Secure Data Provenance in Internet of Things based Networks by Outsourcing Attribute based Signatures and using Bloom Filters

Muhammad Shoaib Siddiqui[1], Atiqur Rahman[2], Adnan Nadeem[3], Ali M. Alzahrani[4]
Faculty of Computer and Information Systems
Islamic University of Madinah
Kingdom of Saudi Arabia

***Abstract*—With the dawn of autonomous organization and network and service management, the integration of existing networks with Internet of Things (IoT) based networks is becoming a reality. With minimal human interaction, the security of IoT data moving through the network becomes prone to attacks. IoT networks require a secure provenance mechanism, which is efficient and lightweight because of the scarce computing and storage resources at the IoT nodes. In this paper, we have proposed a secure mechanism to sign and authenticate provenance messages using Ciphertext-Policy Attribute Based Encryption (CP-ABE) based signatures. The proposed technique uses Bloom filters to reduce storage requirements and an outsourced ABE mechanism to use lessen the computational requirements at the IoT devices. The proposed technique helps in reducing the storage requirements and computation time in IoT devices. The performance of the proposed mechanism is evaluated and the results show that the proposed solution is best suited for resourced constrained IoT network.**

Keywords—Data provenance; bloom filter; ciphertext policy attribute based encryption; IoT

I. Introduction

The reason behind widespread adoption of ubiquitous computing is the use of smart devices and the Internet of Things (IoT). There are now more than 20 billion smartphones and IoT devices [1]. With the existing Wi-Fi and 3G/4G connections and 5G in the near future, Internet is available as on-the-go gateway to the rest of the world, making IoT devices an important factor in most solutions. IoT applications, such as, telemedicine, healthcare, sensing and diagnostic reporting, remote doctor consultation, security, surveillance, industrial automation and monitoring, telemetry, asset tracking, etc., with IoT-based sensor networks providing remote monitoring, are making significant progress. According to Statista 2018 [1], there would be more than 75 billion IoT devices connected to the Internet by 2025 (see Fig. 1).

One of the major characteristics behind the success of IoT is the least human involvement. With the penetration of smart-devices and intelligent home-appliances, IoT has opened the doors of new prospects. However, to maintain the accuracy of this technology, reliability and integrity of data must be ensured. The data received by these devices is vulnerable to attacks. The most critical attacks in the field of monitoring and surveillance are false data injection and the corruption of data [3]. For ensuring the data integrity and sender's authentication, researchers have proposed data provenance mechanisms, which can identify a change in the data and provide a list of sender and forwarder of the data [6].

However, the data provenance mechanisms are also prone to attacks in which the attacker manipulates the provenance data to spoof its identity or hides the change in the data [4]. By attacking the provenance mechanism, the adversary can hide the loss of data integrity. Therefore, it is important to develop secure data provenance mechanism, which can provide data integrity and ensure that the mechanism itself is safe from attacks. However, the secure data provenance schemes have high cost of storage and computation for resource constrained IoT devices.

As the solution to the above-mentioned problem, we propose a secure provenance mechanism based on the attribute-based encryption [14]. We used ciphertext-policy attribute-based encryption (CP-ABE) [14] instead of Public key Cryptography [17] for designing our secure provenance mechanism. In [2], the authors have discussed how the algorithm of CP-ABE can be distributed. We used the algorithm, with some modifications, to convert our secure provenance mechanism into a distributed mechanism by outsourcing most of the computational load on fog/edge nodes. Due to the scarce resources at IoT nodes, we needed to devise a solution which required low storage for keeping the record of each data packet. For this, we have utilized the Bloom filters.

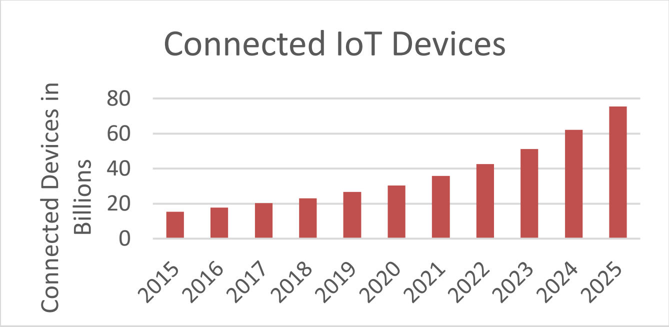

Fig. 1. The Number of IoT Devices Connected to the Internet.

The research is sponsored by the Deanship of Research, Islamic University of Madinah, Kingdom of Saudi Arabia.

The rest of the paper is articulated as follows. Section II presents the literature review by discussing the related work in the field of secure provenance. Section III presents the proposed mechanism. Section IV discusses the details of experimentation and results, while Section V concludes the paper.

II. Related Work

Data provenance provides the source of information and the historical path it has been forwarded from. Scientist is interested in the origin and historical transformation of the data as it contains critical information about the reliability of the data [5]. From it, one can determine the quality of the data based on data origin and past derivations, trace the sources of errors, automated re-creation of the derivations to update the data and provide the attribution of the data sources. Provenance is also essential for the commercial domain, where it can be used to deepen the data source in a data warehouse, trace the creation of intellectual property and provide an audit trail for regulatory purposes.

The use of a data source in a distributed system has been used to track records using a data source, in identifying the data flow from a subset of the original inputs, and in mending the data flow. To do so, one needs to keep track of the set of inputs for each operator used to derive each output. Although there are many ways for data provenance, such as, copy-provenance and how-provenance [4] [6], the information needed is a simple form of why-provenance, or lineage, as defined by Cui et al. [7].

In IoT based networks, reliability of the data received is a critical issue as a lot of attacks are made to corrupt data or inject false data in the networks. Data provenance is a mechanism to ensure that the data is forwarded by reliable nodes such that data integrity is maintained; however, the data provenance mechanisms are also prone to attacks. By attacking the provenance mechanism, the attacker can hide the loss of data integrity. Therefore, it is important to develop secure data provenance mechanism, which can ensure data integrity and is safe from attacks.

In [8], authors have argued that the one of requirement for IoT networks is to allow a user to trust the data regarding its origin and location. They provide a secure provenance mechanism based on secret key cryptography; however, they didn't mention the key distribution and revocation mechanisms.

In [3], the authors have identified the requirements and challenges of implementing a secure provenance system for IoT based networks. They claim that if data traces of IoT devices are recorded then provenance can play a vital role as it solves many issues related to data trustworthiness, decision-making, data reconciliation and data replication. The challenges of implementing a secure provenance mechanism in IoT, WSN, and RFID based networks are identified as data storage and processing, as these devices lack computational power.

Sultan et al. have presented a secure provenance scheme for wireless sensor network (WSN) in [9]. They argued that data provenance in sensor networks introduce some difficult requirements due to their scarce energy resources, and provenance mechanism's high bandwidth consumption. Therefore, it needs efficient storage and secure transport. Their scheme is based on Bloom filters, like [10] and is a light-weight solution; however, the security of the mechanism has a few question marks. Furthermore, their solution may not give the complete provenance path.

Kamal and Tariq provide a solution to provenance in IoT based network, which is also light-weight [11]. They have generated link fingerprints using the Received Signal Strength Indicator (RSSI) value of the IoT devices. The path to source node is identified at server, using the correlated coefficient based on the matching of the link fingerprints. However, they have relied on the physical parameters of RSSI for ensuring security also, without utilizing an authorization mechanism. The attacker can easily spoof the RSSI values to hide its location.

Liang et al. present a block-chain based provenance mechanism in cloud environment, in which they have proposed a framework for gathering and authenticate data provenance in the cloud, by using blockchain of provenance data [12]. It operates mainly in three phases: (1) provenance data collection, (2) provenance data storage, and (3) provenance data validation.

In our proposed schemes, we aim to develop a secure and light-weight provenance mechanism, specifically design for IoT devices, which are a part of a Machine to Machine Network. Our scheme uses Bloom filter to store the provenance information to reduce the storage requirement significantly at the resource constrained IoT devices. We have utilized the solution provided by [13] to outsource the signature generation using attribute-based encryption to lessen the load of signing and verification from the IoT nodes also. Following section discuss the proposed mechanism in detail.

III. Proposed System

A. Communication Model

The communication model is a Cloud environment with IoT devices connected to the cloud. For the users of the cloud, some fog/edge nodes would be assigned that would be closed to the IoT devices as compared to the main servers of the Cloud. These dedicated edge nodes are computationally powerful and close to the IoT devices; therefore, IoT devices can delegate their signature generation and verification responsibilities to these nodes. Fig. 2 shows the scenario of the proposed system.

B. Threat Model

The threat model for the proposed system is as follows:

- There is no physical protection for the IoT nodes. An attacker can access the IoT device and initiate a physical attack. The enemy's purpose is to access the memory of the device and compromise the data integrity.
- An attacker can mimic as an authorized IoT node. In this way attacker can insert wrong or fallacious data

into the system for compromising the accuracy of the system.

- An attacker can snoop, alter, reiterate, and infuse invalid data and messages.
- An attacker can modify data sent from an IoT node to the sink and compromise the provenance mechanism.

C. Secure Provenance based on CP-ABE

For securing the provenance mechanism, Ciphertext-Policy Attribute based Encryption (CP-ABE) is used [14] [15]. Fig. 3 shows the conceptual model of implementing the CP-ABE in IoT based Cloud environment. The distribution of CP-ABE algorithm is adopted from [13], where the authors have implemented the load sharing of cryptographical computation by outsourcing the signing and signature verification to the edge nodes.

By using the distributed CP-ABE algorithm, the partial signature is created by the Fog/Edge node, which takes up most of the computational overhead. The partial signature is sent to the IoT node which creates the complete signature.

Whenever an IoT node wants to send a message to another node, it creates the signature for the message and attaches it with the message. By using a signature, it is ensured that the message is from the authorized user and the data integrity is maintained. Using our proposed mechanism, we ensure the correctness of the data, which includes the identity of the sender, the identity of the forwarding node, and the original data of the IoT node.

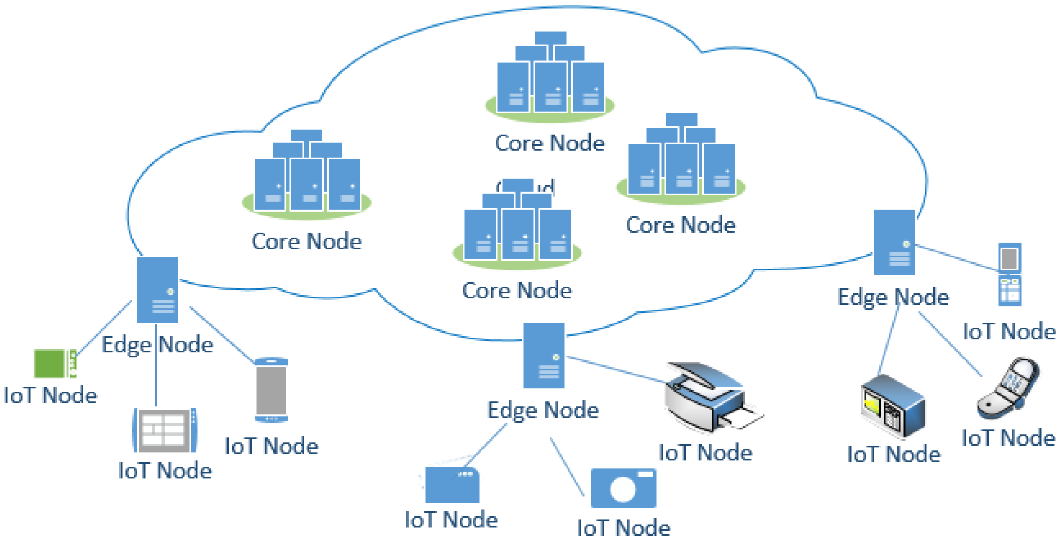

Fig. 2. A Conceptual Scenario of the Cloud based IoT Network with Edge/Fog Node Supporting the Computational Load.

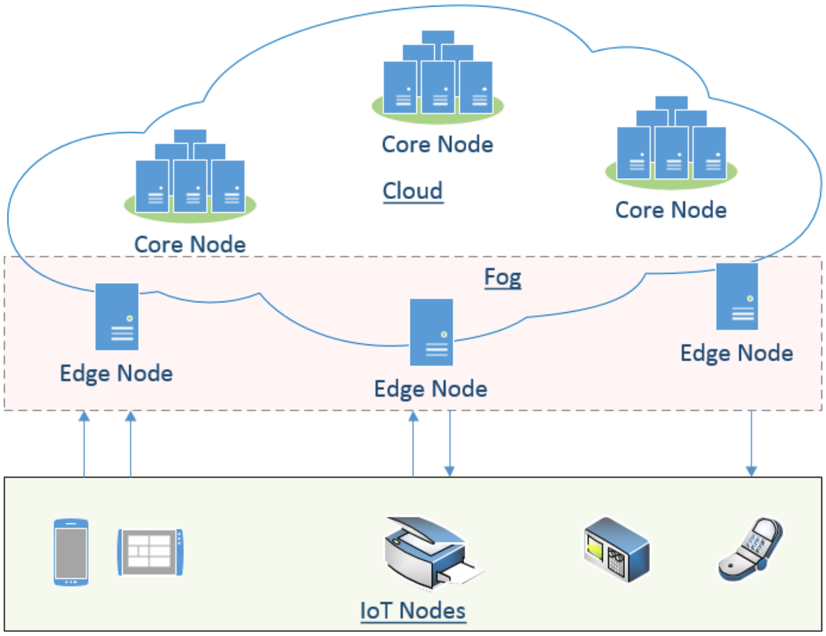

Fig. 3. Conceptual Model – IoT based Cloud/Fog Computing Scenario.

Fig. 4 shows the signing process used to send the provenance data from the IoT node. The signing process is mainly divided into five steps. These five steps are; Setup, Key Generation, Signing-outsourced and Signing. Signing-outsourced has most of the computational load that is why it is performed by the fog/edge node, while the step with significantly lower overhead, i.e. Signing is performed by the IoT node. The details of each step are as follows:

1) Setup: this step is the first step which is performed at the attribute authority, IoT node, or the edge node. The inputs to this step are the security parameter δ, a universal set of attributes μ, and an auxiliary information α. The setup process provides the master key Ψ and a public key U.

2) Key generation: This is the second step, which is to be performed by the attribute authority. The IoT node would initiate the process of signing by requesting the attribute authority for secret key for the IoT node with attribute set *A*, and the corresponding edge/fog node. The key generation phase is provided the master key *Ψ*, and attribute set *A*, for which it outputs the secret key *P* for the IoT node and *Θ* for the edge /fog node, which would use it to create the partial signature.

3) $Sign_{outsourced}$: This phase is executed by the edge/fog node, which takes input the secret key *Θ* attribute set *A*, and the predicate *Γ*. This is the main part of the algorithm in which the CP-ABE is used to create the partial signature $\sigma_{partial}$. This partial signature is sent to the IoT node, which would calculate the final signature; however, the computational overhead is bared by the edge/fog node.

4) Sign: the final signing phase that is performed by the IoT node outputs the signature for the message *M*. It uses the partial signature $\sigma_{partial}$, the secret key *P* (provided by the attribute authority), and the predicate *Γ*.

After the message is signed by the IoT node, the message *M* and the signature *σ* are forwarded to the next node. The message *M* includes the identity of the sender, the identity of the forwarding node, and the original data of the IoT node, while signature *σ* is attached to verify if there was a change in the message or if the message is forwarded by an adversary node.

To verify the integrity of the message and authenticate the sender, the verification step is performed, which is also shown in Fig. 5.

5) Verify: The verify phase takes input the message M, the signature *σ*, the public key *U*, and the predicate *Γ*. It validates if the signature is correct or invalid.

If the verification process fails then the secure provenance mechanism can identify the break in provenance chain. It would ask the sender node to resent the data. If the verification step is successful then the data about the message/packet is saved at the receiving node before that data is forwarded to the next node. In this way the messages between the two hops/nodes are ensured to be secure and verified. Before the data/message is forwarded to the next hop/node, the data is stored at the current node with provenance information for each data packet/message.

As there would be a lot of message that passes from an intermediate node, therefore, it would take up a lot of storage and if these intermediate nodes are IoT nodes then they do not have the luxury of high storage capacity. For solving the issue of storage our mechanism incorporates Bloom filters as discussed in the next subsection.

D. Reducing Storage Requirment using Bloom Filters

The provenance framework stores information of the data whenever it passes from a node. We have utilized Bloom-filters [16] to store the information of the data at each node. Bloom filter does not take a lot of storage; rather it uses an array of bits to store the data. First the secured data packet with sender's digital signature is passed through the hash functions. Three hash functions provide indexes to the 32-bit Bloom filter array, where the bits are turned on (1). Similarly, all the data packets are stored in the Bloom filter. Fig. 6 shows the storage mechanism where i=32 corresponds to the IP address of the sender.

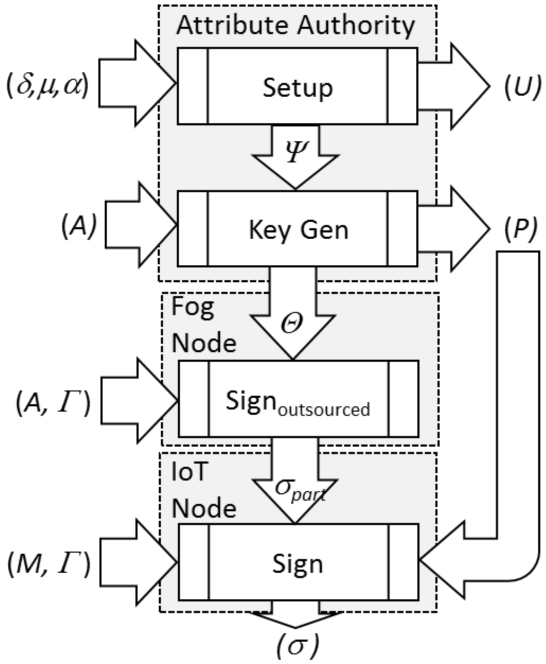

Fig. 4. The Signing Process using Outsourced Attribute based Encryption.

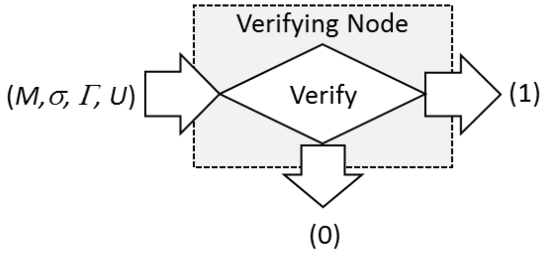

Fig. 5. The Signature Verification Process using Outsourced Attribute based Encryption.

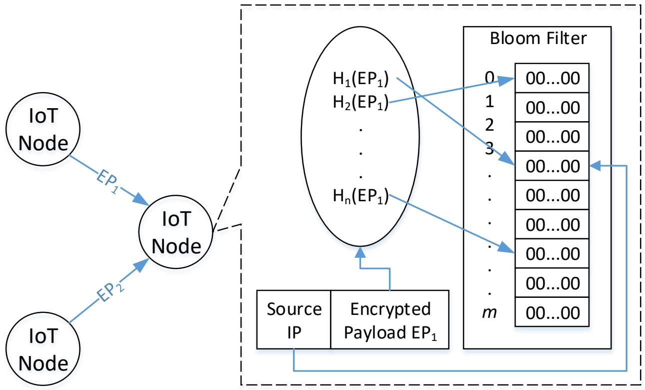

Fig. 6. The Logging of Provenance Data at the IoT Node.

Fig. 7 shows how the membership is checked for a data packet. When the provenance tracking is performed, a request for provenance checking is send to the neighboring nodes. To check if the data packet was passed through a node, the data packet is again passed through the hash functions. The hash function would provide the indexes and if any index is found to be set to (1), then we would know that the data packet was passed through the node. The 32-bit value at that index would provide the IP address of the sender of this packet. If any of the indexes is found to be 'not-set' (0) then the packet was not passed from this node. Now, the provenance tracking request is forwarded to that node which performs the similar action to identify the node that forwarded the said packet to this node. Following this scheme, the provenance path is constructed.

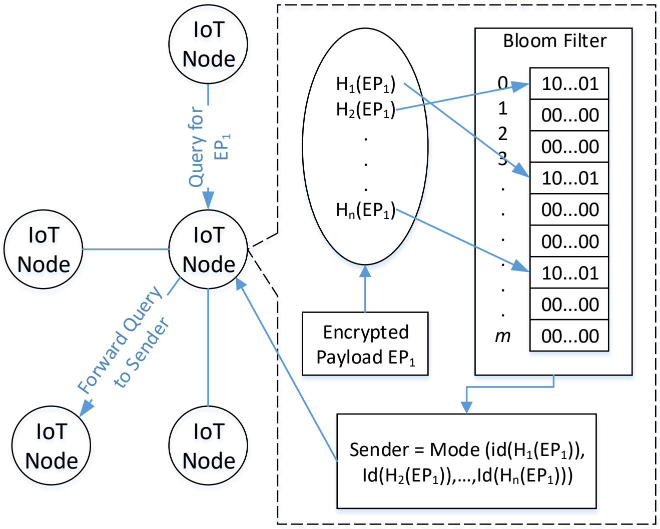

Fig. 7. The Query Mechanism for Checking the Membership of the said Data Packet.

IV. Simulation and Evaluation

The main objective of this research was to reduce the computational load from the IoT nodes and reduce the storage requirements for storing provenance information. We have performed simulation for our proposed scheme using a server as a cloud, an Nvidia Jetson tk2 device as an edge node and multiple instances of a Java application on a laptop to simulate as IoT nodes. We have generated random packets between the IoT nodes, which have utilized the service from the edge node to generate partial signature by the edge /fog nodes and the rest of the signature (with less computational load) by the IoT nodes. We have calculated the storage requirement and computational load at each IoT node to evaluate our proposed scheme.

Fig. 8 shows the storage measurements based on the number of IoT nodes and increased number of packets. We have selected the Bloom filter array with size m = 1024. With the increase in membership, the probability for false positive in the Bloom filter also increase, hence, after 1024/4 = 256 packets, the Bloom filter array is archived, and a new array is allocated to the Bloom filter. As each entry is of 32 bits, therefor, the size of each Bloom filter array is 1024 x 32 bits = 4KB.

Fig. 9 shows the computational load measurement of the proposed scheme for performing key generation and signature with the support of the edge node and without the support of edge node. This computational load also affects the battery power of an IoT device; hence, the battery drainage issue can also be resolve using outsourced signature technique.

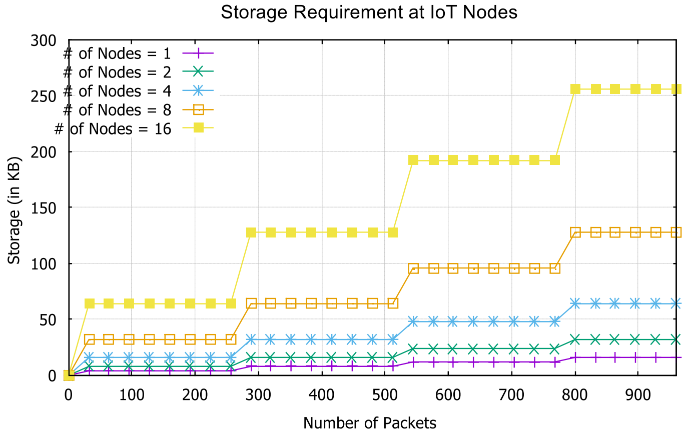

Fig. 8. Required Storage in KBs with Increased Number of Packets for different Number of Nodes.

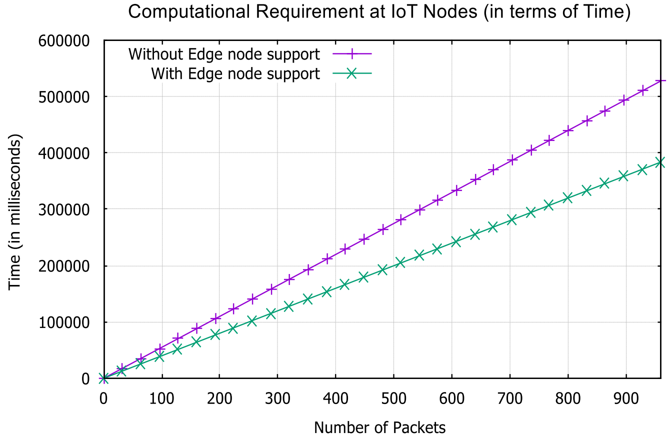

Fig. 9. Time Required to Generate Key and Signature Computation with Increased Number of Packets with and without Edge Node Calculating Partial Digital Signatures.

V. Conclusion

In this paper, we have proposed a secure provenance tracking mechanism which uses the outsourced signing technique using Attribute Based Encryption by offloading the IoT node. The IoT node only calculates the partial digital signature, while the high-overhead computations are performed by the edge node. Our scheme also reduces the storage requirement at the IoT devices by storing the provenance data in a multi-array Bloom filter. We have discussed in detail the mechanism of the storage techniques using Bloom filter, which uses hash function to save the information about the packets received and forwarded by the node. The hash-based searching has the complexity of O(1); therefore, it decreases the provenance tracking time. The results show that our proposed scheme uses less storage and takes less time in terms of computation as compared to the normal secure provenance mechanism.

In future, we aim to enhance the proposed scheme by implementing it on Graphics Processing Units (GPUs) available on mobile devices by using parallel computing, which will reduce the cost in terms of time . We are also investigating for a possible solution using Blockchain technologies, which has inherent characteristics to sim in data provenance.

Acknowledgment

This work was supported by the Deanship of Research, Islamic University of Madinah, Kingdom of Saudi Arab [Project Title: “Secure Provenance in IoT based Machine-to-machine Networks for Reliable Monitoring & Surveillance”, 56/40 ~ 1439-1440].

References

[1] Joyia, Gulraiz J., Rao M. Liaqat, Aftab Farooq, and Saad Rehman. "Internet of Medical Things (IOMT): applications, benefits and future challenges in healthcare domain." J Commun 12, no. 4 (2017): 240-7.

[2] Chen, Xiaofeng, “Secure Outsourced Attribute-Based Signatures”, IEEE transactions on parallel and distributed systems, 2014, Volume: 25 Issue: 12 Page: 3285-3294

[3] Sabah Suhail, Zuhaib Uddin Ahmad, Choong Seon Hong, "Introducing Secure Provenance in IoT: Requirements and Challenges" International Workshop on Secure Internet of Things (SIoT 2016), Sep. 26-30, 2016, Heraklion, Crete, Greece

[4] Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Data provenance: Some basic issues. In Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science, FST TCS 2000, pages 87–93, London, UK, UK, 2000. Springer-Verlag

[5] Pasquier, Thomas; Lau, Matthew K.; Trisovic, Ana; Boose, Emery R.; Couturier, Ben; Crosas, Mercè; Ellison, Aaron M.; Gibson, Valerie; Jones, Chris R.; Seltzer, Margo (5 September 2017). "If these data could talk". Scientific Data. 4: 170114. doi:10.1038/sdata.2017.114.

[6] Robert Ikeda and Jennifer Widom. Data lineage: A survey. Technical report, Stanford University, 2009.

[7] Y. Cui and J. Widom. Lineage tracing for general data warehouse transformations. VLDB Journal, 12(1), 2003.

[8] Muhammad Naveed Aman, Kee Chaing Chua, Biplab Sikdar, Secure Data Provenance for the Internet of Things, Proceedings of the 3rd ACM International Workshop on IoT Privacy, Trust, and Security, April 02-02, 2017, Abu Dhabi, United Arab Emirates

[9] S. Sultana, G. Ghinita, E. Bertino, M. Shehab, "A lightweight secure provenance scheme for wireless sensor networks", Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS, pp. 101-108, 2012.

[10] Muhammad Shoaib Siddiqui, Syed Obaid Amin, and Choong Seon Hong, "Hop-by-hop Traceback in Wireless Sensor Networks", IEEE Communication Letters, Vol. 16, No.2, pp.242-245, February 2012

[11] Mohsin Kamal and Muhammad Tariq, "Light-Weight Security and Data Provenance for Multi-Hop Internet of Things", in IEEE Access, Volume: 6, 2018.

[12] ProvChain: A Blockchain-based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability, in the Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Madrid, Spain. May 2017, Pages 468-477

[13] Asim, Muhammad, Milan Petkovic, and Tanya Ignatenko. "Attribute-based encryption with encryption and decryption outsourcing." (2014).

[14] Bethencourt, J.; Sahai, A.; Waters, B. (2007-05-01). "Ciphertext-Policy Attribute-Based Encryption". 2007 IEEE Symposium on Security and Privacy (SP '07): 321–334. doi:10.1109/SP.2007.11.

[15] M. Ambrosin, A. Anzanpour, M. Conti, T. Dargahi, S. R. Moosavi, A. Rahmani, P. Liljeberg, "On the Feasibility of Attribute-Based Encryption on Internet of Things Devices", IEEE Micro Magazine, vol. 36, no. 6, pp. 25-35, December 2016.

[16] A. Broder, M. Mitzenmacher, "Network applications of bloom filters: a survey", Internet Mathematics, vol. 1, no. 4, pp. 485-509, 2004.

[17] Stallings, William (1990-05-03). Cryptography and Network Security: Principles and Practice. Prentice Hall. p. 165. ISBN 9780138690175.