	MW
13	Mm	Mm	Pr	RW
14	Mm	Mm	Mm	MW
15	Mm	Mm	Hg	MW
16	Mm	Nr	Pr	MW
17	Hg	Nr	Mm	Mm
18	Hg	Nr	Hg	RStr
19	Hg	Fr	Pr	Mw
20	Hg	Fr	Mm	Mm
21	Hg	Fr	Hg	RStr
22	Hg	Mm	Pr	RStr
23	Hg	Mm	Mm	MStr
24	Hg	Mm	Hg	Str
25	Hg	Nr	Pr	MStr
26	Hg	Nr	Mm	Str
27	Hg	Nr	Hg	Vstr

Algorithm 2 Formation of cluster in NEEF

```
Tn ← Total nodes
TN_CH ← Total CH nodes in a round
k,m ← ID of SN
for each node in TN_CH do
    Broadcast CH_MSG(SN(k)).ID,SN (k).Member Nodes, SN (k). FF1
end for
for each non_CH_Node in Tn do
    if SN(k).type == "N" && SN(k).Energy>0 then
        OPTIMUM_CH ← 0 // Initially No CH is Chosen as Optimum
        OPTIMUM_CH_CHANCE ← 0 // Initialising Chance of each CH to 0
        for m=1 to TN_CH list do
            if SN(m) is within Communication Range of SN(k) then
                Max_Chance= Fuzzy (SN (m).member nodes,distance to CH, SN (m). FF1)
                if Max_Chance > OPTIMUM_CH_CHANCE then
                    OPTIMUM_CH← m //ID of CH node
                    OPTIMUM_CH_CHANCE← Max_Chance
                end if
            end if
        end for
    end if
end for
SN (k).CH← OPTIMUM_CH // Optimum CH is chosen by SN
SN(k) will transmit a join request to OPTIMUM_CH node
OPTIMUM_CH node will Acknowledge SN(k) with TDMA slot.
```



(a) Network Structure for S#1

(b) Network Structure for S#2

FIG. 5.1. *Network structure.*

TABLE 5.1
Simulation Parameters and their values

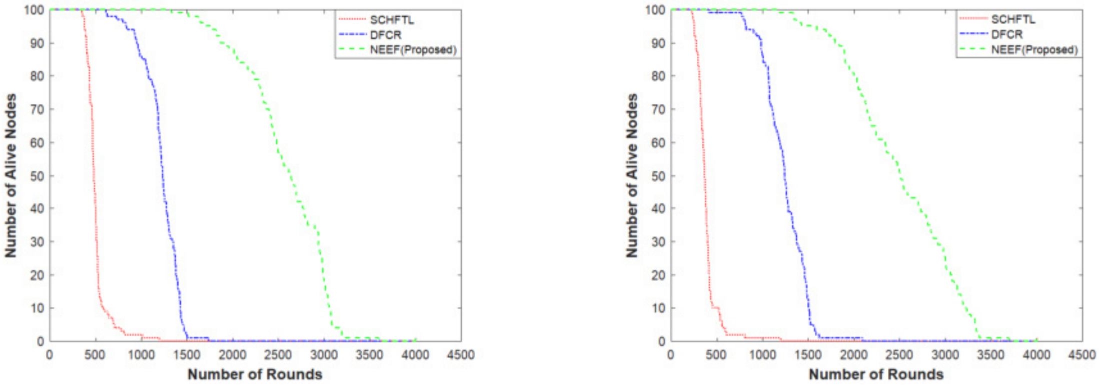
Parameters	Symbol	Values
Total SNs in Field	N	100,200
Amplifier energy for free space	ϵ_{fs}	$10pJ/bit/m^2$
BS location	BS	(50,50),(150,50)
Amplifier energy for multipath	ϵ_{mp}	$0.0013pJ/bit/m^4$
Energy of SN before deployment	E_o	$0.5J/1.0J$
Data packet Size	M	4000bits
Data Fusion	E_{DA}	5nJ/bit/report
Arbitrary constants	α, β	0.7,0.3
Percentage of CH probability	$p\%$	10
Electronic Circuitry	E_{elec}	$50nJ/bit$

QND, HND, average energy of chosen CHs. These metrics will evaluate the protocol from every perspective conforming the enhancement in lifetime of the network. We have carried out simulation more than 50 times and the results were normalised. The graphs depicted are instance of one of the simulations carried out so that clear picture about the performance of all the simulated fuzzy based protocols can be perceived.

5.1. Network structure. In conducting simulation experiments, we have considered two network structure/scenarios as depicted in Fig. 5.1. In scenario 1 (S#1), the BS is positioned at the centre of the field and scenario 2 (S#2) considers the BS located at far off place from the field. The reason for choosing two scenarios is that this protocol can satisfy all the applications of WSN where the BS is either within the vicinity or beyond the vicinity.

5.2. Alive nodes. With the focus on longer lifetime with maximum coverage, alive nodes have huge impact on WSN. More the number of alive nodes, longer will be the lifetime of the network. Fig. 5.2 depicts the number of alive SNs in the field after each rounds for both the scenarios. It can be clearly witnessed that NEEF performs better than SCHFTL and DFCR protocol as it has more alive nodes after each round as equated to SCHFTL and DFCR for both the scenarios. In scenario 1, for up to 1500 rounds, almost all the nodes are dead for SCHFTL and DFCR protocols whereas more than 90% nodes are alive in case of NEEF protocol. For scenario 2, for up to 1500 rounds, more than 90% nodes are alive in the network for NEEF protocol whereas no node is alive for SCHFTL protocol and more than 95% nodes are dead in case of DFCR protocol. Obtained result from Fig.5.2(a,b) clearly unveil the balanced load distribution among the deployed SNs.

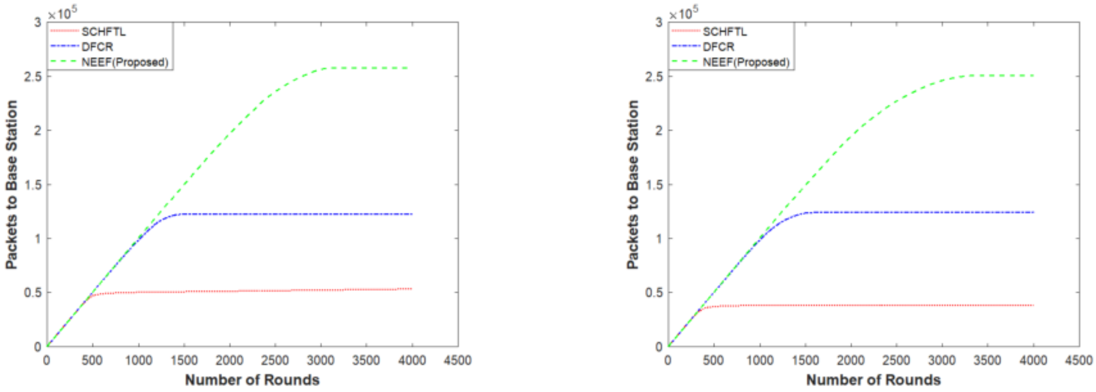
5.3. Throughput. Collecting information from the target area is the ultimate objective of WSN. Successful delivery of more information to the BS reveals better design of protocol. Fig. 5.3 exhibits the number of successful packet delivery to the BS during span of the network. We can see that that NEEF protocol



(a) S#1

(b) S#2

FIG. 5.2. *Alive nodes w.r.t rounds*



(a) S#1

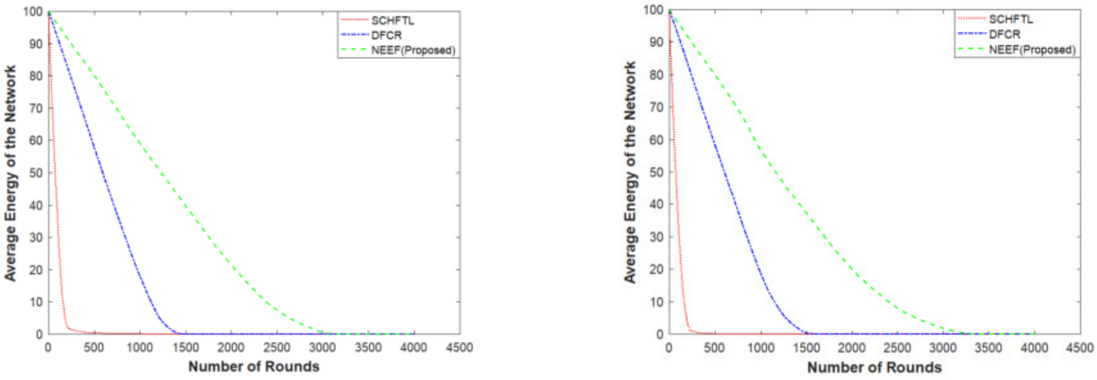
(b) S#2

FIG. 5.3. *Packets to the BS w.r.t. rounds*

has forwarded more packets to the BS as compared to SCHFTL and DFCR protocol. Since, more number of nodes are alive per round in NEEF protocol, therefore, more information is forwarded to the BS. Every WSN is expected to forward as much information as possible with longer lifetime and this purpose is served by NEEF protocol.

5.4. Average energy of network. The dissipation rate of energy of SN may affect the lifetime of the network. If the network is dissipating energy quickly then the lifetime will decrease resulting in incomplete coverage of target area. Fig. 5.4 discuss about average energy of network per round. It can be seen that average energy of NEEF protocol for each round is much more than SCHFTL and DFCR protocol for both the scenarios. The reason behind the better performance is the consideration of crucial parameters like communcation cost to be borne by SN if chosen as CH and its remnant energy level during the selection of CH. This balances the load resulting in more average energy of the network. Stability period or First node dead determines the reliability of the network [8]. If a protocol exhibit better stability period then it's clear that it has more reliability as it ensure the complete coverage of the network because all the SN are alive in the network till that stability period.

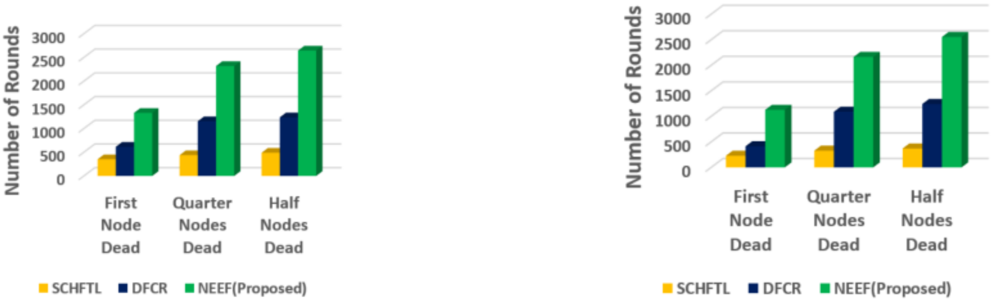
5.5. First Node Dead,Quarter Node Dead and Half Node Dead. FND ensures that the network is reliable as the deployed nodes are covering the target area intactly. QND and HND are checkpoints which determine the rate at which the nodes are expiring. Fig. 5.5 shows the rounds in which the simulated protocols have FND, QND and HND. The reason why we have not considered Last Node Dead (LND) is that after the



(a) S#1

(b) S#2

FIG. 5.4. *Average energy (J) per round*



(a) S#1

(b) S#2

FIG. 5.5. *First Node Dead, Quarter Node Dead and Half Node Dead*

death of 50% node, there is no reliability about the coverage of the target field. From experimental analysis, we have seen that some protocols continue up to 1000 rounds more with their last node which has no relevance as single node is incapable of covering all the target field. For FND, NEEF has improved over SCHFTL and DFCR by 278.63% and 116.09% for scenario 1. In case of scenario 2, the improvement of FND for NEEF is further increased having 398.85% and 172.81% over SCHFTL and DFCR respectively. If we talk about QND, NEEF has shown increment of 429.22% and 100.86% over SCHFTL and DFCR respectively for scenario 1 and 561.16% and 98.34% over SCHFTL and DFCR respectively for scenario 2. For Half Node Death, in case of Scenario 1, the obtained results are 440.69% and 114.08% better than SCHFTL and DFCR respectively and for Scenario 2, the improvements are 590% and 105.72% over SCHFTL and DFCR respectively.

5.6. Average energy of CH. This performance metric ensures that the nodes which are having better energy levels along with other primary and secondary factors are turned into CH. In Fig. 5.6, we have chosen average energy of CHs in lifetime of the network as one of metric because it depicts how energy bundled SNs are chosen to take the challenging role of CH. It is witnessed that the average energy of CHs for scenario 1 is improved by 986.34% and 59.25% as compared to SCHFTL and DFCR protocol. For scenario 2, NEEF performs 1104.91% and 67.93% better than SCHFTL and DFCR protocol. The cause of the poor performance of SCHFTL is that it has chosen some random values for the parameters like DOS, communication quality and total delay which does not give deep insight of the SN capability while selecting the CH candidate. In case of DFCR protocol, it considers the remnant energy and aloofness from the BS as crucial factors for election of CH where intra cluster communication cost is neglected which results in poor performance as compared to NEEF protocol.

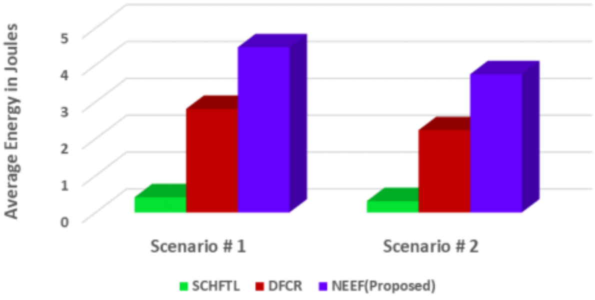


FIG. 5.6. *Average Energy of CH for both scenarios*

6. Conclusion. NEEF is proposed for resource constrained WSN. This protocol has considered two fuzzy fitness values obtained from primary and secondary factors fed to designed Fuzzy Inference System. Primary factors include remnant energy and communication cost whereas secondary factors considered are node density and distance to the base station. With experimental observations, primary and secondary factors are used in appropriate proportion for selection of best cluster head candidates. For balancing the cluster head role, non-cluster head members use fuzzy logic for choosing their cluster heads. NEEF exhibits tremendous improvement over SCHFTL and DFCR protocol in terms of protracted lifetime, stability period, and better average energy of chosen cluster heads and communication of more information to the base station. This protocol is suitable for applications where either the base station can be put at the centre or beyond the target field. In future, we will perform simulation experiments on mobile sensor nodes.

REFERENCES

[1] P.Rawat, K. Singh, H.Chaouchi, and J.M.Bonnin. Wireless sensor networks: a survey on recent developments and potential synergies. *The Journal of Supercomputing*, 68(1):1–48, Apr 2014.

[2] M. Mehdi Afsar and Mohammad H. Tayarani-N. Clustering in sensor networks: A literature survey. *Journal of Network and Computer Applications*, 46:198–226, Nov 2014.

[3] P.S. Mehra, M.N. Doja, and B. Alam. *Enhanced stable period for two level and multilevel heterogeneous model for distant base station in wireless sensor network*, volume 379, page 751–759. 2016.

[4] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, volume vol.1, page 10. IEEE Comput. Soc, 2000.

[5] P. S. Mehra, M. N. Doja, and B. Alam. Stability Enhancement in LEACH (SE-LEACH) for Homogeneous WSN. *EAI Endorsed Transactions on Scalable Information Systems*, 6(20):e5, 2019.

[6] S. Lindsey and C.S. Raghavendra. Pegasis: Power-efficient gathering in sensor information systems. In *Proceedings, IEEE Aerospace Conference*, volume 3, pages 3–1125–3–1130. IEEE, 2002.

[7] O.Younis and S.Fahmy. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *Mobile Computing, IEEE Transactions on*, 3(4):366–379, 2004.

[8] G. Smaragdakis, I. Matta, and A. Bestavros. SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. *Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004)*, 2004.

[9] P. S. Mehra, M. N. Doja, and B. Alam. Low energy adaptive stable energy efficient (LEASE) protocol for wireless sensor network. In *2015 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE 2015*, page 484–488, 2015.

[10] H.Zhou, Y.Wu, Y.Hu, and G.Xie. A novel stable selection and reliable transmission protocol for clustered heterogeneous wireless sensor networks. *Computer Communications*, 33(15):1843–1849, Sep 2010.

[11] P.S. Mehra, M.N. Doja, and B. Alam. *Stable Period Enhancement for Zonal (SPEZ)-Based Clustering in Heterogeneous WSN*, volume 79, page 887–896. Springer, Singapore, 2018.

[12] I. Gupta, D. Riordan, and S. Sampalli. Cluster-head election using fuzzy logic for wireless sensor networks. In *3rd Annual Communication Networks and Services Research Conference (CNSR'05)*, page 255–260. IEEE, 2005.

[13] J Kim, S Park, Y Han, and T Chung. Chef: Cluster head election mechanism using fuzzy logic in wireless sensor networks. In *Proceedings of 10th International Conference on Advanced Communication Technology*, pages 654–659, 2008.

[14] G.Ran, H.Zhang, and S.Gong. Improving on leach protocol of wireless sensor networks using fuzzy logic. *Journal of Information & Computational Science*, 7(3):767–775, 2010.

[15] J.S.Lee and W.L.Cheng. Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication. *IEEE Sensors Journal*, 12(9):2891–2897, Sep 2012.

[16] F.Zhang, Q.Y.Zhang, and Z.M.Sun. Ict2tsk: An improved clustering algorithm for wsn using a type-2 takagi-sugeno-kang fuzzy logic system. In *2013 IEEE Symposium on Wireless Technology & Applications ISWTA)*, page 153–158. IEEE, Sep 2013.

[17] B.Mostafa, C.Saad, and H.Abderrahmane. Fuzzy logic approach to improving stable election protocol for clustered heterogeneous wireless sensor networks. *Journal of Theoretical and Applied Information Technology*, 53(3):334–339, 2013.

[18] C.Li, M.Ye, G.Chen, and J.Wu. An energy-efficient unequal clustering mechanism for wireless sensor networks. In *2nd IEEE International Conference on Mobile Ad-hoc and Sensor Systems, MASS 2005*, volume 2005, page 597–604, 2005.

[19] S.A.Sert, H.Bagci, and A.Yazici. Mofca: Multi-objective fuzzy clustering algorithm for wireless sensor networks. *Applied Soft Computing*, 30:151–165, May 2015.

[20] Z.M.Zahedi, R.Akbari, M.Shokouhifar, F.Safaei, and A.Jalali. Swarm intelligence based fuzzy routing protocol for clustered wireless sensor networks. *Expert Systems with Applications*, 55:313–328, Aug 2016.

[21] R. Logambigai and A. Kannan. Fuzzy logic based unequal clustering for wireless sensor networks. *Wireless Networks*, 22(3):945–957, Apr 2016.

[22] B. Baranidharan and B. Santhi. Ducf: Distributed load balancing unequal clustering in wireless sensor networks using fuzzy approach. *Applied Soft Computing*, 40:495–506, Mar 2016.

[23] P. S. Mehra, M. N. Doja, and B. Alam. Fuzzy based enhanced cluster head selection (FBECS) for wsn. *Journal of King Saud University - Science*, Apr 2018.

[24] D.R.D.Adhikary and D.K.Mallick. An energy aware fuzzy approach to unequal clustering in wireless sensor networks, applied soft computing, 13(4), pp. 1741-1749, 2013. *Journal of ICT Research and Applications*, 11(1):56–77, Apr 2017.

[25] A.Alaybeyoglu and Aysegul. A distributed fuzzy logic-based root selection algorithm for wireless sensor networks. *Computers & Electrical Engineering*, 41(C):216–225, Jan 2015.

[26] Q.Wang, D.Lin, P.Yang, and Z.Zhang. A fuzzy-logic based energy-efficient clustering algorithm for the wireless sensor networks. In *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, page 1–6. IEEE, Sep 2018.

[27] M.Ayati, M.H.Ghayyoumi, and A.K.Mohammadiyan. A fuzzy three-level clustering method for lifetime improvement of wireless sensor networks. *Annales des Telecommunications/Annals of Telecommunications*, 73(7–8):535–546, 2018.

[28] N.Mazumdar and H.Om. Distributed fuzzy approach to unequal clustering and routing algorithm for wireless sensor networks. *International Journal of Communication Systems*, 31(12):e3709, Aug 2018.

[29] P. S. Mehra, M.N. Doja, and B.Alam. Enhanced clustering algorithm based on fuzzy logic (E-CAFL) for wsn. *Scalable Computing: Practice and Experience*, 20(1):41–54, Mar 2019.

[30] H. El Alami and A. Najid, *Fuzzy Logic Based Clustering Algorithm for Wireless Sensor Networks*, International Journal of Fuzzy System Applications (IJFSA), 6 (2017), pp. 63–82.

[31] E.H.Mamdani. Application of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Transactions on Computers*, C–26(12):1182–1191, Dec 1977.

[32] P.Nayak and A.Devulapalli. A fuzzy logic-based clustering algorithm for wsn to extend the network lifetime. *IEEE Sensors Journal*, 16(1):137–144, Jan 2016.

Edited by: Dana Petcu
Received: Apr 28, 2020
Accepted: Jul 6, 2020

AIMS AND SCOPE

The area of scalable computing has matured and reached a point where new issues and trends require a professional forum. SCPE will provide this avenue by publishing original refereed papers that address the present as well as the future of parallel and distributed computing. The journal will focus on algorithm development, implementation and execution on real-world parallel architectures, and application of parallel and distributed computing to the solution of real-life problems. Of particular interest are:

Expressiveness:

- high level languages,
- object oriented techniques,
- compiler technology for parallel computing,
- implementation techniques and their efficiency.

System engineering:

- programming environments,
- debugging tools,
- software libraries.

Performance:

- performance measurement: metrics, evaluation, visualization,
- performance improvement: resource allocation and scheduling, I/O, network throughput.

Applications:

- database,
- control systems,
- embedded systems,
- fault tolerance,
- industrial and business,
- real-time,
- scientific computing,
- visualization.

Future:

- limitations of current approaches,
- engineering trends and their consequences,
- novel parallel architectures.

Taking into account the extremely rapid pace of changes in the field SCPE is committed to fast turnaround of papers and a short publication time of accepted papers.

INSTRUCTIONS FOR CONTRIBUTORS

Proposals of Special Issues should be submitted to the editor-in-chief.

The language of the journal is English. SCPE publishes three categories of papers: overview papers, research papers and short communications. Electronic submissions are preferred. Overview papers and short communications should be submitted to the editor-in-chief. Research papers should be submitted to the editor whose research interests match the subject of the paper most closely. The list of editors' research interests can be found at the journal WWW site (`http://www.scpe.org`). Each paper appropriate to the journal will be refereed by a minimum of two referees.

There is no a priori limit on the length of overview papers. Research papers should be limited to approximately 20 pages, while short communications should not exceed 5 pages. A 50–100 word abstract should be included.

Upon acceptance the authors will be asked to transfer copyright of the article to the publisher. The authors will be required to prepare the text in $\LaTeX 2_\varepsilon$ using the journal document class file (based on the SIAM's `siamltex.clo` document class, available at the journal WWW site). Figures must be prepared in encapsulated PostScript and appropriately incorporated into the text. The bibliography should be formatted using the SIAM convention. Detailed instructions for the Authors are available on the SCPE WWW site at `http://www.scpe.org`.

Contributions are accepted for review on the understanding that the same work has not been published and that it is not being considered for publication elsewhere. Technical reports can be submitted. Substantially revised versions of papers published in not easily accessible conference proceedings can also be submitted. The editor-in-chief should be notified at the time of submission and the author is responsible for obtaining the necessary copyright releases for all copyrighted material.