Bloom Filter Enables Data Provenance Mechanism With Blockchain In WSN

Amthulla Ayesha [#1]
Department of Computer Science and Engineering
C Byregowda Institute Of Technology
563101,Karnataka,India

Subhashini. R[#2]
Department of Computer Science and Engineering
C Byregowda Institute Of Technology Kolar-Srinivaspur Road-Kolar-SrinivaspurRoad- 563101,Karnataka,India.

ABSTRACT

In wireless sensor networks (WSNs), provenance is crucial for having data trustworthiness, which detects the communication failures. The provenance can be encoded through the node IDs along a packet path where the packets are generated, forwarded and/or aggregated. As the WSNs are resource limited networks, most of the provenance schemes are failed due to increase in provenance size. The sensor node uses too much energy to reduce provenance size. Also, it did not provide secure and persistent storage for long term.

To obtain secure storage, a blockchain based data provenance scheme (BCP) of compression free is proposed. The provenance is distributively stored on nodes with packet path. the BS can retrieve the provenance on demand through a query process. With the help of query process on demand base station can retrieve the provenance. Both the simulation and experiment results show that the scheme BCP is more energy efficient and secure than those of the known distributed data provenances.

Keywords- provenance, WSN, blockchain.

I. INTRODUCTION

Wireless sensor networks WSN are very much important ones among many critical systems. Which includes surveillance, cyber-physical systems, health and environmental monitoring, and weather forecasting. Data sources differs from tiny body-worn sensors to external sensors, the example is, video cameras, positioning devices etc. Such a variety of data sources provides the confidence of data authenticity so that only authentic information is provided to applications. Some systems require high level of confidence data to support exact decisions, for example in battlefield surveillance systems and mission critical applications. Provenance is the effective tool for evaluating trustworthiness of data, which gives the history of actions performed on it and the ownership of data. The current research shows the main aspects of provenance in systems such as SCADA systems for critical infrastructure in which the data which is not trusted may leads to wrong controlling of decisions. In a multi-hop sensor network, the base station checks for source and forwarding sensor nodes data provenance on individual data packets from its generation. To provide the trustworthiness and data quality, it is very important to record information such as provenance of each data packet, information of each and every node in the data flow path. Moreover, the constraints on data provenance make challenging such as limitations on energy and bandwidth, tight storage, and resource constraints of sensor nodes. However, many varieties of data provenance schemes have been implemented. In these schemes, data provenance is projected as a directed graph, in which each vertex shows that on the data flow path node is having a provenance record and each edge represents in between two nodes the path of data transfer. One of the major disadvantage with these schemes is that as the number of nodes increases provenance size also increases in the data flow path. This results in bandwidth and energy exhaustion for each packet proving and transmission in the sensor network. To decrease the size of provenance, some method carries lossy compression schemes that will delete critical information when compressing provenance record.In some of the topologies it does not provide accurate packet path because of inappropriate directed connections and not including the edges. Hence, it is necessary to provide the lightweight provenance solution without any overheads. Furthermore, the sensors may lead to attacks because of deployment of sensors in untrusted networks. Hence, it is crucial to give security requirements like confidentiality, integrity and freshness of provenance. The goal is to develop an effective provenance encoding and decoding mechanism which can be able to compress the provenance as much as possible when providing its security. In this project, shows a reliable, scattered and lossless provenance compression scheme, using arithmetic coding each node on a packet's data flow path encodes its provenance record. The base station will decode the provenance when the packet is received and checks for its validity. The sensor networks which is existing in large scale are deployed in various application domains from which data is collected and used in decision making for critical infrastructure. The information is collected and combined from various resources through intermediate processing nodes. When any new node is added the other nodes have to be compromised in the network. Therefore, giving high data trustworthiness is important for accurate decision-making. In evaluating the trustworthiness of data at sensor nodes data provenance is a crucial factor. Provenance management introduces many challenging requirements, such as low energy and bandwidth consumption, efficient storage and secure transmission. There are various challenging requirements like efficient storage, secure transmission, low energy and bandwidth consumption. The issues are the problem of transmitting provenance for sensor networks securely. The works of existing root kit detection is discovering kernel hooks, identify suspicious system call and execution patterns, discovering kernel in various system by using virtual machine for correct system behaviors. Sometime incorrect data is not detected in existing. In existing system there is separate transmission channels for data and provenance whereas, in proposed system only requires single channel for data and provenance. Furthermore, the security solutions for primitive provenance use cryptography and digital signatures, and to store provenance it uses append based data structures which will leads to limited cost. On the other hand, fixed size data structures that compactly represent provenance.

The data structure is fast Message Authentication Code (MAC) schemes and Bloom filters (BF). The Bloom filters make beneficial usage of bandwidth which results in lower rates in error.

II. RELATED WORK

As sensor nodes have less power resources in wireless sensor networks raises critical issues to consume the energy to route data from the source to destination in designing wireless sensor network routing protocols. [1] showed that as the trustworthiness of collected data is very much important to decision makers and the decision-making infrastructure which is highly deployed with sensor networks used in SCADA supervisory control and data acquisition and systems with monitoring battlefield. To solve this problem, we propose a systematic procedure for evaluate the trustworthiness of data items. This approach can access data provenance and evaluating the quantitative measures of trustworthiness for values in trust scores. The result of trust scores are computed with the help of cyclic framework which will represent the interdependency property, the data of trust score will influence on the created nodes which will manipulate the data in network of trust score. The data items present in the trust scores are evaluated from their *value similarity* and *provenance similarity*. C. Wang showed that [2] Data management is increasing rapidly in largescale applications take benefitted by huge storage capacity and slightly coupled resources combined by grid middleware. In this paper the taxonomy is created for characteristics of data provenance and apply it to recent research and apply it to scientific workflow procedures. The main intention of taxonomy divides the provenance system based on why to record provenance, what does it describe, how it store and reflects provenance, and ways to disassemble it. As we know the valuable and confidential data is very much important in all the fields like science, medicine, commerce and government. This valuable data is persistently stored in digital form, so to evaluate the origin of data becomes important shown in [3]. In this paper, we represent how to give strong integrity and confidentiality guarantee for data provenance information. The provenance aware system prototype that deploys tracking of provenance of data at the application layer which makes very much easy to implement. The empirical results show that, the runtime overhead of our method to recording provenance with highly confidential and integrity assurance ranges from 1% - 13% in typical real time environments. In this paper [4], implementation and testing of an in -packet bloom filters of following nodes that deploys cryptographically evaluated identifiers of links is described. The tests of two varieties of cryptographic techniques for the manipulating the link identity have been tested. In this paper [5], the wireless sensor networks have user authentication. The user authentication is very much important problem in developing secure systems. The wireless sensor networks are used in many areas like industries, hospitals, and universities. The users who uses the system should have proper user authentication that user is valid or not. The users can retrieve and send queries to any sensor nodes in the network. To avoid this problem, a dynamic strong password-based solution is proposed which is adapted into wireless sensor network. This solution requires easy operations and less computational load like one-way hash function and exclusive OR operations.

III. SYSTEM ARCHITECTURE

The System architecture of the proposed system is as shown in the figure below.

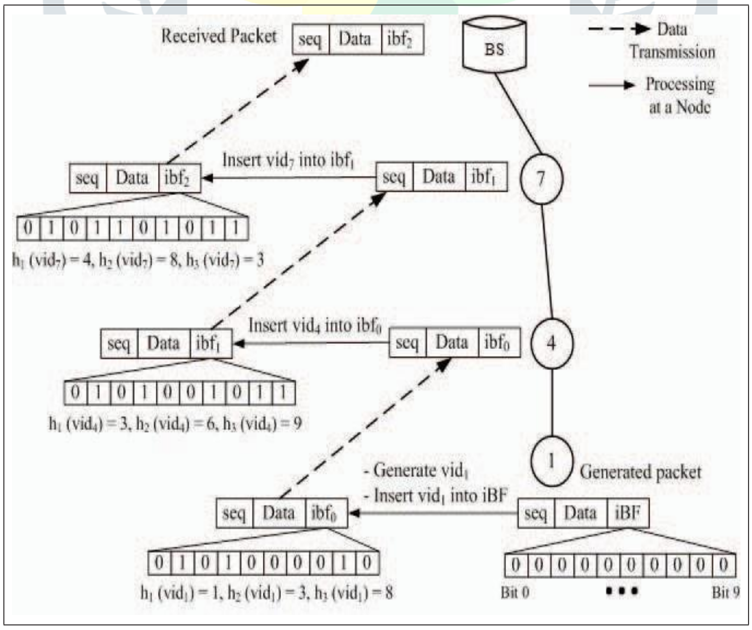

Figure 1: System Architecture

In a data packet provenance encoding maps to creating the vertices in the provenance graph and is inserted into the iBF. Each vertex originates at a node in the data path and the provenance record for the host node is represented. A vertex identified by the vertex ID (VID). The VID is created per-packet which is based on the packet sequence number and the host node has the secret key Ki. A block cipher function is used to generate VID in a secure manner. Hence for a given data packet, the vertex VID representing the node ni is measured as VIDi = generate VID(ni, seq)= EKi (seq). Where E will be a secure block cipher such as AES. When source node creates a packet, along with a packet it also generates the BF which is referred as ibf0 and it is initialized to 0. Then the source generates the vertex with respect to the equation. The VID is inserted into ibf0.

The BF is transmitted as a part of the packet. When the packet is received all the intermediate node nj performs both data and provenance aggregation. When data is sent from single child nj-1 to node nj then it combines the part of the provenance present in the packet along with the own provenance record. In this situation, the iBF ibfj−1 which belongs to the packet which is received represents a partial provenance, that is the provenance graph of the sub-path from the source.

IV. IMPLEMENTATION

The modules used for Implementation has some five modules whose methodology is followed:

1. Base station

The encryption key is generated by the base station and distributed to all the nodes in the network. All the nodes in the network will receive the key and keep it. when the data is received from node it will be performing Attack Analyzing task. The task is used to test whether the data is affected in packet drop attack or forgery attack. When the total number of packets are not matching with the received packets in that case packet drop attack will be happened. When the created ibf and received ibf is not matching then it is said to be affected by forgery attack.

Algorithm 1- Base Station Process

```
Input: Number of nodes in the Network
Output: Key Distribution
a.  Start
b.  Let N be Number of Nodes
c.  For I = 1 To N
•   Generate Unique Encryption Key (EK)
•   Generate Unique Three Hash Keys
•   Send EK, Hash key To Ith Node
d.  Next I
e.  Stop
```

Algorithm 2- Provence Forgery Verification

```
Input: Received Packet
Output: Verification Status
1.    Start
2.    BS Receive data packet from SN
3.    Extract path sequence, Encrypted data and ibf1
4.    Let N be number node in path sequence
5.    Initialise ibf2 Bits
6.    For K = 1 to N
a.    For I = 1 to 3
i.    Perform HASHing on encrypted data with Ith HASH key of Current Node
ii.   From message digest get first byte
iii.  Convert it into Decimal & get last digit value (LDV)
iv.   Replace "1" in LDV of ibf2
b.    Next I
7.    Next K
8.    If ibf1==ibf2
i.    Display packet is safe
b.    Else
i.    Display Path verification failed
9.    Stop
```

2. Network Formation:

Nodes are generated and the path is established between those nodes. Every node must consist of Node id, Group id, IP address, sending port, receiving port. Shortest path is searched by listing all its paths with corresponding energy. Consider six nodes in the network the shortest path from source to destination is identified using this module.

3. Sender Module:

Generate the sequence number for the path which is selected. Find the next hop node by adding plus one from current node. Sender will select the text file and transfer to another node. The initial bloom filter (ibf) packet is set to 0~0~0~0~0~0~0~0~0~0. For the selected path the sequence numbers are generated. The pass the file content and one AES key and three hash code, sequence number and next node id. The file content, one AES key, three hash code, sequence number and next node id is passed. The sequence number is encrypted using the key and cipher text is generated. The hash code task is performed and ibf data is replaced with '1', then new ibf is generated. The file will be converted to number of packet. The packet size is set to 16 byte. When the file content is not divisible by 16 then some characters at end of content are padded. Next 16 bytes of packets are divided. Then data pattern will be file_id##packet_id-current_packet (which is transferring currently)##sequence number##cipher text of packet##new_ibf. Then it will be transferred to next hop node.

Algorithm 2- Data Transfer Process

```
Input: Compose Message and Click Send button
Output: Data transfer to Base Station
1.  Start
2.  Source node (SN) Generate Data
3.  SN Find best Path to Base station
4.  Encrypt data using EK of SN
5.  Intialise ibf Bits
6.  For I = 1 to 3
    a.  Perform HASHing on encrypted data with Ith HASH key of Current Node
    b.  From message digest get first byte
    c.  Convert it into Decimal & get last digit value (LDV)
    d.  Replace "1" in LDV of ibf
    e.  Next I
7.  Form the data packet and transfer to next Node in Sequence
8.  Is recevied node In BS
    a.  If YES
        i.  No. Perform Step 6
    b.  ELSE
        i.  STOP
```

4. Intermediate Node Module

Intermediate nodes will receive the data and the data is split and stored in array format. With the help of intermediate node key, the sequence number is encrypted. Then ibf task is performed which is done in source module. When attacker node is attacking the current node then, it is based on the random number value so only packet drop, forgery node or both attacks need to perform attacks. The AES key is used to encrypt the sequence number. The hash code which is received helps to replace the ibf content. Then it will transfer to next hop node. Finally, the data is received by base station.

5. Attacker Module

In this module randomly selects any node in the network, attack the node and change the order of that node.

V. CONCLUSION AND FUTURE WORK

In this paper, a compression free and energy efficient provenance scheme BCP for WSNs based on the blockchain technology. Every sensor node along a packet path updates its provenance records in a simple way without consuming much energy from the WSN. With the decentralized infrastructure, provenance can be securely stored and verified. Moreover, the WSN is released from energy-consumed in-WSN provenance processing. The results show that this approach is highly beneficial to long-term trustworthiness assessments for WSNs. In BCP, the sensor nodes along a packet path need to slightly update provenance records, from which a little extra energy is consumed. In view of this, the future work is to explore a promoted scheme in which the sensor nodes just perform their routine jobs such as sensing and forwarding environmental data without consuming any extra energy at all. Provenances are encoded or retrieved only in the blockchain infrastructure that monitors the traffic in the WSN.

VI. REFERENCES

[1] I. S. T. He and W. M. Esh, “A survey on Sensor Networks,” IEEE Wireless Communications, no. February, pp. 104–111, 2010.

[2] C. Wang and E. Bertino, “Sensor Network Provenance Compression Using Dynamic Bayesian Networks,” ACM Transactions on Sensor Networks, vol. 13, no. 1, pp. 1–32, 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3027492.2997653.

[3] C. Wang, W. Zheng, and E. Bertino, “Provenance for Wireless Sensor Networks: A Survey,” Data Science and Engineering, vol. 1, no. 3, pp. 189–200, 2016. [Online]. Available: http: //link.springer.com/10.1007/s41019-016-0017-x.

[4] “Vitalik Buterin - Ethereum Blog.” [Online]. Available: https: //blog.ethereum.org/author/vitalik-buterin.

[5] Q. Zhang, X. Zhou, F. Yang, and X. Li, “Contact-based traceback in wireless sensor networks,” 2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007, pp. 2487–2490, 2007.

JETIR