Data Processing and Management in IoT and Wireless Sensor Network

S.P. Sasirekha1, *, A. Priya2, T. Anita3 and P. Sherubha4

1,2Research scholar, Assistant Professor, Research Fellow (computational sciences)
1,4Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu
2Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu
3Makeit Technologies (Centre of Industrial Research), Coimbatore, Tamil Nadu

*Corresponding author e-mail: spsasirekha@gmail.com

Abstract – The deployment of internet over larger scale may introduce huge challenges based on data processing. The enormous amount of IoT based data needs design-based solution for faster data processing and improving its extensibility and adaptability. Based on various IoT based data processing, servicing technologies may provide data-centric models for scalable services. This work concentrates on an extensive review towards the scalable realization and acquisition of data for process. This IoT based services are larger enough to be concentrated. This review gives an insight towards the data processing and management in IoT and sensor-based networking.

1. Introduction

When modelling big data system, data storage, collection, analysis and visualization are essential factors for consideration. In some research activities of big data systems, investigators concentrate on baseline factors for handling big data: certain applications, and technology, practices and standards, i.e. financial, social web, etc). However, there are numerous tools and infrastructure execute above mentioned functionalities of real world [1]. Therefore, data intensive applications are now modelled to offer certain benefits. Based on various investigations, existing approaches has to deal big data system problems (more specifically, infrastructure-based computation for data overflow that comprises cloud computing, granular computing, and bio-inspired computing).

However, there are seven basic factors has to be examined. Along with this, some general overview, concentrating on four stages of big data validation is given [2]. In various survey, investigators concentrate on technical confronts in data acquisition, generation, analysis and storage as separate factor. As well, numerous surveys related to big data was published that resolves various data techniques and platforms, with objective to choose appropriate technological combination in accordance to certain application requirements and technical requirements.

As depicted in prevailing studies, data collection is preliminary steps for constructing big data system. With various data generating sources, WSNs can acquire substantial factors that concern environmental monitoring. WSN comprises of huge amount of SN records and monitors physical factors of WSN and data are accumulated at sink node. WSN are cast off to validate the conditions: sound, temperature, humidity, wind and so on. Moreover, restricted capacity and wireless networks lead to crisis that delivers data to sink node. Nonetheless, an effectual network processing and data aggregation is advantageous to big data. However, there is some necessity of examining research
work connects big data and WSN systems while overcoming WSN deficiencies and enhancing performance.

Figure 1 depicts basic big data as an instance of basic architecture. As provided in Figure 1, sink node gathers data from nodes and gives data to provisional storage for subsequent data aggregation. Aggregated data is influenced by its applications. Certain crisis of investigations and numerous other works are done with integration of WSN as essential source in big data system with above mentioned architecture. Some instances of “Big Data in Ubiquitous WSNs” shows benefits in storage over cloud computing and computation.

![Figure 1. Big data processing in WSN (Source: Ref 1)](image)

However, technical crisis of big data in 3D WSNs, underwater sensor networks, and WMSNs are illustrated and explained. In some previous works, challenges and issues of data collection through WSNs. Researchers concentrated on WSN heterogeneity and energy consumption, that are monitored over big data analysis in WSN [3]. Some issues are addressed in data management crisis of WSNs by initiating diverse algorithm modelled for data aggregation, collection, correlation, prediction and compression, as well, they illustrated feasible and potential application scenarios.

Convergence crisis among (IoT)/WSNs are analyzed and reviewed, where corresponding issues are mentioned: security, data management, convergence problem, and hardware/architecture and interoperability challenges. From data usage perspective, data collected for IoT supplies context aware computing like pervasive and ubiquitous computing. Therefore, big data crisis is WSNs require to be explored effectually. To acquire certain goals, investigators handle open issues and offer IoT based big data insights [4]. Moreover, wireless big data system can handle enormous data in Wireless networks. Indeed, of demonstrating complete system, investigators concentrate on wireless infrastructure like data-driven network optimization, data-aided transmission and applications over layered architecture (as depicted in data layers, transmission, network and applications). Moreover, they illustrate three potential application areas: IoT, smart grids and drones/UAV.

Indeed, of above-mentioned research efforts, comprehensive analysis that offers huge perspective in most existing fields for WSNs as big data source that is lacking. To resolve this research gaps, comprehensive analysis of research that initiates data management in IoT and big data is addressed here. To analyze some drawbacks, existing works that concentrates on big data and WSNs are chosen simultaneously. Even though numerous prevailing research works are applicable to big data.

2. WSN-Big Data system applications
Before commencing detailed investigations, it is extremely desirable to validate data applications are executed and deployed. As WSN is generally constructed to fulfil application requirements, it is sensible to big data application reviews to resolve its crisis. The applications that follow system monitoring can offer advantages from WSNs: environmental monitoring, human body monitoring and smart grids.

In smart grids, SNs are initiated for energy management. The functions over smart grid application comprise energy management, power monitoring, distributed storage coordination and renewable energy generators [5]. In addition, approaches cast off to handle big data constructed by meters and sensors are anticipated. However, practical execution and recommendations of smart grids are analyzed. Some investigators concentrate on handling data and extracting information from the system. As low latency and reliability are two preliminary objectives in grid and fog computing for streaming process is analyzed for real time applications.

The application of observing abnormal functionality of human body is examined with data analysis and ZigBee based monitoring system. To eliminate pulse signal missing, two factors were anticipated:

1) Photo-electricity based dynamic monitoring
2) Anti-jamming approaches

Subsequently, some environmental monitoring examples observe air quality in workplaces. Case studies with respect to big data system that monitor air quality are presented [6]. This study is performed to validate data analytics and visualization effectually. Considerably, these case studies show probability of monitoring safety, as goods quality in supply chain management by merging WSNs. However, probable monitoring applications, security system with human agent robot machine sensor protocol for fire-fight system is offered to offer stable collaboration and faster communication. WSN node is accountable for data generation, while big data centre manages system functionalities in recommended architecture.

Finally, big data for smart cities are anticipated to determine feasible solutions for environment, public administration and urban services. With this, data is converged to mapping e-government services (digital mapping). Generally, e-government services are determined to enhance efficiency satisfaction owing to spatial data infrastructure [7]. Henceforth, it is validated that e-government validated WSN enhances e-service access. However, it is used to prevail over limited resources for managing challenges. This shows application analysis.

3. Technical modelling of Big data via WSN

Data constructed by sensors may exponentially grow. Traditional technologies for data storage, processing (relational databases and servers) is too costly to handle these data. However, it cannot handle processing essential that can be needed for real time process. Specifically, monitoring some events with regular intervals are mainly redundant or least variation leads to huge data storage resource wastage and energy for communication at sensor nodes and relay. This shows that data are redundant, meaningless and of no interest.

Therefore, indeed of WSNs, it is necessary to accumulate and transmit huge amount of data while reducing latency as in Figure 2. However, it is essential to effectually avoid data redundancy and enhance energy efficiency. Overlap among WSN and big data system relies on data processing utilization. For WSN, it may save limited resources [8]. Similarly, non-redundant, clean and appropriate data may diminish excess data volume. Therefore, it may diminish overload by introducing data values.

Research challenges with respective to WSN based big data has to handle securing, clustering, processing and saving energy. Anticipated strategies for handling these confront are dependent on correlation among data clustering, aggregation and energy consumption challenges. Data collection offers data transferring strategy with large scale WSN context and analysis big data collection challenges. For instance, data gathering, analysis and data storage produced by WSN for observing pollution level is depicted. Numerous prototypes were modelled for open source technologies like stream processing, distributed computing with Hadoop and WSN over Arduino platform. This criterion includes network protocol and WSNs applications as in Figure 3 with data processing technologies for big data systems.
Figure 2. Analysis with big data

a. Classification
Investigations over big data system in WSN are extremely classified into two significant fields: data system and network system. Data system is specifically for data processing and network system is for delivering sensor data to big data system. Every investigation concentrates in various research objectives.

Figure 3. Big data based WSN classification

b. Network System
In various research works, investigations are conducted by anticipating either communication protocols or network architectures. Initially, WSN structural construction for data processing is termed as service-oriented architecture and virtualization simulates numerically with direct forwarding comparison for WSN architecture [9]. It comprises of for layers: users, cloud centres, gateway and large scale WSN. Here, there are two serious issues: communication latency and congestion with big data. To handle this issues, data processing and flexible storage with cloud computing is anticipated.
Another confronting challenge for accumulating big data over dense distributed sensor network is acquiring energy efficiency. This approach can be utilized to demonstrate data-gathering and sink node’s trajectory via clustering. Unlike conventional clustering strategy, $k$-centroid clustering is anticipated to maintain energy consumption in continuous iteration. Here, sinks mobility was chosen as an effectual solution to deal with big data collection. However, structure to leverage correlation. Next, utilization of mobile data collector is performed with two approaches: data collection with data mule and SN with AP. These models are categorized with number of hops essential to deal with unexpected partition of data. It also deals with routing protocol in clustered architecture with WSN multi-media data delivery.

It attempts to eliminate network congestion, to enhance data transmission reliability and reduction of packet loss rate [10]. To acquire simultaneous goal, an effectual load balancing approach is anticipated to guarantee balance energy consumption.

Along with data collection, data aggregation strategy for big data is analyzed with information centric networking, where data are acquired from in-network caching and names. The anticipated structure works with following phases:
1) Initializing network and cluster the communication nodes with low-energy adoptive clustering hierarchical protocol;
2) Aggregate collected data from cluster head memory;
3) Aggregated routing approach retrieves data and forwards it to data centre. From this perspective, WSN architecture is designed to eliminate excessive energy consumption.

A framework termed structure fidelity enhances spatial correlation among nodes by eliminating number of active sensor nodes by accomplishing lower structural distortion of data collected. Nodes’ duty cycle is managed by distortion is based on quality assessment factor [11]. Therefore, data fidelity of structural similarity in constant sensing of WSN applications is achieved. As well, compression approach for data collection at primary level and avoiding redundant data produced by second level neighbouring node. $k$-means and $Top-k$ approach uses clustering approach reduces communication cost and improve WSN data mining.

As an instance of routing protocol, a novel routing model is designed to allocate priority in accordance to Quality-of-Service requirements and acquire load distribution by huge amount of SNs in path as in Table I. Specifically, energy efficiency is attained by multiple hops according to accessible resources and energy cost.

| Features     | Disadvantages               | Advantages                  |
|--------------|-----------------------------|-----------------------------|
| Multi-layer  | Higher maintenance cost     | Superior lifecycle          |
| $k$-means    | Higher maintenance cost     | Higher energy efficiency    |
| Mobile sink  | Higher energy consumption   | Finest network partition    |
| QoS          | Higher complexity           | Balancing energy consumption|
| Caching      | Low energy consumption      | Higher energy efficiency    |
| Correlation  | No temporal correlation      | Higher energy efficiency    |

Table 1. Big data categorization

c. Security

While gathering data, WSN carry out both transport and capturing data, it is essential to achieve two tasks in secured way. To resolve this crisis, analysis and discussions over numerous architectural executions that attempts to offer robustness and efficiency for external attack and internal compromise [12]. Then, a novel architecture is anticipated for transportation and data capture as in Figure 4. As well, forged or misleading data gathering happens in WSNs. Henceforth, crucial and sensitive data transmission via secure communication is needed.

While measuring nodes constraint, symmetric cryptography is extremely applicable for WSN owing to efficiency. Moreover, symmetric cryptograph has to work with distributed key management. To diminish distribution overhead centralized state-full connection offers effectual key management for SNs. It can acquire superior balance among security and efficiency is attained with key encryption at nodes’ beginning.
4. Data System

While network system concentrates on delivering data, this system concentrates on effectual data processing that is transmitted through WSN. Data system objectives are extremely diverse than network modelling; it includes data collection, infra-structure, analysis, processing, security and management as depicted in Table II.

| Features                | Advantages                      | Disadvantages     |
|-------------------------|---------------------------------|-------------------|
| Hadoop framework        | Open source                     | Low reliability   |
| Data centric storage    | Highly energy efficient          | Deployment issues |
| 3T data mining          | Efficient response              | Experimental scenarios |

a. Data Collection

Even though numerous approaches for data collection were introduced in network system infrastructure, diverse algorithms are performed for computation [13]. It shows that data collection via mobile sink and static node that are not changed. Moreover, novel procedures and models are depicted in accordance to system requirements.

b. Data Processing

Here, WSN in-network processing is described here: fusion technologies and data aggregation. Discussions and overviews of existing data fusion approaches and data mining were designed for WSNs. It shows that sensor data processing over network is processed further assist in saving limited resources and eliminates excess data duplication as in Table III.

| Characteristics          | Advantages                      | Disadvantages     |
|--------------------------|---------------------------------|-------------------|
| Event awareness          | Cost efficient                  | Weight based computation |
| M-Mobile collector       | Highly energy efficient          | Simulation scenario |
| Mobile collector         | Highly energy efficient          | Infinite storage memory |
| Data collector           | Highly energy efficient          | Dependency on threshold value |
| Mobile data collector    | Reduced latency                 | Flooding based operation |

c. Data Management

However, big data frameworks and tools are initiated to examine query processing and data collection. As major WSN constraint relating data management, energy preservation has to be focussed. In energy efficiency aspects, decentralization emphasizes promising solutions to acquire energy preservation by computation task distribution between sensor nodes.
d. **Data Analysis**

Indeed, of big data analysis, it is essential to demonstrate system architecture. To fulfill this need, a novel framework that comprises of Hive data warehouse and Hadoop over MapReduce. To carry out MapReduce functionality, HiveQL and SQL-like language is carried out to perform queries [14]. With these techniques, data warehouse modelling is anticipated to examine entire data gathered and identify abnormal characteristics is anticipated.

To show data analysis outcome, fast identification and detection of location error in sensor datasets has to be offered. To attain this objective, error detection model utilizes complete computational probable of cloud platform and WSN network features. Here, data errors are defined and classified with WSN cluster. Specifically, error detection over network topology and detection operations are performed in restricted with data blocks or temporal data blocks. Subsequently error detection and big data cleaning for error issues are resolved. Performance is enhanced by utilizing classification and network topology at similar time. However, sensor processing is used for application to construct data cleaning infra-structures.

e. **Security**

To identify correct and efficient intrusion, WSN is generally an appropriate solution; it is defend over insider attack with appropriate trust based approach. Moreover, owing to excess data, trust computation is significantly degraded. To eliminate degradation, trust management with under hierarchical structure is designed. Experimental outcomes demonstrate that anticipated model enhances trust management by identifying malicious node rapidly.

5. **Open Challenges**

This section discusses about the open challenges in the network model.

a. **Internet-of-Things (IoT)**

IoT system has to be engaged with WSNs characteristics and effectual communication systems. Indeed, of typical WSN, IoT system utilizes superior performance of energy efficiency and latency in real implementations as depicted in Table IV. IoT utilizes and gathers real time data to optimize functionalities, identify mal-functionalities, security breaches. Henceforth, an IoT system has to include real time streaming data management, real-time decision making and real-time analytics.

| Research field | Challenges | Future directions |
|----------------|------------|-------------------|
| Architecture   | Mobile sink, aggregation | Application protocols, 5G communication |
| Real time environment | QoS protocol | IoT |
| System framework | WSN platform | Middleware |

b. **Application-Specific needs**

In general, WSN is implemented and deployed to achieve application specific requirements, as certain requirements have to be fulfilled. For instance, certain applications need scalar values, while other vector requirements [14]. In accordance to essential application values, data produced by sensor node holds diverse properties naturally and features. This relies that data system and network for big data has to be modelled with application requirements.

c. **Network modelling**

In WSN, mobile sink, big data systems and clustering are essential research challenges in data aggregation and collection. Moreover, these strategies have to be improved with significant diminishment in energy consumption and offer real time communication among huge data amount [15]. For instance, data aggregation algorithm with respect to PCA has to attain minimal delay and energy consumption is implemented in cluster head.

6. **Conclusions**

In this work, research challenges for managing big data are measured as a major data sources. Before concentrating towards major contributions of this chapter, opportunities related to data management in
big data system with WSN is presented in detail. Specifically, various researches concentrate only on research field and its related objectives. Then, data collection in WSN is concentrated with processing of in-network while determining unique features of network properties. At last, various research issues in managing data and processing data is discussed.

References
[1] Chen, C.L.P.; Zhang, C. Data-intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. *Inf. Sci.* 2014, 275, 314–347
[2] Xiao, F.; Zhang, C.; Han, Z. Editorial: Big Data in Ubiquitous Wireless Sensor Networks. *Int. J. Distrib. Sens. Netw.* 2014
[3] Harb, H.; Idrees, A.; Jaber, A.; Makhoul, A.; Zahwe, O.; Taam, M. Wireless Sensor Networks: A Big Data Source in Internet of Things. *Int. J. Sens.* Wirel. Commun. Control 2017, 7, 141–149.
[4] Sezer, O.; Dogdu, E.; Ozbayoglu, A. Context-Aware Computing, Learning, and Big Data in Internet of Things: A Survey. *IEEE Internet Things J.* 2018, 5, 1–27
[5] Jaradat, M.; Jarrah, M.; Bousselham, A.; Jararweh, Y.; Al-Ayyoub, M. The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid. *Procedia Comput. Sci.* 2015, 56, 592–597
[6] Du, Y.; Hu, F.; Wang, L.; Wang, F. Framework and Challenges for Wireless Body Area Networks Based on Big Data. In *Proceedings of the IEEE International Conference on Digital Signal Processing, Singapore*, 21–24 July 2015
[7] Wu, D.; Shi, H.; Wang, H.; Wang, R.; Fang, H. A Feature-based Learning System for Internet of Things Applications. *IEEE Internet Things J.* 2018.
[8] Fouad, M.; Oweis, N.; Gaber, T.; Ahmed, M.; Snasel, V. Data Mining and Fusion Techniques for WSNs as a Source of the Big Data. *Procedia Comput. Sci.* 2015, 65, 778–786
[9] Song, C. A Novel Wireless Sensor Network Architecture Based on Cloud Computing and Big Data. *Int. J. Online Eng.* 2017, 13, 18–25.
[10] Xu, Z. The Analytics and Applications on Supporting Big Data Framework in Wireless Surveillance Networks. *Int. J. Soc. Humanist. Comput.* 2017, 2, 141–149.
[11] P. Sherubha, “A detailed survey on security attacks in wireless sensor networks”, *International Journal of Soft Computing*, 2016.
[12] P. Sherubha, M. Banu chitra, “Multi class feature selection for breast cancer detection”, *International journal of pure and applied mathematics*, 2018.
[13] P. Sherubha, P. Amudhavalli and S.P. Sasirekha, “Clone Attack Detection using Random Forest and Multi Objective Cuckoo Search Classification”, *International Conference on Communication and Signal Processing*, 2019.
[14] P. Sherubha, N. Mohanasundaram,” An Efficient Intrusion Detection and Authentication Mechanism for Detecting Clone Attack in Wireless Sensor Networks”, *Jour of Adv Research in Dynamical & Control Systems*, Vol. 11, No. 5, 2019.
[15] P. Sherubha, “An Efficient Network Threat Detection and Classification Method using ANP-MVPS Algorithm in Wireless Sensor Networks”, *International Journal of Innovative Technology and Exploring Engineering*, 2019.