IoT and Big Data: An Architecture with Data Flow and Security Issues

Deepak Puthal[1(✉)], Rajiv Ranjan[2], Surya Nepal[3], and Jinjun Chen[4]

[1] School of Computing and Communications, University of Technology Sydney, Ultimo, Australia
deepak.puthal@gmail.com
[2] School of Computing Science, Newcastle University, Newcastle upon Tyne, UK
rranjans@gmail.com
[3] CSIRO Data61, Canberra, Australia
Surya.Nepal@data61.csiro.au
[4] Swinburne Data Science Research Institute, Swinburne University of Technology, Melbourne, Australia
jinjun.chen@gmail.com

Abstract. The Internet of Things (IoT) introduces a future vision where users, computer, computing devices and daily objects possessing sensing and actuating capabilities cooperate with unprecedented convenience and benefits. We are moving towards IoT trend, where the number of smart sensing devices deployed around the world is growing at a rapid speed. With considering the number of sources and types of data from smart sources, the sensed data tends to new trend of research i.e. big data. Security will be a fundamental enabling factor of most IoT applications and big data, mechanisms must also be designed to protect communications enabled by such technologies. This paper analyses existing protocols and mechanisms to secure the IoT and big data, as well as security threats in the domain. We have broadly divided the IoT architecture into several layers to define properties, security issues and related works to solve the security concerns.

Keywords: Internet of Things · Big data · Security · Security threats · Quality of Service

1 Introduction

IoT is a widely-used expression but still a fuzzy one, due to the large number of concepts brought together to a concept. The IoT appears a vision of a future source of data where sensing device, possessing computing and sensorial capabilities can communicate with other devices using Internet protocol. Such applications are expected to bring a large total of sensing and actuating devices, and in significance these costs will be a major factor. On the other hand, cost restrictions dictate constraints in terms of the resources available in sensing platforms, such as memory and computational power. Overall, such factors motivate the design and adoption of communications and security mechanisms

A. Longo et al. (Eds.): IISSC 2017/CN4IoT 2017, LNICST 189, pp. 243–252, 2018.
https://doi.org/10.1007/978-3-319-67636-4_25

optimized for constrained sensing platforms, capable of providing its functionalities efficiently and reliably.

Several of these applications are approaching the bottleneck of current data streaming infrastructures and require real-time processing of very high-volume and high-velocity data streams (also known as big data streams). The complexity of big data is defined through 5Vs: (1) volume– referring to terabytes, petabytes, or even exabytes (1000^6 bytes) of stored data, (2) variety– referring to unstructured, semi-structured and structured data from different sources like sensors, surveillance, image or video, medical records etc., (3) velocity– referring to the high speed at which the data is handled in/out for stream processing, (4) variability– referring to the different characteristics and data value where the data stream is handled, (5) veracity– referring to the quality of data. These features introduce huge open doors and enormous difficulties for big data stream computing. A big data stream is continuous in nature and it is important to perform real-time analysis as the lifetime of the data is often very short (data is accessed only once) [1, 2, 6, 7]. As the volume and velocity of the data is so high, there is not enough space to store and process; hence, the traditional batch computing model is not suitable.

Even though big data stream processing has become an important research topic in the current era, data stream security has received little attention from researchers [1, 2]. Some of these data streams are analysed and used in very critical applications (e.g. surveillance data, military applications, etc.), where data streams need to be secured to detect malicious activities. The problem is exacerbated when thousands to millions of small sensors in self-organising wireless networks become the sources of the data stream. How can we provide the security for big data streams? In addition, compared to conventional store-and-process, these sensors will have limited processing power, storage, bandwidth, and energy.

Big data in IoT environment is gaining lots of interest from global researcher. By focusing current research trend, we have given the data flow between the layers including research issues in IoT generated big data architecture. The main contributions of the paper can be summarized as follows:

- We have proposed IoT generated big data architecture while defining layer wise properties of IoT.
- Followed by, we have highlighted the security threats, issues and solutions of individual IoT layers.
- Finally, we have highlighted the security issues of big data in IoT.

The rest of this paper is organized as follows. Section 2 gives the background IoT layers and their features. Section 3 describes security threats of individual layers in IoT architecture. Section 4 presents the security issues and requirements in IoT generated big data streams. Section 5 concludes the paper.

2 IoT Architecture

The connection of physical things to the Internet makes it possible to access remote sensor data and to control the physical world from a distance. The IoT is based on

this vision. A smart object, which is the building block of the IoT, is just another name for an embedded system that is connected to the Internet [9]. Al-Fuqaha et al. in [10] clearly defined the individual elements of IoT, which includes identification, sensing, communication, computation, services, and semantics. There is another technology that points in the same direction as RFID technology. The novelty of the IoT is not in any new disruptive technology, but in the pervasive deployment of smart objects. IoT system architecture must guarantee the operations of IoT, which bridges the gap between the physical and the virtual worlds. Since things may move geographically and need to interact with others in real-time mode, IoT architecture should be adaptive to make devices interact with other things dynamically and support unambiguous communication of events [11]. We broadly divided the complete architecture of IoT into three different layers, such as source smart sensing device, communication (Networks) layer and cloud data centre as shown in Fig. 1. These layers can be related to the service layer of IoT, where service layer and interface layer are integrated into the data centre in our architecture. The service level architecture of IoT consists of four different layers with functionality such as sensing layer, network layer, service layer, and interfaces layer [11, 12].

- Sensing layer: This layer is integrated with available hardware objects (sensors, RFID, etc.) to sense/control statuses of things.
- Network layer: This layer supports the infrastructure for networking over wireless or wired connections.
- Service layer: This layer creates and manages services requirements according to the user's need.
- Interfaces layer: This layer provides interaction methods to users and applications.

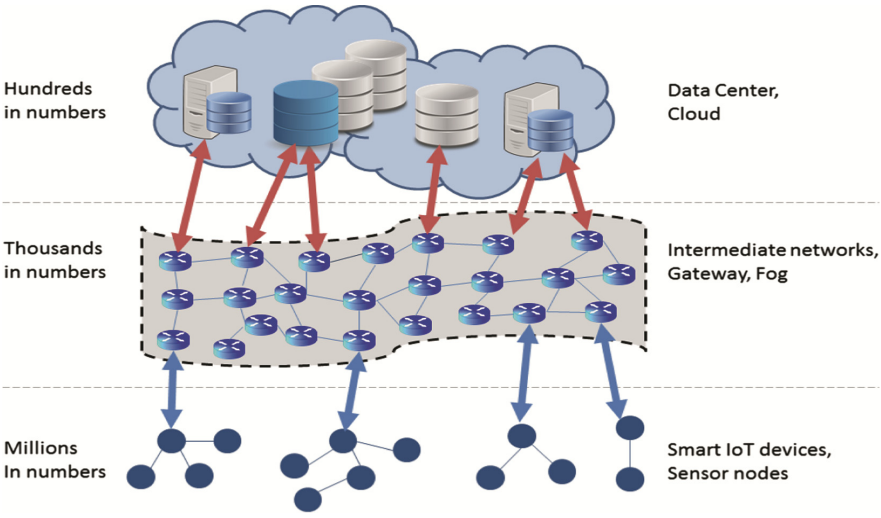

Fig. 1. Layer wise IoT architecture from IoT device to cloud data centre.

2.1 Sensing Layer

IoT is expected to be a world-wide physical inner-connected network, in which things are connected seamlessly and can be controlled remotely. In this layer, more and more devices are equipped with RFID or intelligent sensors, connecting things becomes much easier [13]. Individual objects in IoT hold a digital identity which helps to track easily in the domain. The technique of assigning a unique identity to an object is called a universal unique identifier (UUID). UUID is critical to successful services deployment in a huge network like IoT. The identifiers might refer to names and addresses. There are a few aspects that need to be considered in the sensing layer such as deployment (devices need to deployed randomly or incrementally), heterogeneity (devices have different properties), communication (needs to communicate with each other in order to get access), network (devices maintain different topology for data transmission process), cost, size, resources and energy consumption. As the use of IoT increases day by day, many hardware and software components are involved in it. IoT should have these two important properties: energy efficiency and protocols [11].

- *Energy efficiency:* Sensors should be active all the time to acquire real-time data. This brings the challenge to supply power to sensors; high energy efficiency allows sensors to work for a longer period.
- *Protocols:* Different things existing in IoT provide multiple functions of systems. IoT must support the coexistence of different communications such as ZigBee, 6LoWPAN etc.

2.2 Networking Layer

The role of the networking layer is to connect all things together and allow things to share information with other connected things. In addition, the networking layer is capable of aggregating information from existing IT infrastructures [4], data can then be transmitted to cloud data centre for the high-level complex services. The communication in the network might involve the Quality of Service (QoS) to guarantee reliable services for different users or applications [5]. Automatic assignment of the devices in an IoT environment is one of the major tasks, it enables devices to perform tasks collaboratively. There are some issues related to the networking layer as listed below [11]:

- Network management technologies including managing fixed, wireless, mobile networks
- Network energy efficiency
- Requirements of QoS
- Technologies for mining and searching
- Data and signal processing
- Security and privacy

Among these issues, information confidentiality and human privacy security are critical because of the IoT device deployment, mobility, and complexity. For information confidentiality, the existing encryption technology used in WSNs can be extended and deployed in IoT. Granjal et al. [3] divided the communication layer for IoT

applications into five different parts: Physical layer, MAC layer, Adaptation layer, network/routing layer, application layers. They also mentioned the associated protocols for energy efficiency as shown in Fig. 2.

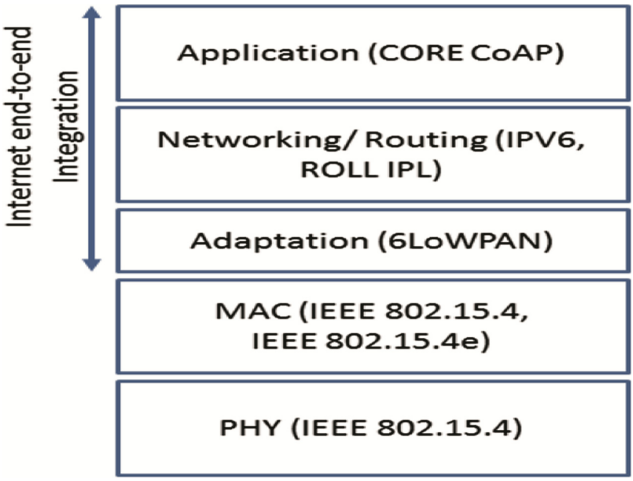

Fig. 2. Communication protocol in IoT.

2.3 Service Layer

A main activity in the service layer involves the service specifications for middleware, which are being developed by various organisations. A well-designed service layer will be able to identify common application requirements.

The service layer relies on the middleware technology, which provides functionalities to integrate services and applications in IoT. The middleware technology provides a cost-effective platform, where the hardware and software platforms can be reused. The services in the service layer run directly on the network to effectively locate new services for an application and retrieve metadata dynamically about services. Most of specifications are undertaken by various standards developed by different organisations. However, a universally accepted service layer is important for IoT. A practical service layer consists of a minimum set of the common requirements of applications, application programming interfaces (APIs), and protocols supporting required applications and services.

2.4 Interface Layer

In IoT, a large number of devices are involved; those devices can be provided by different vendors and hence do not always comply with same standards. The compatibility issue among the heterogeneous things must be addressed for the interactions among things. Compatibility involves information exchanging, communication, and events processing. There is a strong need for an effective interface mechanism to simplify the management

and interconnection of things. An interface profile (IFP) is a subset of service standards that allows a minimal interaction with the applications running on application layers. The interface profiles are used to describe the specifications between applications and services.

3 Security Threats of Each Layer

This subsection lists the security threats and security issues is each individual layer as divided in the above subsections.

3.1 Sensing Layer

The sensing layer is responsible for frequency selection, carrier frequency generation, signal detection, modulation, and data encryption [3, 14]. An adversary may possess a broad range of attack capabilities. A physically damaged or manipulated node used for attack may be less powerful than a normally functioning node. IoT devices use wireless communication because the network's ad hoc, large-scale deployment makes anything else impractical. As with any radio-based medium, there exists the possibility of jamming in IoT. In addition, devices may be deployed in hostile or insecure environments where an attacker has easy physical access. Network jamming and source device tampering are the major types of possible attack in the sensing layer. The features of sensing layers follow from Fig. 2.

Jamming: Interference with the radio frequencies nodes are using and

Tampering: Physical compromise of nodes.

3.2 Network Layer

The security mechanisms designed to protect communications with the previously discussed protocols must provide appropriate assurances in terms of *confidentiality*, *integrity*, *authentication* and *non-repudiation* of the information flows. Other relevant security requirements are *privacy*, *anonymity*, *liability* and *trust*, which will be fundamental for the social acceptance of most of the future IoT applications employing Internet integrated sensing devices. According to the communication protocol in IoT, we divided in five different layer as shown in Fig. 2.

MAC Layer. The MAC layer manages, besides the data service, other operations, namely accesses to the physical channel, validation of frames, guaranteed time slots, node association and security. The standard distinguishes sensing devices by its capabilities and roles in the network. A full-function device (FFD) can coordinate a network of devices, while a reduced-function device (RFD) is only able to communicate with other devices (of RFD or FFD types). By using RFD and FFD, IEEE 802.15.4 support topologies such as peer-to-peer, star and cluster networks [15].

Network Layer. One fundamental characteristic of the Internet architecture is that it enables packets to traverse interconnected networks using heterogeneous link-layer technologies, and the mechanisms and adaptations required to transport IP packets over particular link-layer technologies with appropriate specifications. With a similar goal, the IETF IPv6 over Low-power Wireless Personal Area Networks (6LoWPAN) working group was formed in 2007 to produce a specification enabling the transportation of IPv6 packets over low-energy IEEE 802.15.4 and similar wireless communication environments. 6LoWPAN is currently a key technology to support Internet communications in the IoT, and one that has changed a previous perception of IPv6 as being impractical for low energy wireless communication environments. No security mechanisms are currently defined in the context of the 6LoWPAN adaptation layer, but the relevant documents include discussions on the security vulnerabilities, requirements and approaches to consider for network layer security.

Routing Layer. The Routing Over Low-power and Lossy Networks (ROLL) working group of the IETF was formed with the goal of designing routing solutions for IoT applications. The current approach to routing in 6LoWPAN environments is materialized in the Routing Protocol for Low power and Lossy Networks (RPL) [16] Protocol. The information in the *Security* field indicates the level of security and the cryptographic algorithms employed to process security for the message. What this field doesn't include is the security-related data required to process security for the message, for example a Message Integrity Code (MIC) or a signature. Instead, the security transformation itself states how the cryptographic fields should be employed in the context of the protected message.

Application Layer. As previously discussed, application-layer communications are supported by the CoAP [17] protocol, currently being designed by the Constrained RESTful Environments (CoRE) working group of the IETF. We next discuss the operation of the protocol as well as the mechanisms available to apply security to CoAP communications. The CoAP Protocol [17] defines bindings to DTLS (Datagram Transport-Layer Security) [18] to secure CoAP messages, along with a few mandatory minimal configurations appropriate for constrained environments.

3.3 Service Layer (Middleware Security)

Due to the very large number of technologies normally in place within the IoT paradigm, a type of middleware layer is employed to enforce seamless integration of devices and data within the same information network. Within such middleware, data must be exchanged respecting strict protection constraints. IoT applications are vulnerable to security attacks for several reasons: first, devices are physically vulnerable and are often left unattended; second, is difficult to implement any security countermeasure due to the large scale and the decentralised paradigm; finally, most of the IoT components are devices with limited resources, that can't support complex security schemes [19]. The major security challenge in IoT middleware is to protect data from data integrity, authenticity, and confidentiality attacks [20].

Both the networking and security issues have driven the design and the development of the VIRTUS Middleware, an IoT middleware relying on the open XMPP protocol to provide secure event driven communications within an IoT scenario [19]. Leveraging the standard security features provided by XMPP, the middleware offers a reliable and secure communication channel for distributed applications, protected with both authentication (through TLS protocol) and encryption (SASL protocol) mechanisms.

Security and privacy are responsible for confidentiality, authenticity, and nonrepudiation. Security can be implemented in two ways – (i) secure high-level peer communication which enables higher layers to communicate among peers in a secure and abstract way and (ii) secure topology management which deals with the authentication of new peers, permissions to access the network and protection of routing information exchanged in the network [21]. The major IoT security requirements are data authentication, access control, and client privacy [8]. Several recent works tried to address the presented issues. For example, [22] deals with the problem of task allocation in IoT.

4 Security Issues in IoT Generated Big Data Streams

Applications dealing with large data sets obtained via simulation or actual real-time sensor networks/social network are increasing in abundance [23]. The data obtained from real-time sources may contain certain discrepancies which arise from the dynamic nature of the source. Furthermore, certain computations may not require all the data and hence this data must be filtered before it can be processed. By installing adaptive filters that can be controlled in real-time, we can filter out only the relevant parts of the data thereby improving the overall computation speed.

Nehme et al. [24] proposed a system, StreamShield, designed to address the problem of security and privacy in the data stream. They have clearly highlighted the need for two types of security in data stream i.e. (1) the "data security punctuations" (dsps) describing the data-side security policies, and (2) the "query security punctuations" (qsps) in their paper. The advantages of such a stream-centric security model include flexibility, dynamicity and speed of enforcement. A stream processor can adapt to not only data-related but also to security-related selectivity, which helps reduce waste of resources, when few subjects have access to streaming data.

There are several applications where sensor nodes work as the source of the data stream. Here we list several applications such as real-time health monitoring applications (Health care), industrial monitoring, geo-social networking, home automation, war front monitoring, smart city monitoring, SCADA, event detection, disaster management and emergency management.

From all the above applications, we found data needs to be protected from malicious attacks to maintain originality of data before it reaches a data processing centre [25]. As the data sources is sensor nodes, it is always important to propose lightweight security solutions for data streams [25].

These applications require real-time processing of very high-volume data streams (also known as *big data stream*). The complexity of big data is defined through 5Vs i.e. *volume, variety, velocity, variability, veracity*. These features present significant

opportunities and challenges for big data stream processing. Big data stream is continuous in nature and it is important to perform the real-time analysis as the life time of the data is often very short (applications can access the data only once) [1, 2]. So, it is important to perform security verification of big data streams prior to data evaluation. Following are the important points to consider during data streams security evaluation.

- Security verification is important in data stream to avoid malicious data.
- Another important issue, security verification should perform in near real-time.
- Security verification should not degrade the performance of stream processing engine (SPE). i.e. security verification speed should synchronize with SPE.

5 Conclusion

A glimpse of the IoT may be already visible in current deployments where networks of smart sensing devices are being interconnected with a wireless medium, and IP-based standard technologies will be fundamental in providing a common and well accepted ground for the development and deployment of new IoT applications. According to the 5Vs features of big data, the current data stream heading towards the new term as big data stream where sources are the IoT smart sensing devices. Considering that security may be an enabling factor of many of IoT applications, mechanisms to secure data stream using data in flow for the IoT will be fundamental. With such aspects in mind, this paper an exhaustive analysis on the security protocols and mechanisms available to protect big data streams on IoT applications.

References

1. Puthal, D., Nepal, S., Ranjan, R., Chen, J.: A dynamic prime number based efficient security mechanism for big sensing data streams. J. Comput. Syst. Sci. **83**(1), 22–42 (2017)
2. Puthal, D., Nepal, S., Ranjan, R., Chen, J.: DLSeF: a dynamic key length based efficient real-time security verification model for big data stream. ACM Trans. Embedded Comput. Syst. **16**(2), 51 (2016)
3. Granjal, J., Monteiro, E., Sá Silva, J.: Security for the internet of things: a survey of existing protocols and open research issues. IEEE Commun. Surv. Tutor. **17**(3), 1294–1312 (2015)
4. Tien, J.: Big data: unleashing information. J. Syst. Sci. Syst. Eng. **22**(2), 127–151 (2013)
5. Boldyreva, A., Fischlin, M., Palacio, A., Warinschi, B.: A closer look at PKI: security and efficiency. In: Okamoto, T., Wang, X. (eds.) PKC 2007. LNCS, vol. 4450, pp. 458–475. Springer, Heidelberg (2007). doi:10.1007/978-3-540-71677-8_30
6. Puthal, D., Nepal, S., Ranjan, R., Chen, J.: A dynamic key length based approach for real-time security verification of big sensing data stream. In: Wang, J., Cellary, W., Wang, D., Wang, H., Chen, S.-C., Li, T., Zhang, Y. (eds.) WISE 2015. LNCS, vol. 9419, pp. 93–108. Springer, Cham (2015). doi:10.1007/978-3-319-26187-4_7
7. Puthal, D., Nepal, S., Ranjan, R., Chen, J.: DPBSV- an efficient and secure scheme for big sensing data stream. In: 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, pp. 246–253 (2015)
8. Weber, R.: Internet of things-new security and privacy challenges. Comput. Law Secur. Rev. **26**(1), 23–30 (2010)

9. Kopetz, H.: Internet of things. In: Kopetz, H. (ed.) Real-Time Systems. Real-Time Systems Series. Springer, Boston (2011). doi:10.1007/978-1-4419-8237-7_13
10. Al-Fuqaha, A., et al.: Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. **17**(4), 2347–2376 (2015)
11. Li, S., Xu, L., Zhao, S.: The internet of things: a survey. Inf. Syst. Front. **17**(2), 243–259 (2015)
12. Xu, L., He, W., Li, S.: Internet of things in industries: a survey. IEEE Trans. Industr. Inf. **10**(4), 2233–2243 (2014)
13. Ilie-Zudor, E., et al.: A survey of applications and requirements of unique identification systems and RFID techniques. Comput. Ind. **62**(3), 227–252 (2011)
14. Wang, Y., Attebury, G., Ramamurthy, B.: A survey of security issues in wireless sensor networks. IEEE Commun. Surv. Tutor. **8**(2), 2–23 (2006)
15. IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC Sublayer, IEEE Std. 802.15.4e-2012 (Amendment to IEEE Std. 802.15.4–2011), (2011), pp. 1–225 (2012)
16. Thubert, P.: Objective function zero for the routing protocol for low-power and lossy networks (RPL). RFC 6550 (2012)
17. Bormann, C., Castellani, A., Shelby, Z.: Coap: an application protocol for billions of tiny internet nodes. IEEE Internet Comput. **16**(2), 62 (2012)
18. Zheng, T., Ayadi, A., Jiang, X.: TCP over 6LoWPAN for industrial applications: an experimental study. In: 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1–4 (2011)
19. Conzon, D., Bolognesi, T., Brizzi, P., Lotito, A., Tomasi, R., Spirito, M.: The virtus middleware: an XMPP based architecture for secure IoT communications. In: 21st International Conference on Computer Communications and Networks, pp. 1–6 (2012)
20. Sicari, S., Rizzardi, A., Grieco, L., Coen-Porisini, A.: Security, privacy and trust in internet of things: the road ahead. Comput. Netw. **76**, 146–164 (2015)
21. Bandyopadhyay, S., Sengupta, M., Maiti, S., Dutta, S.: A survey of middleware for internet of things. In: Özcan, A., Zizka, J., Nagamalai, D. (eds.) CoNeCo/WiMo -2011. CCIS, vol. 162, pp. 288–296. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21937-5_27
22. Colistra, G., Pilloni, V., Atzori, L.: The problem of task allocation in the internet of things and the consensus-based approach. Comput. Netw. **73**, 98–111 (2014)
23. Fox, G., et al.: High performance data streaming in service architecture. Technical report, Indiana University and University of Illinois at Chicago (2004)
24. Nehme, R., Lim, H., Bertino, E., Rundensteiner, E.: StreamShield: a stream-centric approach towards security and privacy in data stream environments. In: ACM SIGMOD International Conference on Management of data, pp. 1027–1030 (2009)
25. Chen, P., Wang, X., Wu, Y., Su, J., Zhou, H.: POSTER: iPKI: identity-based private key infrastructure for securing BGP protocol. In: ACM CCS, pp. 1632–1634 (2015)