A Survey on Optimizing Data Gathering in Internet of Things

Prof. Garima Bhatt[1] Prof. Ganga Singh[2]
[1]Research Scholar [2]Head of Department (CSE) & Assistant Professor
[1]OPJS University, Jaipur, India [2]Bagula Mukhi College of Technology, Bhopal (M.P.), India

Abstract— Internet of things (I.O.T) facilitates trillions of devices to be activated with network connectivity to collect and exchange real-time information for providing intelligent services. Thus, IOT allows connected devices to be controlled and accessed remotely in the presence of adequate network infrastructure. Unfortunately, traditional network technologies such as enterprise networks and classic timeout-based transport protocols are not capable of handling such requirements of IOT in an efficient, scalable, seamless, and cost-effective manner. Besides, the advent of software-defined networking (SDN) introduces features that allow the network operators and users to control and access the network devices remotely, while leveraging the global view of the network. In this respect, we provide a comprehensive survey of different SDN-based technologies, which are useful to fulfill the requirements of IOT, from different networking aspects-edge, access, core, and data center networking. In these areas, the utility of SDN-based technologies is discussed, while presenting different challenges and requirements of the same in the context of IOT applications. We present a synthesized overview of the current state of IOT development. We also highlight some of the future research directions and open research issues based on the limitations of the existing SDN-based technologies.

Keywords: IoT, Network Access, ML, AI, DM Energy Optimization Adaptive Neural Learning Principal Component Analysis, Routing Protocol

I. INTRODUCTION

Internet of Things (IoT) is one of the most recent technologies that are reshaping our future in a distinguish way. The IOT paradigm permits trillions and billions of devices to communicate and connect through the internet. These devices include sensor nodes, actuators, etc. which can work in cooperation with other such devices with minimum or no human involvement in their operation [1]. The advent of IOT has revolutionized the performance of various machines in different applications such as wireless sensing, healthcare, security ,environmental protection, transport services and many more [2].Wireless Sensor Network (WSN) is one of the major contributor to the data gathering schemes for IOT. Information is sensed and processed in a distributed manner and is sent to a centralized source (base station) for further processing. Although, data gathering in traditional WSNs and data gathering in IOT is a bit different from each other as the sensors in the latter are smarter and independent of humans. The IOT based data path ring requires the sensor not only to be able to sense and process the information but also to take critical decisions on their own within the inherent constraints of battery and bandwidth. Additionally, the significantly high number of devices imposes severe complications in managing and controlling the sensors involved in IOT data gathering. A large number of devices in IOT, continuously sense and generate a massive amount of data that needs to be regularly and quickly processed and analyzed for eective decision making and improve the IOT framework for intelligent processing [3]. Optimally, the raw data from the numerous connectivity and data forwarding.

1) Network Formulation and Communication Establishment: The primary objective of this module is to deliver the gathered data to the desired destination.
2) Data and Platform: This module facilitates data storage on the cloud and accessibility to the stored data.

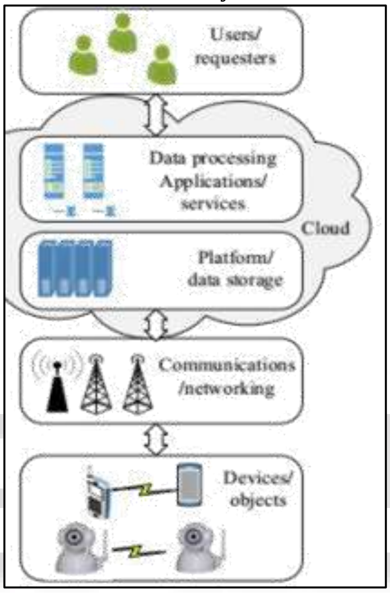

Fig. 1: IOT Architecture

3) Management and Processing: The final module of the IOT architecture facilitates the end uses to avail the IOT services.

II. LITERATURE SURVEY

A. The Ethics of Safety Critical Systems by Jonathan

Bowen Author has developed some ethics in order to solve the problems of testing of safety critical Embedded Systems. Developer has freedom to choose different techniques and methodologies to test and to develop the software. Aircraft as example for Safety Critical systems contains 10-9 accident per hours where software testing is not sufficient to solving of problems [2]. Software diversity likes, N-version programming and verification and validation of software can add some extra reliability and measurements on the effectiveness of such approaches are difficult to obtain. Computer program having change of single bit can lead unpredictable effect on its normal operation. Author explained some of the pitfalls and issues in choosing a suitable approach [1].

B. Neural-Network Based Test Cases Generation Using Genetic Algorithm by Ruilian Zhao & Shanshan Lv

In the paper Neural-Network based Test case Generation using Genetic Algorithm, Author focuses on major problem while performing Black Box Testing, where Black box testing is how to select test cases from input domain on the basis of specification. In some software instead of input domain test cases are generated by output domains. Author presents a new approach for automated test case generation. Model developed via neural network and improved genetic algorithms. Here the classification is based on specification and the specification is responsible to select test cases [3].

C. Model- Based Test cases prioritization using Neural Network Classification by Nida Gökçe1& Mübariz Eminli

Running of test cases of differential Embedded Systems at a time may create large complexity and problems. If we consider Model based testing for real life software systems, we have to create a large number of tests and execution of test cases randomly will not give appropriate results. There is need of prioritization of tests before execution according to their importance or testers requirements. This paper solves the problem by providing classification approach to the results. Here parameter for classification is important index and frequency of events occurring. For approach classification Neural network is used and for all test cases classification is performed by multi layer perceptron neural network in artificial intelligence [4].

D. Automated Test case generation based on coverage analysis by Tim A.Majchrzak & Herbert Kuchen

Paper focuses on automated generations of unit tests. They have developed tool, which is symbolically executed java byte code for searching execution path through program, for searching choice-point generation, constraint solving and backtracking was used .They have proposes a novel way to reduce the test cases, which was based on their contribution to the global coverage of the data flow and control flow. Basic aim behind work is elimination of repeated and meaningless test cases. For example a method for handling AVL trees should only be tested by AVL trees rather than with arbitrary trees [5].

E. Testing Embedded Real Time Systems with TTCN-3by Juergen Grossmann, Diana Serbanescu.

In this paper author introduces a real time test technology and methodology for the automotive domain, based on a standardized test technology. These enable the exchange of test requirements, test cases, and test procedures between different parties involved in automotive systems development. Both methodology and technology are based on TTCN-3 and will contain several enhancements that adapt the already existing TTCN-3 standard to the needs of the automotive domain and especially to the emerging AUTOSAR standard. The paper started with the specification of real-time enhancements to TTCN-3 that provide systematic support for testing timing requirements of automotive software applications under real-time conditions. Author presented the real-time language features that were introduced to TTCN-3. They described their syntax and semantics and discussed their application by means of an example that reflects typical automotive timing requirements [6].

F. Parameter extraction and optimization using Levenberg –Marquardt algorithm by Le Due-Hung,Pham Cong-Kha.

Parameter extraction is use to determine the values of device model that minimize the difference between a set of measured characters and the result obtained by evaluation of device model characteristic to measure the data. A combination of a local optimization and the group device extraction strategy is adopted for parameter extraction. Here comparison between simulated and experimental data is carried out for the model fitting. In this paper author proposes the Levenberg-Marquardt algorithm for optimization and parameter extraction of EKV MOSFET model. This algorithm is the combination between the steepest gradient and the Gauss-Newton algorithms. Mat lab environment is used for design project implementation [7].

III. MACHINE LEARNING FOR DATA ANALYSIS

For accurate decision making from smart data analytics, it is trivial to determine the ex-act nature of the task to be accomplished. For example, the task of structure discovery in unstructured data, clustering algorithms are the most satisfactory solutions. K-means clus-tering, for instance, is capable of handling big data sets with a wide range of data types. The algorithms proposed in [16], [17] are an example of such a clustering applied to Smart City application of IoT. Another popular approach for identifying the structure of unstructured data is the DB-scan algorithm for clustering and has been widely used in several IoT appli-cations [17], [18], [19]. A different task, which can be achieved through machine learning, is identifying unusual data points and anomalies in smart data. Support Vector Machine (SVM) and Principle Component Analysis (PCA), in this regard, have been widely applied to optimize the noisy data for achieving high efficiency. For instance, the schemes proposed in [17] [20] use one class SVM to identify the anomalous behaviour of humans. Similarly, linear regressions can be applied to the IoT data to infer values and classify sequenced data[21], [22]. Some other examples include neural networks for categorizing data. An instance of such an application is the feed forward neural network used for inferring the behavior of data generation in the future and optimizing the energy consumption of the network [23], [24] and [25]. The learning based algorithms take a set of input, termed as `training data' to learn the behavior of the data. Based on the training set, the algorithm makes prediction and inferences for future behaviors. Therefore, based on the type of learning, the algorithms are classified as [26], [27] and [28]:

1) Supervised: The training set comprises of input vectors along with corresponding target vectors, popularly known as labels.
2) Unsupervised: The training set does not have target vectors/labels.
3) Reinforcement: Deals with specific situation condition.

IV. NOTEWORTHY CONTRIBUTIONS IN THE FIELD OF PROPOSED WORK.

WSNs are the main components of IoT which collect data from the environment and transmit the data to the sink nodes. For long service time and low maintenance cost, WSNs require adaptive and robust designs to address many issues, e.g., data collection, topology formation, packet forwarding, resource and power optimization, coverage optimization, efficient task allocation, and security. For these issues, sensors have to make optimal decisions from current capabilities and available strategies to achieve desirable goals. IoT continues to extend the current Internet by providing connectivity and interaction between the physical and cyber worlds. In addition to increased volume, the IoT generates Big Data characterized by velocity in terms of time and location dependency, with a variety of multiple modalities and varying data quality. Intelligent processing and analysis of this Big Data is the key to developing smart IoT applications.

Driven by the need of accurate and energy efficient solutions for data gathering and analysis in IoT applications, the objectives of this work are defined as follows:

1) Economic and Pricing Model for Data Gathering

As data is the basis for the extraction of knowledge, IoT applications require efficient data gathering methods. Since the volume of data produced by any IoT application is vast, the data gathering plays a vital role in determining the accuracy and efficiency with which the data is gathered and processed. Considering, the wide range of hard-ware and technologies that might be involved in the gathering, the effctiveness of the gathering schemes is a trivial task in IoT applications. Although, a wide range of solutions have been explored in the recent years, however, the diversity of the Big Data and the need of real-time processing, impose severe limitations on the efficiency of such approaches.

2) Economic and Pricing Model for Topology Formation

The data gathering in IoT applications is highly dependent on the routing structure which varies significantly with the applications of IoT. Since the applications of IoT are extensive, and not all the designated hardware for the data gathering share the same data type, routing algorithms, etc., it is vital to explore a solution which can support several IoT applications and is compatible with different gathering schemes.

V. MACHINE LEARNING FOR DATA ANALYSIS

Algorithms designed for the analytical analysis of smart data should support Big Data analytic. In simpler terms, since IoT applications generate a considerable volume of data which requires immediate processing, analytical algorithms for IoT applications should be able to cope up with the real-time processing of data from different type of hardware. For example, the deep learning algorithms can generate highly accurate results given the volume of data is large and they have enough time to process it. However, the real-time processing of such large volumes requires some optimization to get the desired result. Additionally, such algorithms do not fare well with noisy smart data. Similarly, semi-supervised learning algorithms can be optimized to incorporate their use in IOT and Big Data analytic.

1) Validating the proposed schemes for IoT

To achieve the above objectives, a set of schemes need to be developed and validated with extensive simulations and comparative analysis with some of the state of the art algorithms.

2) Proposed methodology during the tenure of the research work.

*we can use various machine learning algorithms like svm, cart, lda, ANN etc

The plan of work can be roughly divided into the following headings:

1) Literature survey: An extensive survey of the previous work would enable us to havea clear picture of the problems associated with the data gathering in IoT.
2) Identifying objectives: A set of objectives must be defined to ensure the progress of the research. Objective identification is necessary to implement the step-by-step work plan. Issues in IoT are many, but identifying a particular problem and related issues with it, would be the next step of this work.
3) Propose solutions for achieving the objectives: In order to achieve the objectives of the work, schemes based on economic and pricing models are required for optimizing the data gathering and routing issues in IoT. Additionally, a machine learning based data analytic approach will also be explored for analyzing the data thus produced. The proposed solutions will aim at rectifying the identifed issues and meeting up with the objectives of the work.
4) Comparative analysis and result description: Finally, a comparative analysis of the previous approaches and the proposed solutions will be done. The comparative analysis will justify the proposed solutions for the identified problems.

Fig. 2: Proposed Research Plan

Activity Plan of the Research Work

An expected activity plan of the research work is shown in figure 3. The chart is divided into six equal halves. Each half shows a duration of six months. Various tasks that will be achieved in this duration are:

- 1st Six Months: An extensive survey of the related literature will be done.
- 2nd Six Months: Based on the literature study, a problem will be identified.
- 3rd Six Months: Expected solutions for the problem will be finalized.
- 4th Six Months: The solutions will be implemented and simulation results will be obtained.

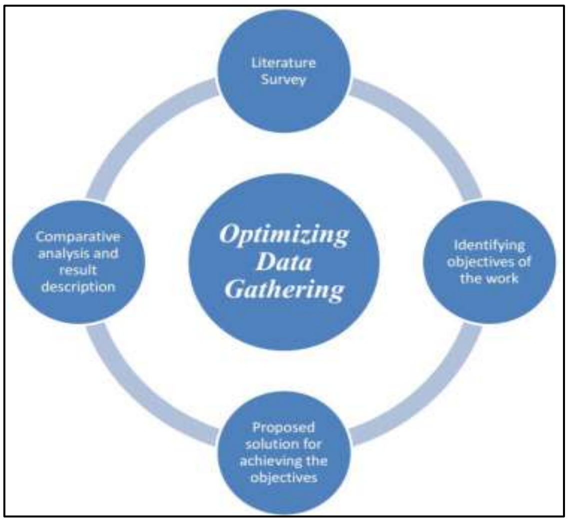

Fig. 3: Proposed Activity Plan

– 5th Six Months: Based on the work done, research papers will be written and will be submitted for publication.
– 6th Six Months: Thesis writing.

VI. EXPECTED OUTCOME

Successful completion of this work would result in the following:

1) An efficient scheme for data gathering based on economic and pricing models.
2) An efficient scheme for topology formation based on economic and pricing models.
3) A machine learning based scheme to analyze the large volumes of data generated by IoT systems.
4) Validation of the proposed schemes through extensive simulations to prove the efficacy of the schemes over existing approaches.

Sources on the network is processed for the extraction of knowledge. Data science, as such, is used in diferent IoT applications to make the process accurate, quick and1 intelligent. With the wide range of machine learning and mining algorithms available, the collaboration of data science and IoT produces intelligent solutions to improve the quality of IoT applications. The collaboration of data science and IoT involves identifying and ending solutions to the much needed aspects such as the diff erent data types, the three V'sof IoT (volume, velocity, and variety). These are suitable models for such data types such as artifcial neural networks, classifcations etc. and energy ecient algorithms for combining and matching the broad characteristics of data [4]. As a result, new and ecient approaches for scalable, accurate and energy ecient data gathering and processing are being explored to improve the quality of the wide range of services provided by the IoT.

REFERENCES

[1] R. Minerva, A. Biru, and D. Rotondi, “Towards a definition of the internet of things (iot)," IEEE Internet Initiative, no. 1, 2015.
[2] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (iot): A vision, architectural elements, and future directions," Future generation computer sys-tems, vol. 29, no. 7, pp. 1645-1660, 2013.
[3] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey," Computer networks, vol. 54, no. 15, pp. 2787-2805, 2010.
[4] C. Cecchinel, M. Jimenez, S. Mosser, and M. Riveill, “An architecture to support the collection of big data in the internet of things," in Services (SERVICES), 2014 IEEE World Congress on. IEEE, 2014, pp. 442-449.
[5] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woel∟e, “Vision and challenges for realising the internet of things," Cluster of European Research Projects on the Internet of Things, European Commision, vol. 3, no. 3, pp. 34-36, 2010.
[6] M. Quarterly, “The internet of things," McKinsey&Company?2010, 2010.
[7] J. A. Stankovic, “Research directions for the internet of things," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 3-9, 2014.
[8] E. Borgia, “The internet of things vision: Key features, applications and open issues," Computer Communications, vol. 54, pp. 1-31, 2014.
[9] D. Uckelmann, M. Harrison, and F. Michahelles, “An architectural approach towards the future internet of things," in Architecting the internet of things. Springer, 2011, pp. 1-24
[10] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, “A survey on facilities for experimental internet of things research," IEEE Communications Magazine, vol. 49, no. 11, 2011.
[11] N. Cao, S. Brahma, and P. K. Varshney, \An incentive-based mechanism for location estimation in wireless sensor networks," in Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE. IEEE, 2013, pp. 157-160.
[12] F. Zhao, J. Shin, and J. Reich, “Information-driven dynamic sensor collaboration,"IEEE Signal processing magazine, vol. 19, no. 2, pp. 61{72, 2002.
[13] S. Y. Shah and B. K. Szymanski, ”Price based routing for event driven prioritized traffic in wireless sensor networks," in Network Science Workshop (NSW), 2013 IEEE 2nd.IEEE, 2013, pp. 1-8.
[14] G. He,“Destination-sequenced distance vector (dsdv) protocol," Networking Laboratory,Helsinki University of Technology, pp. 1-9, 2002.
[15] A. E. Al-Fagih, F. M. Al-Turjman, W. M. Alsalih, and H. S. Hassanein, “A priced public sensing framework for heterogeneous iot architectures," IEEE Transactions on Emerging Topics in Computing, vol. 1, no. 1, pp. 133-147, 2013.
[16] A. M. Souza and J. R. Amazonas, “An outlier detect algorithm using big data processing and internet of things architecture," Procedia Computer Science, vol. 52, pp. 1010-1015,2015.
[17] M. Shukla, Y. Kosta, and P. Chauhan, “Analysis and evaluation of outlier detection algorithms in data streams," in Computer, Communication and Control (IC4), 2015 International Conference on. IEEE, 2015, pp. 1-8.
[18] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, “Mining smart card data for transitriders travel patterns," Transportation Research Part C: Emerging Technologies, vol. 36,pp. 1-12, 2013.
[19] M. A. Khan, A. Khan, M. N. Khan, and S. Anwar, “A novel learning method to classify data streams in the internet of things," in Software Engineering Conference (NSEC),2014 National. IEEE, 2014, pp. 61-66.
[20] A. Shilton, S. Rajasegarar, C. Leckie, and M. Palaniswami, “Dp1svm: A dynamic planar one-class support vector machine for internet of things environment," in Recent Advances in Internet of Things (RIoT), 2015 International Conference on. IEEE, 2015,pp. 1-6.
[21] P. Ni, C. Zhang, and Y. Ji, “A hybrid method for short-term sensor data forecasting in internet of things," in Fuzzy Systems and Knowledge Discovery (FSKD),

2014 11th International Conference on. IEEE, 2014, pp. 369-373.
[22] W. Derguech, E. Bruke, and E. Curry, "An autonomic approach to real-time predic-tive analytics using open data and internet of things," in Ubiquitous Intelligence and Computing, 2014 IEEE 11th Intl Conf on and IEEE 11th Intl Conf on and Autonomic and Trusted Computing, and IEEE 14th Intl Conf on Scalable Computing and Com-munications and Its Associated Workshops (UTC-ATC-ScalCom). IEEE, 2014, pp.204-211.
[23] F. Ramalho, A. Neto, K. Santos, N. Agoulmine et al., "Enhancing ehealth smart ap-plications: A fog-enabled approach," in E-health Networking, Application & Services (HealthCom), 2015 17th International Conference on. IEEE, 2015, pp. 323-328.
[24] I. Kotenko, I. Saenko, F. Skorik, and S. Bushuev, "Neural network approach to forecast the state of the internet of things elements," in Soft Computing and Measurements (SCM), 2015 XVIII International Conference on. IEEE, 2015, pp. 133-135.
[25] S. Hu, "Research on data fusion of the internet of things," in Logistics, Informatics and Service Sciences (LISS), 2015 International Conference on. IEEE, 2015, pp. 1-5.
[26] D. Barber, Bayesian reasoning and machine learning. Cambridge University Press,2012.
[27] C. M. Bishop, Pattern recognition and machine learning. springer, 2006.
[28] K. P. Murphy, Machine learning: a probabilistic perspective. MIT press, 2012

