Sensor Searching Techniques in Internet of Things: A Survey, Taxonomy, and Challenges

Farzana Noshin Choudhury, Fahrin Rahman, Md. Rafsan Jamil, and Nafees Mansoor

Abstract Along with the fast development of sensing technologies and smart devices, endless automation opportunities are expected in every sphere of modern life. Hence, the IoT concept plays a vital role in managing and controlling the wireless devices over the Internet. Therefore, IoT is considered to be the third wave of information and communication technology (ICT) after eras of the Internet and cellular networks. Besides, the post-COVID-19 world is expected to more sensor-centric to ensure lesser human interactions. Since a humongous number of sensors are intended to be deployed in IoT, efficient data communication requires the network to be clustered physically or logically. Consequently, the selection of the appropriate sensor(s) for data processing and gathering is vital in IoT. Several sensor selection techniques in IoT have been proposed recently, still, sensor searching remains quite a new research field. Therefore, several sensor searching methods for IoT are studied and presented in this paper. Moreover, the strengths and limitations of the existing sensor searching techniques are also outlined in the paper. Hence, the new performance metrics are presented in the paper, where the existing techniques for searching are analyzed.

Keywords Internet of things · Routing protocols · Sensor searching algorithm · IoT survey · Sustainable cities and communities

F. N. Choudhury · Md. R. Jamil
Ahsanullah University of Science & Technology, Dhaka, Bangladesh
e-mail: noshin.eee@gmail.com

Md. R. Jamil
e-mail: rafsan.shovon@gmail.com

F. Rahman
Islamic University of Technology, Dhaka, Bangladesh
e-mail: fahrin_rahman@yahoo.com

N. Mansoor (✉)
University of Liberal Arts Bangladesh (ULAB), Dhaka, Bangladesh
e-mail: nafees@ieee.org

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
S. Fong et al. (eds.), ICT Analysis and Applications, Lecture Notes in Networks and Systems 154, https://doi.org/10.1007/978-981-15-8354-4_74
1 Introduction

Internet of things (IoT), a pervasive technology, is getting noticed in the communication research circle for the past few years [1]. The term “IoT” came into the spotlight after Kevin Ashton coined it into 1999. IoT is the wide network containing physical objects or “Things” planted with electronics, software, and connectivity. The objective of IoT can be described as enabling and achieving bigger value and service through exchange of data with the manufacturer, operator, and/or other connected devices. Exchanging information or data is based on the infrastructure of the International Telecommunication Union’s Global Standards Initiative [2]. The “Things” in the IoT can be considered as an animal having a biochip transponder or a device with a built-in sensor. Hence, these sensors are implemented around various geographical locations. Those sensors are also managed by different organizations.

Search engines are considered as the most vital part of the Internet. People need search engines to find their desired information in the shortest possible time. Searching in IoT is more challenging in case of searching for documents on the Web using search engines. Complex descriptions can be a feature of IoT. Thematic, spatial, and temporal dimensions can be the mode of representation of the data produced by IoT objects [3]. Again, loss of communication or malfunctioned wireless sensor nodes can bring change in the status of IoT objects. So, traditional searching techniques are not appropriate and sufficient for IoT-based applications.

Sensors implemented around the world are increasing day by day. The observation and maintenance (O&M) data of the sensors are of high importance in the Internet of things. The research and survey considering sensor searching techniques in IoT are not sufficient so far [4]. The existing searching techniques in IoT, their advantages and drawbacks, and their comparative performance analysis have been presented here.

Organization of the paper can be mentioned as the following: In Sect. 2, a discussion on different recently developed sensor searching techniques for IoT is presented. The proposed performance metrics are discussed in Sect. 3. An analysis of the existing techniques based on the proposed metrics is described in Sect. 4. Conclusion and future works have been discussed in Sect. 5.

2 Existing Techniques

In recent years, different searching techniques have been introduced to the field of IoT. A large volume of data is generated from numerous sensors. Appropriate searching technique matters a lot in collection and processing of that data. This section represents a brief discussion on the existing techniques.

Context-aware sensor search, selection and ranking model, abbreviated as CASSARAM, is a user-priority-based sensor searching technique in IoT [5]. In CASSARAM, requirements of the users are divided into the following categories:
point-based and proximity-based. In this algorithm of CASSARAM, point-based requirements that are identified and specified by the user are non-negotiable entities in the CASSARAM technology. Moreover, in CASSARAM, proximity-based requirements are considered to be negotiable where these requirements are not always considered to be user-specific. Proximity-based requirements include characteristics of sensors that include reliability, accuracy, etc., which refer to context property. Semantic Sensor Network Ontology (SSNO) is the modeling technique of sensor descriptions and context properties. SSNO stores the context properties in original measurement units (accuracy in percentage or latency). For consistency, the context properties are normalized to [0, 1]. Moreover, relational expression-based filtering is used to reduce the complexity of large queries. Finally, CPWI decides the rank of the sensor on proximity-based user requirements. To overcome the limitation of CASSARAM, comparative priority-based heuristic filtering (CPHF) is applied which performs the function of removing the sensors positioned far away from the user-defined ideal sensors [5]. In this case, distributed sensor searching can improve scalability and efficiency.

Searching with textual metadata does not work well in practice. Because search input can be in different forms. When we enter data, there are human-based mistakes. Different terms are used to describe the similar concept or important metadata is not entered. Adopting the traditional search-by-example approach can solve this problem to some extent [6]. A time series of sensor values are specified by the user for performing specific search. A comparison is made between the time series and the indexed fuzzy sets. A fuzzy set is a mapping of a set of real numbers that allows partial membership instead of binary membership. It is computed by the sensor itself. Fuzzy sets-based method can be used to compute a similarity score for each indexed sensor based on their output. When the similarity score is higher, then the measurement ranges overlap more, and more the sensor measurements belong to the fuzzy set defined by the output of the given sensor [6].

Numerous sensors are functioned for Internet access purpose. But there are some problems to ensure efficient searching. User has to model sensor data as Bayesian network (BN). This BN model is essential for forming perfect prediction of desired sensors and should be constructed, and then the probability table can be formed. BN is such a model by which a small probability table can be formed at every node which have all nodes connected to the network. Correlations can be modeled as the dependency relationship between pair of nodes. Sought state, number of matching sensors, total set of sensors, and set of sensors (contracted) are the parameters here. Bayesian network table is also used here to separate the sensors which are non-contacted [7].

For network modeling, distributed regression is used which is an efficient and general framework. The algorithm follows kernel linear regression. Shape, size of sensor data, and actual complete sensor data can be extracted through this research method [8]. The complexity and requirements can be limited so that the regression method can work perfectly to express the perfect structure and type of data. Thus, it is reducing the cost. There are two layers—junction tree and routing tree, and they are for probability function and data collection. From this model, the particular
data or model coefficient can be transmitted. This is a full algorithmic work where sensor functions and basis matrices can be implemented to get dot product matrix and project measurement vector. From these, particular coefficient can be found. The routing tree can perform transmission. It is also making it possible to access new locations. Distributed regression algorithm is such an algorithm that is built on a distributed application of Gaussian elimination. The complexity of this algorithm depends on the cluster size of each node but does not depend on the number of parameters. This research method can make connection between two sensor nodes even though they are not neighbors. Error correction can be performed at the time of implementation [8].

Shodan is a powerful IP search engine which enables notification for the IP exposure for the IoT devices. This web-based notification system is developed because the techniques for preventing IP-related security threats have been proven insufficient. Searching data for IP addresses is provided by Shodan. In just one month, Shodan engine collects approximate 500 + million IoT devices [2]. Computers that are connected to the Internet are found by the user by means of various filters in Shodan. It provides two types of information that is Web page and script console using IP search engine application programming interface (API). The Web page provides only the basic data (IP address, connected country, city etc.) and also a server name. But script console contains more useful data like port number, affiliation, latitude, longitude, etc., which are inherently more useful than data from Web page. The script API of Shodan is used to set up the IP exposure notification system. In this system, exposed IoT device information is objectified. Object is marked on Google Maps from longitude and latitude information. From these marks, it can be identified how many IP exposures are there in that particular area. If advantage taken, dangerous IP exposure can be denoted.

3 Performance Metrics

To get a bird’s eye view detail of the research about existing searching techniques, it is convenient in defining some parameters or dimensions against which the existing techniques can be compared. The metrics are presented as follows:

- **User’s criterion**: A technique for providing the opportunity to input user’s criterion that can generate output matching user’s requirements. A user-friendly technique consumes less time to perform the searching in IoT as the user gets easily acquainted with the method.
- **Realistic application**: It describes whether the working prototype supports experiments using realistic applications or not. Experiments using realistic applications demonstrate how much the technique is effective in a realistic environment.
- **Provision for sensing as a service**: Sensing-as-a-service model is one of the greatest priorities in IoT today. It refers to the new Internet of things business model. Till now infrastructure as a service (IaaS), software as a service (SaaS),
etc., have been introduced in IoT. Provision for sensing as a service enables pay-per-use system and free and paid system for sharing.

- Scalability: Scalability means the capacity of handling a growing amount of work by a system, network, process, or application. A scalable searching technique can deal with millions of sensor nodes which is an advantage for IoT.

- Robustness: Millions of “Things” are connected to IoT. So, the searching technique needs to be robust as any kind of failure in the sensor networks affects the data collection process. Robust techniques encourage effective searching in IoT for a long time period without disruption.

- Machine learning: Machine learning refers to artificial intelligence. In this process, computers become able to learn without being explicitly programmed. IoT devices are generating a large amount of data. In this case, machine learning can help to improve efficiency and reduce costs by pursuing those data.

- Access to the new location: In a particular technique, a sensor network should identify the new location. It should have the ability to access various locations where sensors are deployed.

- Threat detection and prevention: Security is a very important issue in IoT because all the devices are connected through the Internet. Any kind of harm can reach the devices as IP-related threats are increasing. Hence, threat detection and prevention are essential.

4 Performance Analysis

This section represents the overall comparative analysis of the techniques. Basically, the analysis can help the user to select the best searching method satisfying his requirements (Table 1).

In CASSARAM, context information is used to select sensors matching user’s criteria. The algorithm used in this technique can capture user priorities.

| Table 1 Summary of analysis                                      |
|-------------------------------------------------------------------|
| Technique            | User’s criterion | Sensing as a service | Scalability          | Machine learning | Access to location | Threat detection |
|----------------------|------------------|----------------------|----------------------|------------------|--------------------|------------------|
| CASSARAM             | Yes              | Yes                  | Not enhanced         | Supported        | No                 | No               |
| Fuzzy sets           | Yes              | No                   | Enhanced             | Not supported    | Yes                | No               |
| Bayesian network     | Yes              | No                   | Enhanced             | Not supported    | Yes                | No               |
| Distributed regression | Yes            | No                   | Enhanced             | Not supported    | Yes                | No               |
| Shodan               | Yes              | No                   | Not enhanced         | Not supported    | No                 | Yes              |
CASSARAM also supports realistic application and is helpful for broadening Sensing-as-a-service vision. Scalability along with efficiency is enhanced in this method. It has the ability of running for a long time without failure. Future investigation is needed to incorporate machine learning by heuristic algorithm in CASSARAM. Access to the new location and threat detection and prevention are not provided by CASSARAM.

Fuzzy set-based sensor search: In sensor similarity searching using fuzzy sets, sensor gives exact outputs as earlier. The technique can be used for realistic application such as searching videos and images. Sensors give output as per user demands satisfying user’s criteria. Suppose, if a user wants to know the places with same climate condition, some temperature sensors with the same output will be generated. This method does not support provision for sensing as a service, machine learning, and threat detection and prevention. This method is not highly scalable but efficient in this sense that it returns the search result to the user with a fast response. Fuzzy set represents output with a few bytes and needs more work to search among large number of sensors.

Bayesian network-based sensor search: This method supports user’s criteria. Bayesian network is essential for forming perfect prediction of desired sensors. This method supports realistic applications. Access to the new location is supported here. Provision for sensing as a service is not supported here. Sensors are read out randomly until it reaches to a point where matches can be found for a given number. Remote sensor readouts can be reduced accurately in this approach, and it improves searching for sensors that exhibits a given state when it is time for query for forming a prediction model. High scalability, machine learning, and threat detection are not provided here.

Distributed regression: User can define his requirements in this approach. This method is useful for drawing out more absolute information regarding the shape and structure of sensor data which ensures accuracy. Network nodes can fail due to lossy wireless transmission. Luckily, the distributed regression can be formed robust to such failures. This algorithm is really scalable as it can deal with large sensor networks consisting of hundreds of sensors. By building contour plots, analyzing number of sophisticated in-network applications, it is possible to have practical implications. This method does not support provision of sensing-as-a-service and machine learning techniques.

Shodan search engine is used to ensure threat detection and prevention by introducing IP exposure notification system. The whole working mechanism can be carried out using real-life IoT devices so it supports realistic applications. The system can work with millions of IoT devices so it is definitely scalable. Access to the new location and machine learning technique is not supported in this system. The system prevents the threats for exposed IoT devices. Exposed IP address can be found, and the level of exposure can be estimated in this system. Thus, the risk can be reduced. IP exposure threats can be reduced in state institutions, ranging from personal to national.
5 Open Research Area

Though the existing techniques can provide us with the best utilization of the sensor networks deployed in IoT, the scope of future research can be elaborated to the following sectors:

Scalability: Scalability in wireless sensor network (WSN) actually refers to scaling a network with a range of high node numbers and a bit of high node density [9, 10]. Implementation with global sensor networking (GSN) system may be approachable for developing a generic platform for deploying sensor networks and processing data produced by the sensor network in a distributed fashion.

Sensing as a service: Since humongous devices and sensors are getting connected to the Internet, IoT is becoming a major topic in this technology driven era. Some major vital measurements can be obtained through device management, the network, and the collected data. The management and provisioning of such sensor devices and data can create new business opportunities with some new challenges. Industry and academic—both the sectors—need to manage the devices that are interconnected and utilize the opportunity represented by the huge amount of data generated. Again, sensor network infrastructure involves huge investment as well as high maintenance cost. It holds users back from setting up their own IoT systems and Web applications that exploit sensor data. Mobile devices or sensor networks are occupied with sensor data through some sort of operations so that the data supplied by the sensor providers can be subscribed and published. Cloud-based sensing as a service (CSaaS) can be proposed as it can engage and manage various types of sensors on IoT devices by using the layer of virtualization; for in-depth analysis regarding cloud system, data is collected from sensors and sensor networks. In SaaS approach, real time and historical data can be accessed for analysis, and it can be referred as a convenient and standardized sensing data service for the consumers and third-party applications that are data-dependent. Dynamic provisioning for users is enabled here to leverage the huge pool of resources on demand [11].

Machine learning: Machine learning is becoming a major key point in IoT enhancing artificial intelligence between IoT devices as well as resulting in energy savings. WSN consumes a major percentage of its energy. Specially, for higher required accuracy, higher computational requirements are needed, hence higher energy consumptions occur [12]. For large-scale energy-constrained sensor networks, transmitting all data directly to the sink cannot be considered as an efficient process. There is an effective and efficient solution for passing data to a local aggregator which is termed as a cluster head (CH) [13]. CH collects data from all the sensors within its cluster and then transmits data to the sink. Proper CH selection is very significant as it will reduce energy consumption and also enhance the network’s lifetime [14]. Distributed machine learning techniques may be suitable for limited resource devices such as WSNs. Distributed machine learning methods need less computational power and also smaller memory footprint, compared to centralized learning algorithms [15]. This technique can help selecting proper CH, and
through this method, the system may be energy efficient and machine learning may be enhanced.

Threat detection and prevention: It is crucial to improve the process of threat detection and prevention in IoT as a numerous number of sensors are connected. However, it is becoming challenging task because there are many factors that can cause faults and can influence data.

Body area network: Some intelligent sensor nodes should capture various physiological signals such as—sensing, sampling, processing, and communicating physiological signals. The sensor used in wireless body area network (WBAN) would have to be low energy consuming, less complex, having smaller form factor as well as light in weight, power efficient, user-friendly, and reconfigurable. Data existing on multiple mobile devices and wireless patient nodes needs to be collected and analyzed in a seamless fashion. ZigBee (802.15.4) or Bluetooth (802.15.1) can be used for designing the nodes to sample vital signs and transferring relevant data to a personal server through a wireless personal network. Different platforms with variety of wide area network (WAN) access possibilities for Internet access can be used to accommodate a personal server application. [16]. Selection of platform is system-specific and needs to be selected to reduce obtrusiveness for a given user. An algorithm can be proposed for routing in WBAN.

6 Conclusion

In the near future, billions of objects are going to be included in IoT. A lot of sensors are expected to be attached with those objects resulting in increasing number of sensors around us. To get desired data from those sensors and processing them for future applications, knowledge of appropriate sensor searching method is very crucial. For sensor searching criteria, sensors should be paired with IoT applications. Clustered sensor networks raise the challenge of searching and selecting sensors appropriately. For any query, the algorithm should search and select the right sensors for capturing data and combining together with numeric and semantic reasoning. The algorithm in a searching technique should be defined the in such a way so that the users do not need to access all the sensors for collecting data. Again, search algorithms can broaden the research area by focusing on the priority issues. Categorization of searching results on the basis of priorities can be added as a major feature of the search engines. From this research, the users can be benefitted by getting a clear concept of the overall analysis and comparison of the existing techniques including sensor search, sensor similarity search, and providing security.
References

1. Ashton, K. (2009). That internet of things thing. *RFID Journal, 22*(7), 97–114.
2. Ko, Y.-S., Ra, I.-K., & Kim, C.-S. (2015). A study on IP exposure notification system for IoT devices using IP search engine shodan. *International Journal of Multimedia and Ubiquitous Engineering, 10*(12), 61–66.
3. Farin, N. J., Rahman, A., Mansoor, N., & Hossain, S. (2016). Wotcoms: A novel cross-layered web-of-things based framework for course management system. In *Proceedings of the First International Conference on Advanced Information and Communication Technology (ICAICT-16)*.
4. Zhou, Y., De, S., Wang, W., & Moessner, K. (2016). Search techniques for the web of things: A taxonomy and survey. *Sensors, 16*(5), 600.
5. Perera, C., Zaslavsky, A., Christen, P., Compton, M., & Georgakopoulos, D. (2013). Context-aware sensor search, selection and ranking model for internet of things middleware. In *2013 IEEE 14th International Conference on Mobile Data Management (MDM)* (Vol. 1, pp. 314–322). IEEE.
6. Truong, C., Romer, K., & Chen K. (2012). Fuzzy-based sensor search in the web of things. In *2012 3rd International Conference on Internet of Things (IOT)* (pp. 127–134). IEEE.
7. Mietz, R., & Römer, K. (2011). Exploiting correlations for efficient content-based sensor search. In *SENSORS, 2011*. IEEE.
8. Guestrin, C., Bodik, P., Thibaux, R., Paskin, M., & Madden, S. (2004) Distributed regression: An efficient framework for modeling sensor network data. In *Information Processing in Sensor Networks, 2004. IPSN 2004. Third International Symposium* (pp. 1–10). IEEE.
9. Liao, W.-H., Chen, C. C. (2010). A multi-dimensional data storage algorithm in wireless sensor networks. In *Network Operations and Management Symposium (NOMS)*. IEEE.
10. Rumín, A. C., Pascual, M. U., Ortega, R. R., & López, D. L. (2010). Data centric storage technologies: Analysis and enhancement. *Sensors (Basel), 10*(4): 3023–3056. Published online 2010 Mar 30. https://doi.org/10.3390/s100403023.
11. Kim, M., Asthana, M., Bhargava, S., Iyyer, K. K., Tangadpalliwar R., et al. (2016). Developing an on-demand cloud-based sensing-as-a-service system for internet of things. *Journal of Computer Networks and Communications*, 3292783, 17.
12. Meshram, N. A., & Thakare, V. M. (2015). Secured wireless sensors network using machine learning approach. In *IEEE Conference Paper, Advanced Technologies in Computing and Networking*.
13. Tasmim, S., Kamal, A. H., Tusher, M. O., & Mansoor, N. (2020). DEB: A delay and energy-based routing protocol for cognitive radio ad hoc networks. Algorithms for intelligent systems. In *Processing of International Joint Conference on Computational Intelligence* (pp. 643–654). Springer.
14. Mansoor, N., Islam, A. M., Zareei, M., Baharun, S., & Komaki, S. (2016, November). A novel on-demand routing protocol for cluster-based cognitive radio ad-hoc network. In *2016 IEEE Region 10 Conference (TENCON)* (pp. 632–636). IEEE.
15. Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H.-P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials, 16*(4) 1996–2018. arxiv.org/abs/1405.4463
16. Otto, C. (2006). An Implementation of a wireless body area network for ambulatory health monitoring. Huntsville, Alabama