Volume 5, Issue 10, October-2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Filtering Schemes for Detecting False Data Injection Attacks in Wireless Sensor Networks: A Survey

Sanjana R. Heralgi*
M. E. Computer Engineering,
D. Y. Patil COE, Akurdi, Pune,
Maharashtra, India

Deepali Gothawal
Department of Computer Engineering
D. Y. Patil COE, Akurdi, Pune,
Maharashtra, India

Abstract— The sensor nodes operate in unattended environment. Individual sensor nodes are subjected to the security compromises. Due to the cost-constraints they lack in tamper-resistant hardware, thus increasing the chances of being attacked. Adversary can compromise sensor nodes and inject bogus data into the network that threaten system's security and also consumes network resources. To enhance the security of the system, many filtering mechanisms have been designed. The filtering schemes have been developed in the light of either symmetric or asymmetric cryptography. These techniques depends on statically configured routes and node localization, thus does not achieve high resilience to the node compromise. Hence, Polynomial-based en-route filtering scheme have been proposed that does not rely on node localization and static routes and achieve high resilience to the compromised nodes. This paper focuses on en-route filtering schemes and polynomial based en-route filtering scheme.

Keywords— False Data Injection Attack, En-route Filtering Schemes, Polynomial-based En-route Filtering Schemes, Wireless Sensor networks, Cyber-Physical network systems.

I. INTRODUCTION

With the rapid advancement in the electromechanical systems and wireless network technologies, the wireless sensor networks have gained increasing attention owing it's wide range of applications from battle field surveillance to civilian applications [1]. For example, home automation systems, military and national defence, medical care, environmental monitoring, wildlife tracking, weather checking applications, traffic management and in many other areas [2]. The wireless sensor network comprises of sensor nodes which are deployed densely throughout vast province to monitor the events such as wildfire, vehicles of enemy in the battle field, etc. The sensor nodes are not provided with the tamper-resistant hardware due to the cost constraints, thereby increasing the chances of being compromised by the adversary through many types of security attacks, such as selective forwarding, wormholes, and sybil attacks [6].

The false data injection attack is very prominent and serious menace to the sensor networks and is said as *insider attack*. In this type of attack, the adversary hi-jack subset of sensor nodes and send altered or false data [3] to the sink or inject their own sensor nodes into the network and induce the network to accept them as legitimate nodes [4]. Detection of false data injected is very essential, otherwise it may lead to false alarm at the data collection unit making it to take wrong decisions and thus devastating the limited energy of the sensor nodes. Therefore, the false data must be detected and filtered before it reaches to the sink, this largely reduces the energy consumption of the sensor nodes and wastage of network resources.

For example, an attacker could make a fire event or give false fire location information to the data collection unit, and then high resources will be used to save workers by sending them to wrong location. Therefore, it is extremely important to filter the false data as precisely as possible in sensor networks. If all the false data are overflowing towards the sink synchronously, then enormous energy will be devastated at the en-route nodes and also heavy verification burden will undoubtedly fall on the sink. As a result, the whole network could be paralysed quickly [5]. Therefore, false data must be filtered as early as possible to mitigate the energy waste [5]. Likewise false data injection attack, compromised nodes can also plant many other different attacks. It can stop the generation of reports for true events, block valid reports from reaching the destination node or record and replay the old reports.

Many filtering mechanisms to handle the false data injection attacks have been developed. Many of them are based on symmetric cryptography, i.e., once a node is attacked it is difficult to identify it. The attacked node can misuse its keys and create false data and thus the performance of the filtering scheme degrades. In this paper, some of the en-route filtering mechanisms have been discussed.

II. RELATED WORK

Recently, many filtering mechanisms [7], [8], [9], [10], [11] have been developed to tackle the false data injection attacks in the sensor networks. Many of them are based on symmetry key distribution mechanism, in which once the node is compromised it is difficult to identify the node. Many of the filtering mechanisms are based on static routes and node localization. For example SEF [7] and IHA [8] are the en-route filtering schemes were the first two proposed mechanisms, but they have threshold limitation. The LBRS [10] and LEDS [9] are the en-route filtering mechanisms that

overcome threashold limitation by using node localization and association and generating cell-based generation of reports and using techniques of location aware key generation based on statically configured routes or conforming to beam model [10].

Some other filtering mechanisms GRSEF [11] achieve low resilience to the compromised nodes. CCEF [13] scheme introduced the concept of cipher to filter bogus information at the en-route nodes instead of using symmetric key. EAB [14] used authentication bit map instead of MAC for report verification. Lin et al. [15] examined the efficiency of all the recently developed en-route filtering schemes and concluded that all these schemes have either threshold limitation or depend on node association and localization, static routes. Some of the filtering schemes are discussed briefly below.

A. Statistical En-route Filtering (SEF)

In SEF [7], large numbers of sensor nodes are deployed densely. Each of the sensing nodes generate a keyed MAC (Message Authentication Code) [16]. It is assumed that multiple sensor nodes can detect the sane event, so each detected event is assigned with multiple MACs. SEF limits the quantity of the confidential information to be assigned to each node to protect the network break down through a single compromised node and relying on the mutual decision made by the neighbouring nodes that have detected the same event.

Each of the sensing node recite the sensed signal density and one of the sensor node is chosen as the Centre-of-stimulus (CoS) node [7]. The CoS then compose and encapsulate the detecting event and make a combined report. The report is forwarded to the sink node by passing through large number of hops. Each sensing node in the en-route verifies the correctness of the report by checking the MACs with pre-defined probability and rejects the report with invalid MAC. With the increasing number of hops, the probability of detecting invalid MACs also increases. Some of the reports may escape from the en-route filtering and reach the destination. The destination node (sink) further verifies the correctness of the MAC and drops the false reports.

SEF has the following goals:

(i) Early Detection of false data reports: Detecting and dropping the false reports early can save the energy and bandwidth of the network.

(ii) Low computation and communication overhead: Taking the resource constraints of the sensor nodes into consideration, the asymmetric cryptography has been eliminated and only the efficient one-way hash function is used.

SEF limits the quantity of confidential information to be assigned to each node. The more the security information provided to the node the more successful transit sifting can be. For the authentication of the event, certain amount of confidential information must be provided to node, the adversary can access the security information by compromising only a single node. Thus to overcome this problem, SEF divides its global key pool into partitions and each partition is assigned certain number of keys.

B. Interleaved Hop-by-Hop Authentication (IHA)

The IHA [8] scheme focus on detecting and filtering the false data at en-route nodes or at the base station or sink. This scheme is used in large-scale wireless sensor network, where the sensed data need to be sent through several hops before reaching to the base station. In IHA scheme, the sensor nodes are systematized into clusters with a unique cluster ID. Each cluster comprises of t+1 sensor nodes, where is the predefined threshold and one of the node is selected as the cluster head. This scheme assures that the base station detects the false data injected by the adversary, if there are less than t compromised nodes. Given a threshold t, and upper bound B is provided, for number of hops the report to be forwarded before it is detected and rejected.

The network is bidirectional, that is, if node a can communicate with node b, then node b can also communicate with node a. And every node in the network must share its master key with the base station. Each node cognise its one-hop neighbouring node and has to establish pair-wise key. Each node has two association nodes, one is lower association node and other upper association node. The report is forwarded ahead only if it is verified by the lower association node.

This scheme has the following properties:

(i) If a false data has been interpolated by the compromised nodes, then the base station is capable enough to detect them.

(ii) The number of hops that a false data is to be detected and dropped must be smaller.

(iii) Then mechanism is efficient in computation and communication concerning the security.

(iv) The scheme should be robust towards the node failure.

The IHA requires the fixed route to transmit messages between the cluster head and the base station. Thus, making it unsuitable for the network topologies that frequently changes. Moreover, if the association fails, then security cannot be guaranteed.

C. Location-based Resilient Security (LBRS)

The LBRS [10] scheme overcomes the problem of threshold that were encountered in previous schemes. This is location based mechanism in which the secret keys are confined to the geographic location and few keys are stored within its own location. This location-binding technique limits the scope of the key misuse. In this scheme, the terrain is disunited into square cells and each of these cells is provided with some cell keys based on its location. Each node in the

cell contains two types of keys, one to authenticate the report within the cell and other keys are randomly chosen for remote cells. This scheme uses two technique: one is location-binding keys and other is location-based key assignment. It has high resilience to the number of compromised nodes for the following reasons:

(i) It averts the attacker from arbitrarily misusing a compromised key, because the keys are limited to its geographic location and can be used only within its location.
(ii) Limits damage if the adversary compromises multiple nodes and accumulate the keys, because group of keys are assigned to a different location and cannot be used together.
(iii) At last, limits the number of keys to be stored into the node as each node is assigned only a few keys depending upon its location.

This scheme has a severe drawback, it is assumed that individual node can determine its location and can generate location-based keys in a short time slot. But this task cannot be finished in such a short time slot. And the process of localization itself is susceptible to various attacks. This scheme does not provide end-to-end data security.

D. Location-aware End-to-End Security (LEDS)

Previous discussed filtering schemes provide hop-by-hop security and does not ensure end-to-end data security. To overcome this problem Ren et al. proposed LEDS [9] scheme that provides end-to-end data security and taking the advantage of broadcast nature of the sensor network, LEDS also provides one-to-many data delivery approach. This scheme exploits static and location-aware security paradigm in sensor network. It consists of two techniques: a location-aware key management framework, and an end-to-end data security mechanism.

1. Location-aware key management framework: Management of keys in LEDS depends on the static and location-aware nature of sensor nodes. A robust and efficient location-aware key management is achieved by lodging the location information into the keys. In LEDS, each node has three types of keys: (1) two unique secret keys, shared between the sensor nodes and the sink. (2) a cell key, shared among the sensor nodes within the cell. (3) a set of authentication keys, shared with report-auth cells, that provide authentication between cells and filter false data. All these keys are computed by each individual node independently and locally.
2. End-to-End data security mechanism: It ensures data confidentiality by encrypting the report with the cell key of the corresponding cell. Since, the cell key is shared among all the sensor nodes within that cell and with the sink, data confidentiality is ensured until no node in the cell is compromised. It also ensured data availability by providing strong shield against report demote and selective forwarding attacks.

LEDS has following features:

(i) The targeted area is divided virtually into number of cells using virtual geographic grid. Then the location information is provided with each key owned by a node.
(ii) Guarantees end-to-end data security, by encrypting unique key with each report. Furthermore, the authenticity of data can be individually verified by the sink.
(iii) Have efficient en-route filtering mechanism to filter the false data injected by compromised nodes.
(iv) Finally, it assures high level data availability by counteracting both report disruption [10] and selective forwarding attacks [17] simultaneously.

Like LBRS, LEDS also assumes that sensor nodes can determine their location and generate their own location-based keys in a short time slot, which is not possible. LEDS addresses selective forwarding attacks by allowing all the nodes within the cell to forward the reports, which results in communication overhead. This scheme also has t-threshold limitation, if the number of compromised nodes are more than t, then security cannot be guaranteed.

E. Group-based Resilient Statistical En-route Filtering Scheme (GRSEF)

Yu et al. proposed GRSEF [11] scheme that tries to overcome the problems of SEF mechanism. In GRSEF, the nodes are divided into t-groups after the sensor nodes are deployed. Thus, increase the probability of covering any position by different groups. Before deployment, every node is provided with a global master key, which is used to compute the group master key. The group master key is then integrated with the multiple axes-based methods [11] to compute authentication keys.

When an event occurs, the nodes detect the event and generate MAC using authentication keys. It then sends the group number and encrypted MAC to the cluster head. The cluster head is responsible for collecting all the group numbers and MACs from the neighbouring nodes and then make a combined report and send it to base station. Each node in the path that receives the report uses the event location to verify the MAC by deriving the partition-binding keys. If a node in the forwarding path receives invalid MAC or the number of MACs is less than the predefined threshold T, then the report is discarded. When the report is received by the base station, it derives all the keys by global master key and verifies the MAC in the report.

However, GRSEF has some drawbacks; it does not provide resilience to selective forwarding attack and report disruption. It has threshold limitations. It is not suitable for the topologies that changes frequently. Computing and maintaining multiple axes-based keys increases communication and cost overhead.

III. DISCUSSIONS

All the techniques that are discussed in the literature survey have drawbacks of T-threshold limitation and some depend on static routes and node localization, which are not suitable for Cyber Physical Network Systems (CPNS) [18].

The CPNS have achieved renewed attention because of the progression of the wireless sensor network technology and advancement in cyber physical systems. The CPNS consists of sensor nodes, controller, actuators and wireless networks, that is widely been used to monitor the local and remote physical entities [19], [20]. Application areas of CPNS include, vehicular networks, transportation networks, network of unmanned vehicles, etc. [12].

In CPNS, the sensor nodes monitor the physical component and obtain measurement, process those measurements and then the measured data is forwarded to the controller. The controller after receiving the measurement reckons the state of the system and transmits the commands to the actuator to handle the system's operation. The wireless sensor networks are used to estimate the state of the CPNS. The sensor nodes in the CPNS lack in tamper-resistant hardware and thus are prone to various attacks like node impersonating attacks, node replication attacks, etc. And false data injection attack is one of the attacks that thwart the security of the system. Hence, the false data must be detected and dropped before it reaches the controller and making the controller to estimate wrong physical state of the system. The schemes that are discussed in literature survey have some limitations and thus are not suitable for CPNS. Therefore, yang et al. proposed Polynomial-based Compromise Resilient En-route Filtering (PCREF) [12] scheme that uses polynomial for the report authentication instead of using MACs that ensure high resilience to the number of compromised nodes without depending on the static data routed and node localization and achieve high filtering probability of false injected data.

PCREF scheme consists of two types of sensor nodes, one is sensing node and other is forwarding node. The sensing nodes are responsible for sensing the measurement, endorse and forward the report of sensed measurement along the path. Forwarding node is used to just forward the received reports towards the controller. The typical scenario of PCREF is shown in the figure 3.1.

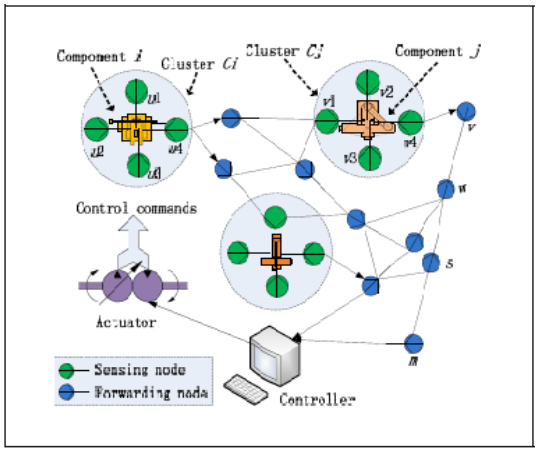

Figure 3.1: Simple scenario of PCREF [12]

Each node in the network stores two types of polynomials; Authentication polynomial and Check polynomial. Sensing node contains the authentication polynomial of local cluster and check polynomial of remote cluster with some predefined probability. The forwarding node stores only the check polynomial with same probability of each cluster. The polynomials are tied with the node ID and the polynomials are derived by the primitive polynomial pool. This scheme is independent of the static route because, the statistical pre-assignment is used to share the authentication data instead of using node association. Besides, the PCREF uses cluster-based polynomial assignment to compute the authentication and check polynomial, thus each cluster is assigned different polynomials. So that, if some nodes are compromised in a cluster, those cannot affect the security of other cluster, thus it limits the effect of attacked nodes within a particular cluster. PCREF scheme has two key components: (i) authentication information assignment, that is used to assign the keys, authentication and check polynomials and sensing nodes local ID, and (ii) data security management, that is used for the detection and filtering of false measurement reports. These components are discussed briefly in the following sub-sections.

A. Authentication Information Management

In this phase, a master key and a global polynomial pool need to be prepared before the sensor nodes are deployed. The master key is generated and stored in the sensor node before deployment, which is further used for computing the cluster key for each cluster in the network. The polynomial pool consists of various ternary polynomials that are created randomly before sensor node deployment. There is a hash function *H (.),* whose domain is encrypted measurement by sensing nodes and range is the set of positive integers. This phase has four steps.

1. Cluster Organisation: Sensing nodes are organised into clusters and they monitor the physical component. The sensing nodes can be deployed in a place close to the component. A node communicates with each other and stores the node Ids of only other n-1 nodes within a cluster. A node cannot store the node ID of other cluster even if it is only one hop away.
2. Authentication information assignment. In this phase, the network and all other nodes are initialized with the parameter $\{K_C, f(x, y, z), T, H(.)\}$, where K_C is master key, $f(x, y, z)$ is the set of polynomial with parameters x representing all the sensing node Ids, y representing all forwarding node Ids, and z representing all the measurement reports. T is the threshold and $H(.)$ is the hash function. Then using these parameters the designer computes the authentication polynomial and check polynomial.

3. Key generation: In this phase, the sensing uses the master key and generates the cluster key. The master key and the cluster head ID is concatenated and a new string is generated that is used as the cluster key. The master key is erased after the network deployment.
4. Local ID assignment: Cluster head assigns local ID to each sensing node. This phase is carried out after the sensor network deployment.

B. Data Security Management

This phase is concerned with the detection and filtering of the false measurement reports. This is carried out by following four steps.

1. Sensing report generation: The sensing node monitors the physical component, measures the data and creates a measurement report r. This report consists of the encrypted measured data, local ID, node ID and Message Authentication Polynomial (MAP). For the same measurement report, sensing node generates various MAPs using node ID and authentication polynomial. Once the sensing node generates the report, it is then forwarded to the cluster head.
2. Measurement report generation and transmission: Once the cluster head receives all the measurement reports r generated by the sensing node, it then randomly selects T reports among them and combines the selected reports to an integrated measurement report R and then the cluster heads forward this report R to the controller along the route.
3. En-route filtering: The en-route filtering is done on the false reports injected by the compromised nodes. The measurement report is transmitted hop-by-hop to the controller. The intermediate node that is having the check polynomial verifies the correctness of the received measurement report.
4. Controller authentication: After the measurement report is received by the controller, the controller authenticates the report in the same way as the intermediate node does. As the controller has all the polynomials, cluster keys and master key, it can verify all the measurement reports and filter out the false reports.

IV. CONCLUSIONS

In order to detect the false data injection attacks many filtering schemes have been proposed. But most of them either have T-threshold limitation or rely on static routes and node localization that are not suitable for CPNS. Hence, PCREF scheme has been proposed by yang et al. which can filter false data effectively and have resilience to the compromised nodes without depending on static routes and node localization. PCREF uses polynomials instead of MACs for verifying the reports and uses cluster based polynomial assignment. Yang et al. concluded that this scheme has better filtering capability and high resilience to the compromised nodes as compared with existing schemes.

ACKNOWLEDGMENT

The authors would like to thank the publishers, researchers for making their resources available and teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1] Wang, Ding, and Ping Wang. "On the anonymity of two-factor authentication schemes for wireless sensor networks: Attacks, principle and solutions." Computer Networks 73 (2014): 41-57.

[2] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Compuer .Networks. 52 (12) (2008) 2292–2330,

[3] Mo, Yilin, et al. "False data injection attacks against state estimation in wireless sensor networks." Decision and Control (CDC), 2010 49th IEEE Conference on. IEEE, 2010.

[4] H.Chan and A. Perrig, "Security and Privacy in Sensor Networks," *IEEE Computer*, October 2003.

[5] Lu, Rongxing, et al. "BECAN: a bandwidth-efficient cooperative authentication scheme for filtering injected false data in wireless sensor networks." Parallel and Distributed Systems, IEEE Transactions on 23.1 (2012): 32-43.

[6] V.C. Giruka, M. Singhal, J. Royalty, and S. Varanasi, "Security in Wireless Sensor Networks," Wireless Comm. and Mobile Computing, vol. 8, no. 1, pp. 1-24, Jan. 2008.

[7] Ye, Fan, et al. "Statistical en-route filtering of injected false data in sensor networks." *Selected Areas in Communications, IEEE Journal on* 23.4 (2005): 839-850.

[8] Zhu, Sencun, et al. "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks." *Security and privacy, 2004. Proceedings. 2004 IEEE symposium on*. IEEE, 2004.

[9] Ren, Kui, Wenjing Lou, and Yanchao Zhang. "LEDS: Providing location-aware end-to-end data security in wireless sensor networks." *Mobile Computing, IEEE Transactions on* 7.5 (2008): 585-598.

[10] Yang, Hao, et al. "Toward resilient security in wireless sensor networks." Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing. ACM, 2005.

[11] Yu, Lei, and Jianzhong Li. "Grouping-based resilient statistical en-route filtering for sensor networks." *INFOCOM 2009, IEEE*. IEEE, 2009.

[12] Yang, Xinyu, et al. "A novel en-route filtering scheme against false data injection attacks in cyber-physical networked systems." *Computers, IEEE Transactions on* 64.1 (2015): 4-18.

[13] Yang, Hao, and Songwu Lu. "Commutative cipher based en-route filtering in wireless sensor networks." *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*. Vol. 2. IEEE, 2004.

[14] Chen, Yu-Shian, and Chin-Laung Lei. "Filtering false messages en-route in wireless multi-hop networks." *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*. IEEE, 2010.

[15] Lin, Jie, et al. "Towards effective en-route filtering against injected false data in wireless sensor networks." *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*. IEEE, 2011.

[16] Deng, Jing, Richard Han, and Shivakant Mishra. *Enhancing base station security in wireless sensor networks*. Technical Report CU-CS-951-03, Department of Computer Science, University of Colorado, 2003.

[17] C. Karlof and D. Wagner, .Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures., Ad Hoc Networks, 1(2), 2003.

[18] Wu, Fang-Jing, Yu-Fen Kao, and Yu-Chee Tseng. "From wireless sensor networks towards cyber physical systems." *Pervasive and Mobile Computing* 7.4 (2011): 397-413.

[19] Cardenas, Alvaro A., Saurabh Amin, and Shankar Sastry. "Secure control: Towards survivable cyber-physical systems." *The 28th International Conference on Distributed Computing Systems Workshops*. IEEE, 2008.

[20] Pajic, Miroslav, Alexander Chernoguzov, and Rahul Mangharam. "Robust architectures for embedded wireless network control and actuation." *ACM Transactions on Embedded Computing Systems (TECS)* 11.4 (2012): 82.

