Survey paper on Provenance-based encoding and decoding scheme for detecting packet drop attack in WSN

Prof. Pavan Kulkarni[1], Rajashree Sangole[2]
[1, 2]Department of Computer Engineering
[1, 2]TCOER,Computer Department & Savitribai Phule Pune University Kondhwa,Pune-48,Maharashtra,India.

Abstract- *Wireless sensor systems are deployed in various application areas, and the data they gather are utilized as a part of decision-making for critical infrastructures. A malicious adversary may present extra nodes in the network or compromise existing ones. Therefore, guaranteeing high data trustworthiness is crucial for right decision-making. Data provenance represents to a key factor in evaluating the reliability of sensor data. Provenance management for sensor networks presents a few challenging requirements, for example, low energy and data transmission utilization i.e. (bandwidth consumption), effective capacity and secure transmission.*

In this paper, we propose a scheme to securely transmit provenance for sensor data. The proposed method depends on in packet Bloom filters to encode provenance. We present productive mechanisms for provenance reconstruction and verification at the base station. In addition, we expand the protected provenance scheme with functionality to detect packet drop organized by malicious data sending nodes, and the outcomes demonstrate the adequacy and efficiency of the lightweight secure provenance scheme in detecting packet forgery attacks.

Keywords— Bloom filters, publish/subscribe, multicast, Provenance, Security, Sensor Networks.

I. INTRODUCTION

Many application domains, such as environmental monitoring, power grids and cyber physical infrastructure systems etc. Data get produced from huge number of sensor node sources and processed at intermediate hops in network on their way to a base station that helps in decision-making. Data sources are of different diversities that creates the basic need of assurity of trustworthy data, so only trustworthy information has been considered in decision process.

Data provenance is an operative technique to evaluate data trustworthiness, since it précises the history of proprietorship and the actions accomplished on the data.

Although provenance forming, gathering, and querying have been considered extensively for workflows and databases, provenance has not been properly addressed in sensor networks.

In this paper, we examine the problem of secure and effectual provenance transmission and processing for sensor networks. In a multi-hop sensor network, data provenance permits the base station to detect the source and forwarding path of an individual data packet since its generation. Provenance must be documented for each data packet, but important challenges get up due to the tight storage, bandwidth and energy constraints of the sensor nodes. Therefore, it is essential to develop a light-weight provenance solution which does not lead significant overhead. Besides, sensors often function in an untrusted environment, where they may be focus to attacks. Hence, it is essential to address security necessities such as integrity, confidentiality and freshness of provenance.

II. LITERATURE REVIEW

In packet Bloom filters encrypt scheme is used for data provenance. Data provenance characterizes a key factor in evaluating the reliability of sensor data. Provenance organization for sensor networks announces several challenging wants, such as low energy and bandwidth ingestion, efficient storage and secure transmission. Announce efficient mechanisms for provenance verification and reconstruction at the base station. In addition, extended the secure provenance scheme with mechanism to detect packet drop attacks staged by malicious data forwarding node[1] and evaluated the proposed system both analytically and empirically, and the results prove the effectiveness and efficiency of the lightweight secure provenance scheme in detecting packet forgery and loss attacks.

It specifies a systematic method for assessing the trustworthiness of data items. This methodology uses the data provenance as well as their values in calculating trust scores [2], that is, measureable measures of trustworthiness. To

obtain trust scores, proposed a cyclic framework which well reflects the inter-dependency property: the trust score of the data affects the trust score of the network nodes that formed and handled the data, and vice-verse.

In a Wireless Ad Hoc Network, nodes collaborate in assistant the network functionality [3]. The effect of malicious nodes can lead to Packet Dropping, which interrupt the infrastructures of hypothetically any node within the ad hoc networking field. Link errors can affect packet dropping, so does the insider attack, or the collective effect of link errors and malicious nodes cause packet dropping. In the most severe form, the malicious node simply stops progressing every packet received from upstream nodes, completely disorderly the path between the source and the destination.

Paper [4] suggested a mechanism in which sensor data is labeled with its provenance data spontaneously and provenance data can be improved from this tagged data. Experimentations with various scenarios proved robustness of this scheme. Special feature of this scheme is that, the provenance data is embedded into actual sensor data. Proposed system does not deliver any way to provide security to provenance data.

Paper [5] focused on provenance controlling and proposed a novel secure provenance communication scheme in which provenance is surrounded into inter packet timing domain and paper also considered limitations, requirements of WSN. Proposed scheme is different from traditional watermarking schemes. The scheme surrounds provenance data into interpacket delays and not in actual sensor data. As provenance data is not directly embedded into actual data, data quality degradation issue is solved. Provenance information is improved using optimal threshold bases mechanism to reduce the provenance retrieval errors. Proposed scheme is based on the spread spectrum watermarking technique and it is efficient against various sensor network or flow watermarking attacks. This scheme adopts that provenance data remains same for flow of the packets.

Paper [6] defined the architecture of the bloom filter data arrangement and its adeptness. Bloom filter is vector of n bits. When data is encoded into bloom filter, set of hash functions is used. Data to be encoded is hashed using hash function. Result of the hash function will be integer values. Primarily bit vector encompasses all bit value equal to 0 bit. At output, integer index is set to 1. Main purpose of bloom filter is to check the membership of element i.e. once element is encoded; membership of the data can be checked. [6] Discoursed the potential network presentations of bloom filter data structure and described appropriateness of the bloom filter for network applications.

III. PROBLEM STATEMENT

Data provenance characterizes a crucial factor in assessing the trustworthiness of sensor data. Provenance administration for sensor networks brings together several challenging requirements, such as bandwidth consumption, secure transmission efficient storage and low energy. In this project, we are recommending securely transmit provenance for sensor data by a novel lightweight scheme.

.

IV. PROPOSED SYSTEM

We are devising a provenance encoding and decoding mechanism which fulfils security and performance needs. We put forward a provenance encoding strategy whereby each node on the path of a data packet securely embeds provenance information within a Bloom filter (BF) which is transmitted along with the data. Upon receiving the packet, the BS extracts and verifies the provenance information. We also design an extension of the provenance encoding scheme which allows the BS to identify if a packet drop attack was staged by a malicious node.

We utilize only Bloom filters and fast message authentication code (MAC) schemes which are fixed-size data structures that efficiently represent provenance. Bloom filters make effective use of bandwidth, and yield low error rates in practice. We articulate the problem of secure provenance transmission in sensor networks, and identify the challenges specific to this context. We advise an in-packet Bloom filter (iBF) provenance-encoding scheme.

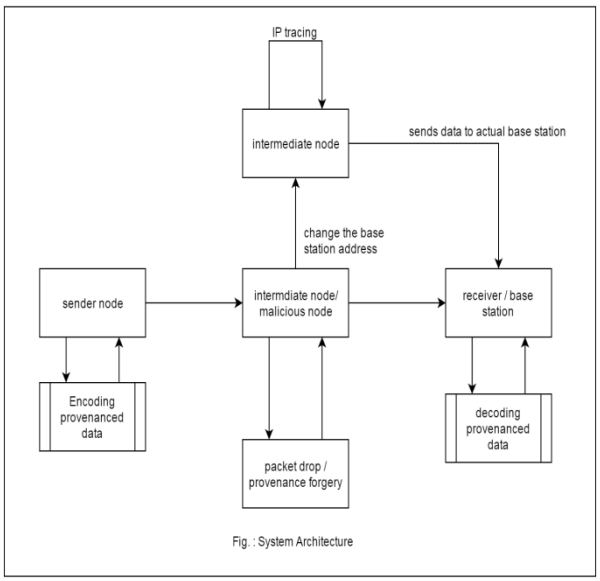

Fig. : System Architecture

1) Network Model:

We took into consideration a multihop wireless sensor network, consisting of a number of sensor nodes and a base station (BS) that collects data from the network.

The network is modelled as a graph $(N,)$, where $N = \{ni|, 1 \leq i \leq |N|\}$ is the set of nodes, and L is the set of links, containing an element for each pair of nodes ni and nj that are communicating directly with each other. Sensor nodes are static after deployment, but routing paths may alter over time, e.g. in case of node failure. Each node reports its neighbouring i.e. one hop node information to the BS after deployment. The BS allocates each node a unique identifier $nodeID$ and a symmetric cryptographic key Ki. In addition, a set of hash functions, $H = \{h1, h2,....., hk\}$ are broadcast to the nodes to use during provenance embedding.

2) Data Model:

We consider a multiple-round process of data collection. Each sensor node produces data from time to time, and individual values are routed and aggregated towards the BS using any existing hierarchical i.e. tree-based dissemination scheme. A data path of p hops is represented as $< nl, n1, n2, ..., np >$, where nl is a leaf node representing the data source, and node ni is i hops away from nl. Each non-leaf node in the path aggregates the received data and provenance with its own locally-generated data and provenance.

Every data packet contains:

i. A unique packet sequence number.
ii. A data value and
iii. Provenance.

The sequence number is attached to the packet by the data source, and all nodes use the same sequence number for a given round. The sequence number integrity is ensured through message authentication codes (MAC).

3) Provenance Model

Consider a node-level provenance, which encodes the nodes that are involved at every step of data processing. This illustration is used in previous research for trust management and for detecting selective forwarding attacks.

Given a data packet d, its provenance is modelled as a directed acyclic graph (V, E) where each vertex $v \in V$ is qualified to a specific node $HOS(v) = n$ and denotes the provenance record i.e. node ID for that node. Each vertex in the provenance graph is uniquely identified by a vertex ID (VID),produced by the host node using cryptographic hash functions. The edge set E comprises directed edges that connect sensor nodes.

- **Provenance Encoding:**

Provenance encoding for data packet refers to generating the vertices in the provenance graph and inserting them into the iBF. Each vertex originates at a node in the data path and represents the provenance record of the host node.

- **Provenance Verification:**

When the BS accepts a data packet, it implements the provenance verification process, which assumes that the BS distinguishes what the data path should be, and checks the iBF to see whether the correct path has been followed. However, right after network positioning, as well as when the topology changes (e.g., due to node failure), the path of a packet led by a source may not be known to the BS.

- **Provenance Collection:**

A provenance collection development is necessary, which retrieves provenance from the received iBF and thus the BS acquires the data path from a source node. Subsequently, upon getting a packet, it is sufficient for the BS to validate its information of provenance with that encoded in the packet

V. ADVANTAGES

1. Our design is effective approach for provenance decoding and verification at the base station.
2. We extend the secure provenance encoding scheme and devise a mechanism that identifies packet drop attacks done by malicious forwarding sensor nodes.
3. We implement a thorough security analysis and performance evaluation of the proposed provenance encoding scheme and packet loss detection mechanism.
4. We need only a single channel for both transmission channels for data and provenance.

VI. CONCLUSION

We addressed issues of securely transmitting provenance for sensor networks, and suggested a light-weight provenance encoding and decoding scheme based on Bloom filters. The scheme ensures integrity, confidentiality, reliability and freshness of provenance. We extended the scheme to incorporate data-provenance binding, and contain packet sequence information that supports discovery of packet

loss attacks. Experimental and analytical evaluation outcomes show that the proposed scheme is effective, light-weight and scalable. In future work, we plan to implement a real system prototype of our secure provenance scheme, and to optimize the accuracy of packet loss detection, especially in the case of multiple consecutive malicious sensor nodes.

ACKNOWLEDGMENT

I consider myself most fortunate and to have worked under guidance of Prof. Pavan Kulkarni Faculty and Computing Department. I would like to thank to all other teachers and friends who are really helping me to make project successfully.

REFERENCES

[1] Salmin Sultana, Gabriel Ghinita, Member, IEEE , Elisa Bertino, Fellow, IEEE , and Mohamed Shehab, Member, IEEE Computer Society, "A Lightweight Secure Scheme for Detecting Provenance Forgery and Packet Drop Attacks in Wireless Sensor Networks", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 12, NO. 3, MAY/JUNE 2015.

[2] H. Lim, Y. Moon, and E. Bertino, "Provenance-based trustworthiness assessment in sensor networks," in Proc. of Data Management for Sensor Networks, 2010, pp. 2–7

[3] Tao Shu and Marwan Krunz, "Privacy-Preserving and Truthful Detection of Packet Dropping Attacks in Wireless Ad Hoc Networks" IEEE Transactions on Mobile Computing, Volume: 14 Issue: 4

[4] Sankardas Roy, Mauro Conti, Sanjeev Setia, and Sushil Jajodia," Secure Data Aggregation in Wireless Sensor Networks" 2013

[5] Foster, J. Vockler, M. Wilde, and Y. Zhao, "Chimera: A virtual data system for representing, querying, and automating data derivation," in Proc. of the Conf. on Scientific and Statistical Database Management, 2002, pp. 37–46.

[6] K. Muniswamy-Reddy, D. Holland, U. Braun, and M. Seltzer, "Provenance-aware storage systems," in Proc. of the USENIX Annual Technical Conf., 2006, pp. 4–4.

[7] Y. Simmhan, B. Plale, and D. Gannon, "A survey of data provenance in e-science," SIGMOD Record, vol. 34, pp. 31–36, 2005.

[8] R. Hasan, R. Sion, and M. Winslett, "The case of the fake picasso: Preventing history forgery with secure provenance," in Proc. Of FAST, 2009, pp. 1–14.

[9] S. Madden, J. Franklin, J. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for ad-hoc sensor networks," SIGOPS Operating Systems Review, no. SI, Dec. 2002.