Secure Data Aggregation in Wireless Sensor Networks: Enumeration Attack and Countermeasure

Aishah Aseeri
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716
aaseeri@udel.edu

Rui Zhang
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716
ruizhang@udel.edu

Abstract—**Data aggregation is a key primitive in wireless sensor networks and refers to the process in which the sensed data are processed and aggregated en-route by intermediate sensor nodes. Since sensor nodes are commonly resource constrained, they may be compromised by attackers and instructed to launch various attacks. Despite the rich literature on secure data aggregation, most of the prior work focuses on detecting intermediate nodes from modifying partial aggregation results with two security challenges remaining. First, a compromised sensor node can report arbitrary reading of its own, which is fundamentally difficult to detect but widely considered to have limited impact on the final aggregation result. Second, a compromised sensor node can repeatedly attack the aggregation process to prevent the base station from receiving correct aggregation results, leading to a special form of Denial-of-Service attack. VMAT [1] (published in ICDCS 2011) is a representative secure data aggregation scheme with the capability of pinpointing and revoking compromised sensor nodes, which relies on a secure MIN aggregation scheme and converts other additive aggregation functions such as SUM and COUNT to MIN aggregations. In this paper, we introduce a novel enumeration attack against VMAT to highlight the security vulnerability of a sensor node reporting an arbitrary reading of its own. The enumeration attack allows a single compromised sensor node to significantly inflate the final aggregation result without being detected. As a countermeasure, we also introduce an effective defense against the enumeration attack. Theoretical analysis and simulation studies confirm the severe impact of the enumeration attack and the effectiveness of the countermeasure.**

I. Introduction

Wireless sensor networks play a key role in the emerging IoT paradigm where millions of sensors are expected to be deployed throughout the physical space, which continuously sense the surrounding environment and generate an unprecedented amount of data. A typical wireless sensor network is a multi-hop wireless network formed by many resource-constrained sensor nodes and a base station, where sensed data are forwarded to the base station with Internet connectivity via intermediate sensor nodes. Exemplary applications of wireless sensor networks include manufacture plant monitoring, asset tracking, traffic monitoring, environmental monitoring, public safety, and so on [2].

In-network data aggregation [3], [4] a key functionality in wireless sensor networks and refers to the process in which the sensed data are processed and aggregated en-route by intermediate sensor nodes. Since sensor nodes are commonly battery powered with limited communication and computation resources, forwarding every sensor reading to the base station would quickly deplete the energy of intermediate nodes. In-network data aggregation allows the base station to learn statistic aggregates of the sensed data while greatly reducing the energy consumption and prolonging the network's lifetime. Consider the SUM aggregation as an example. Sensor nodes first form an aggregation tree rooted at the base station. During the aggregation process, every node sums up the readings from its children and its own and forwards the partial sum to its parent. The base station is able to obtain the sum of all readings at the end of the process. Other common aggregate functions such as MAX/MIN, COUNT, and AVERAGE can be realized in a similar fashion.

As an important network primitive, in-network data aggregation faces several critical security challenges. Since sensor nodes are resource-constrained, they may be physically captured or compromised by attackers and instructed to launch various attacks. For example, a compromised sensor node may modify its partial aggregation result to significantly inflate or deflate the final aggregation result at the base station. Second, even if the base station is able to detect and reject the false aggregation result, a compromised sensor node can launch persistent attacks to prevent the base station from receiving the correct aggregation result, leading to a special form of Denial-of-service attack. Last but not the least, a compromised sensor node may report an arbitrary reading of its own while following the aggregation protocol.

Secure data aggregation in wireless sensor networks has been studied extensively in the past. A common assumption held in the literature is that a single compromised sensor node forging its own reading is fundamentally difficult to detect but has limited impact on the final aggregation result for robust aggregation functions like SUM and COUNT [5]. Most of the research efforts have focused on detecting intermediate nodes manipulating partial aggregation results. Existing solutions can be broadly classified into two categories. The first category such as [6]–[9] can provide accurate aggregation results and detect malicious sensor nodes manipulating partial aggregation results via commitment verification. The second category such as [1], [10]–[14] offers statistical estimations of the aggregate results via probabilistic sampling. As mentioned above, a single malicious sensor node can keep attacking the aggregation process to prevent the base station from obtaining the correct aggregate. There are a very few attempts addressing the identification and revocation of compromised nodes with VMAT [1] being a representative. VMAT relies on verifiable MIN aggregation and converts other additive aggregation functions such as SUM and COUNT into MIN aggregation via verifiable sampling.

In this paper, we introduce a novel enumeration attack against VMAT [1] to highlight the vulnerability of converting additive aggregation functions to MIN aggregation via probabilistic sampling. We observe that a compromised sensor node can exploit the vulnerability of probabilistic sampling by enumerating all possible readings to find the one that leads to a significantly inflated aggregation result. In other words, the long-held view that a single compromised node falsifying its local value has a limited impact on final aggregation results does not always hold. While VMAT has incorporated a verifiable random number generation mechanism to prevent compromised sensor nodes from generating arbitrary random samples, we show that such mechanism is necessary but inadequate. As a countermeasure, we also introduce an effective defense against the enumeration attack. Our contributions in this paper can be summarized as follows.

- We introduce a novel enumeration attack against VMAT to highlight the danger of converting additive aggregation into MIN aggregation, whereby a small number of compromised sensors could severely manipulate the final aggregation result.
- We theoretically analyze the impact of enumeration attacks and validate our analysis using simulation studies.
- We introduce an effective countermeasure against enumeration attacks by requiring every sensor node to commit to its reading prior to knowing the random seed for generating random synopsis. We confirm the efficacy and efficiency of the countermeasure via simulation studies.

The rest of this paper is structured as follows. Section II discusses the related work. Section III presents the network and adversary models. Section IV reviews the VMAT scheme. Section V presents the enumeration attack and its evaluation. Section VI presents a defense against the enumeration attack and evaluates its performance. Section VII finally concludes this paper.

II. Related Work

Secure data aggregation in wireless sensor networks and related systems has been studied extensively in the past.

Existing solutions can be generally classified into two categories. The first category such as [6]–[9] provides accurate aggregation result at the base station. Most of these schemes [6]–[8] ensure aggregation-result integrity by requiring intermediate nodes to commit to partial aggregation-results through cryptographic means. SIES [9] explores homomorphic encryption to detect intermediate nodes modifying partial aggregation results. The second category such as [1], [10], [11], [13], [14] aims to provide a statistical estimation of the aggregate result with probabilistic guarantee. SIA [10] considers a single-aggregator model and statistically detects false aggregation results via random sampling and interactive proof, which is subsequently improved in [11] to realize secure approximate-median aggregation. A secure aggregation scheme based on verifiable set sampling was introduced in [13]. Synopsis diffusion [12] is a robust aggregation framework against packet loss that explores multi-path routing and duplicate-insensitive aggregation, which is improved in [15] to enable detection of false subaggregates and [14] to tolerate false subaggregates.

While most of these solutions [1], [6]–[11], [13], [14], [16] focus on detecting intermediate nodes manipulating partial aggregation results, there are a few attempts aiming at identifying compromised nodes during data aggregation in addition to VMAT [1]. Early proposals [17], [18] rely on expensive public-key cryptography operations and group testing to identify malicious nodes. Xu *et al.* [19] proposed an improvement for SDAP [7] to identify malicious nodes via statistical abnormality detection and random node grouping. Their scheme is ineffective if the attacker adopts its behavior according to the statistical detection rules. In [20], a secure aggregation scheme was introduced to pinpoint intermediate nodes that drop partial aggregation results. The approach, unfortunately, incurs a communication overhead linear to the total number of sensor nodes, which largely nullifies the benefit of in-network aggregation. In [21], Li *et al.* introduced a secure SUM aggregation protocol to misbehaving intermediate aggregators by having every intermediate node's partial aggregation result checked by its children and parent, which is ineffective against two colluding parent and child nodes. In addition, there is a general consensus [6], [7], [14], [15] that a compromised node forging its own reading is fundamentally difficult to detect but has limited impact on robust aggregation functions such as SUM and COUNT [5].

III. Network and Adversary Models

In this section, we introduce our system and adversary models.

A. Network Model

We consider a multi-hop wireless sensor network comprising a base station and n sensor nodes. Each sensor node i has a sensed reading d_i in the range $\{1, \ldots, k\}$. The base station intends to learn $f(d_1, \ldots, d_n)$, where $f(\cdot, \ldots, \cdot)$ is some aggregation function such as MAX/MIN, SUM, AVERAGE, and COUNT. The aggregation is performed over an aggregation tree, which is the directed tree rooted at the base station formed by the unique path from every sensor node to the base station.

B. Adversary Model

We assume that the base station has adequate computation and energy resources and is safeguarded from possible attacks. In contrast, sensor nodes are constrained in computation and communication resources and may be compromised by the attacker, e.g., through physical capture. Once compromised, all the information stored at the sensor node such as cryptographic keys is revealed to the attacker. The attacker aims to have the base station accept a significantly inflated aggregation result without being detected. We consider the following two attacks in this paper.

- A compromised node may falsify its own sensed reading, which may or may not be in the valid reading range.
- A compromised node may modify or drop a partial aggregation result.

We further assume that the attacker can compromise up to c sensor nodes and that all the compromised nodes can collude in an arbitrary fashion under the instruction of the attacker. We focus on the attacks targeting data aggregation in this paper and refer to the rich literature (e.g., [22]–[27]) for other possible attacks on wireless sensor networks.

IV. Review of VMAT

In this section, we briefly review the VMAT scheme and how to convert additive aggregation functions into MAX aggregation.

VMAT [1] is a representative secure aggregation scheme built upon efficient symmetric-key operations with the capabilities of pinpointing and revoking malicious nodes. Under VMAT, each node shares one or multiple secret keys, called edge keys, with each of its neighbor, and a distinct secret key with the base station. The key component of VMAT is a secure MIN aggregation scheme. During the aggregation phase, each sensor node creates a message consisting of its node ID, sensor reading, and a MAC encrypted with an edge key shared with its parent. Each intermediate node receives the messages from its children and forwards the message with the smallest reading among all the messages from its children and itself. At the end of the aggregation phase, the base station obtains the minimal reading among all sensor nodes and verifies whether this minimal reading has a valid MAC. During the confirmation phase, the base station uses authenticated broadcast to announce the minimum value it received. If the minimum value is higher than the true minimal value, then the sensor node with the true minimal value can detect it and issue a veto message to be flooded back to the base station.The base station can then revoke one of the edge keys used by the reporter sensor through finding out between which neighboring sensors the value contained in the veto was dropped without an even smaller value being forwarded. We refer readers to [1] for more details of the secure MIN aggregation protocol.

VMAT explores the distributed randomized algorithm proposed in [28] to convert additive aggregation such as SUM and COUNT into MIN aggregation. Consider SUM aggregation as an example. To compute $S = \sum_{i=1}^{n} d_i$, each node i with reading d_i generates m mutually independent random synopses $s_{i,1}, s_{i,2}, \ldots, s_{i,m}$ from an exponential distribution $\mathsf{Exp}(d_i)$ with mean $1/d_i$. All n sensor nodes then participate in m parallel instances of secure MIN aggregation to allow the base station to obtain $s_1^{\min}, s_2^{\min}, \ldots, s_m^{\min}$, where $s_j^{\min} = \min(s_{1,j}, s_{2,j}, \ldots, s_{n,j})$ for all $1 \leq j \leq m$. The sum of all d_i can then be estimated as

$$\hat{S} = \frac{m}{\sum_{j=1}^{m} s_j^{\min}},$$

which has been shown [28] to be an unbiased estimator of S. In addition, when $m = \Theta(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$, $\hat{S}$ is within $((1-\epsilon)S, (1+\epsilon)S)$ with probability at least $1-\delta$. AVERAGE and COUNT aggregates can be realized in a similar fashion.

To prevent a compromised node from generating arbitrarily small synopsis, VMAT uses a deterministic pseudorandom number generator to ensure that any synopsis must correspond to a valid reading in range. In particular, the deterministic pseudorandom number generator takes the sensor reading d_i, node ID i, and a nonce s as input and outputs m synopsis $s_{i,1}, \ldots, s_{i,m}$. On receiving $s_1^{\min}, s_2^{\min}, \ldots, s_m^{\min}$, the base station can verify that every minimal synopsis is indeed generated from a valid reading. Unfortunately, we will show in the next section that this mechanism alone is necessary but inadequate.

V. Enumeration Attack

In this section, we use SUM aggregation as an example to introduce a novel data enumeration attack.

A. Attack

In enumeration attack, a compromised sensor node aims to inflate the final aggregate at the base station. In comparison to the naive attack in which a compromised node simply reports the maximum reading in range, enumeration attack is more effective by causing the aggregation result significantly deviating from the true aggregation result.

Enumeration attack exploits the vulnerability that a compromised sensor node can report arbitrary reading of its own. Recall that in VMAT, every node i with reading d_i generates m independent synopsis from an exponential distribution with mean $1/d_i$, and the aggregation result is computed from the m minimal synopsis across all the sensor nodes. Recall that a valid sensed reading is in the range $\{1, \ldots, k\}$. If the sensor node simply reports the maximum reading k, each of its m synopsis is an exponential random variable with mean $1/k$. In enumeration attack, a compromised sensor node attacks one synopsis of its choice. Consider as an example that a compromised sensor node i attacks synopsis $s_{i,1}$. Node i can compute one synopsis for each possible reading $1, \ldots, k$ using the verifiable random number generator $\mathsf{DRNG}(s, d, ID)$ to find the reading d^* that leads to the smallest synopsis s_d as

$$d^* = \arg\min \mathsf{DRNG}(s, d, ID).$$

It then faithfully participates in the secure MIN aggregation with s_d.

We say the enumeration attack succeeds if s_d happens to be smaller than all the synopsis s_{j_1} generated by non-compromised sensor nodes. It is easy to see that under enumeration attack, the synopsis $s_{i,1}$ is the minimal of k independent exponential random variables with means $1, 1/2, \ldots, 1/k$, respectively, which is smaller than the one generated from the maximum reading k with high probability. In other words, enumeration attack allows a sensor node to generate a much smaller synopsis with high probability.

Multiple compromised sensor nodes can collude to maximize the impact of the enumeration attack. In particular, if the attacker has $c > 1$ sensor nodes, the attacker can instruct each compromised sensor node to attack one distinct synopsis or evenly allocate the compromised sensor nodes across m synopsis if $c > m$. In the worst case, if enumeration attack succeeds for every synopsis, then the final aggregation result computed by the base station is independent from any of the non-compromised sensor nodes' reading.

B. Theoretical Analysis

We first analyze the probability that a single compromised sensor node can succeed in launching enumeration attack. Without loss of generality, we consider one compromised sensor node i and g non-compromised sensor nodes and assume that node i intends to attack synopsis $s_1^{\min}$. We have the following theorem regarding the success probability of a single node attacking one synopsis.

Theorem 1. *Assume that there are g non-compromised sensor nodes. Further assume that the readings of non-compromised sensor nodes are i.i.d. random variables with probability distribution $\mathsf{Pr}(d_j = x) = p_x$ where $1 \le x \le k$. The probability that a single compromised node can successfully launch enumeration attack against a single synopsis is given by*

$$P_{\mathit{succ}} = \int_0^\infty \lambda e^{-\lambda t} \cdot (\sum_{y=1}^{k} p_y e^{-yt})^g dt. \tag{1}$$

Proof: Without loss of generality, assume that a compromised sensor node i aims to attack synopsis $s_1^{\min}$. The enumeration attack succeeds if node i can find a reading $d_i \in \{1, \dots, k\}$ that results in its synopsis $s_{i,1}$, being the minimum among all $s_{1,1}, \dots, s_{n,1}$. Let s_{em} be the synopsis generated by node i under enumeration attack. We can see that

$$s_{\text{em}} = \min(s[1], s[2] \dots, s[k]),$$

where $s[1], s[2] \dots, s[k]$ are mutually independent exponential distributed random variables with means $1, 1/2, \dots, 1/k$, respectively. It follows that s_{em} is an exponential random variable with p.d.f.

$$f(s_{\text{em}} = t) = \lambda e^{-\lambda t}$$

for $t \ge 0$, where $\lambda = k(k+1)/2$.

Assume that there are g non-compromised sensor nodes. Let $s_{j,1}$ be the synopsis generated by a non-compromised sensor node j. It follows that

$$\begin{aligned}\mathsf{Pr}(s_{j,1} \le t) &= \sum_{x=1}^{k} \mathsf{Pr}(s_{j,1} \le t | d_j = x) \cdot \mathsf{Pr}(d_j = x) \\ &= \sum_{x=1}^{k} (1 - e^{-xt}) p_x \\ &= 1 - \sum_{x=1}^{k} p_x e^{-xt}.\end{aligned}$$

Let $s_g^{\min}$ be the minimal synopsis among g non-compromised sensor nodes. We have

$$\begin{aligned}\mathsf{Pr}(s_g^{\min} \le t) &= 1 - \mathsf{Pr}(s_g^{\min} > t) \\ &= 1 - \prod_{j=1}^{g} \mathsf{Pr}(s_{j,1} > t) \\ &= 1 - (\sum_{y=1}^{k} p_y e^{-xy})^g.\end{aligned}$$

We finally have

$$\begin{aligned}P_{\text{succ}} &= \mathsf{Pr}(s_{\text{em}} < s_g^{\min}) \\ &= \int_0^\infty \lambda e^{-\lambda t} \cdot \mathsf{Pr}(s_g^{\min} > t) dt \\ &= \int_0^\infty \lambda e^{-\lambda t} \cdot (\sum_{y=1}^{k} p_y e^{-yt})^g dt\end{aligned}$$

■

We also have the following theorems regarding the expected number of synopsis successfully attacked and the optimal strategy of allocating compromised nodes to synopsis.

Theorem 2. *Assume that there are c compromised sensor nodes and g non-compromised sensor nodes. Suppose that the attacker allocate c_j nodes to attack the jth synopsis for $1 \le j \le m$, where $\sum_{j=1}^{m} c_j = c$. The expected number of synopsis successfully attacked is given by*

$$\mathsf{E}(\hat{m}) = m - \sum_{j=1}^{m} (1 - P_{\mathit{succ}})^{c_j},$$

where P_{succ} is given in Eq. (1).

Theorem 3. *Assume that there are c compromised sensor nodes. The optimal attack strategy is to assign the compromised nodes to synopsis in a round robin fashion, i.e., the ith compromised node to attack the jth synopsis, where*

$$j = i \mod m.$$

The proofs of the two theorems are straightforward and omitted here due to space constraints.

C. Simulation Results

We conduct simulation studies to validate our theoretical analysis. Specifically, we consider $n = 1000$ sensor nodes and $m = 50$ synopsis as the default setting and evaluate the impact of several parameters. We also consider four probability distributions of non-compromised nodes' readings. Every point is the average of 500 runs, each with a distinct random seed.

Figs. 1(a) to 1(c) illustrate the impact of valid reading range and the number of non-compromised nodes on P_{succ}, where we assume that the readings from non-compromised sensor nodes follow four uniform distributions $\mathsf{U}(5, 15), \mathsf{U}(25, 35), \mathsf{U}(45, 55)$ and $\mathsf{U}(65, 75)$ with mean $10, 30, 50$ and 70, respectively. First of all, we can see that the theoretical results match the simulation results very well, which validate our theoretical analysis. We can see from Fig. 1(a) that the success probability increases as the reading range increases. This is expected, as the larger the reading range, the more readings the compromised sensor node can try to find the minimal possible synopsis, the higher the probability that its synopsis is smaller than all the synopsis generated by the non-compromised sensor nodes, and vice versa. In addition, the larger the expectation of the non-compromised node's reading, the lower the success probability. This is because it is more likely for non-compromised nodes to generate smaller synopsis with larger readings. We can see from Fig. 1(b) that the success probability decreases as the number of non-compromised nodes increases. This is also anticipated, as the more non-compromised nodes, the smaller the minimal synopsis among all the synopsis generated by the non-compromised nodes. Finally, we can see from Fig. 1(c) that the number of synopsis successfully attacked increases as the number of compromised sensor nodes increases. We can also observe that the pace of increasing slows down after the number of compromised nodes exceeds the number of synopsis.

Figs. 2(a) to 2(c) compares the relative estimation errors under enumeration attack and naive attack where every compromised sensor node simply reports the maximum reading in range. The relative estimation error is defined as $|\hat{S}_{\text{att}} - \hat{S}|/\hat{S}$, where $\hat{S}_{\text{att}}$ and $\hat{S}$ are the sums estimated by the base station under attack and under no attack, respectively. We assume that the average readings of non-compromised sensor nodes are 50, 100, and 150, respectively. We can see from Fig. 2(a) that the relative estimation error increases as the number of compromised nodes increases under both naive and enumeration attacks, which is anticipated. In addition, the relative estimation error under the naive attack is very limited, which is in line with the long-held view and conclusions in [5]. However, the relative estimation error under enumeration attack is always significantly higher than that under the naive attack. For example, enumeration attack can inflate the sum aggregation result by 40% and 100% with 25 and 50 compromised sensor nodes, respectively. Such large aggregation errors highlight the severe impact of the enumeration attack. Moreover, the larger the average reading of non-compromised nodes, the smaller the impact of both naive attack and enumeration attack. We can also see from Fig. 2(b) that the relative estimation error decreases as the number of synopsis increases. This is expected, as if the number of compromised nodes remains the same, the proportion of the synopsis successfully attacked decreases as the number of synopsis increases. When the number of synopsis exceeds 115, the relative estimation error under enumeration attack is about the same as that under the naive attack. Finally, Fig. 2(c) shows that the aggregation error decreases as the number of non-compromised nodes increases. This is because the more non-compromised nodes results, the lower the success probability, the fewer synopsis successfully attacked, and vice versa.

VI. Countermeasure

In this section, we introduce an effective countermeasure against the enumeration attack.

A. Countermeasure

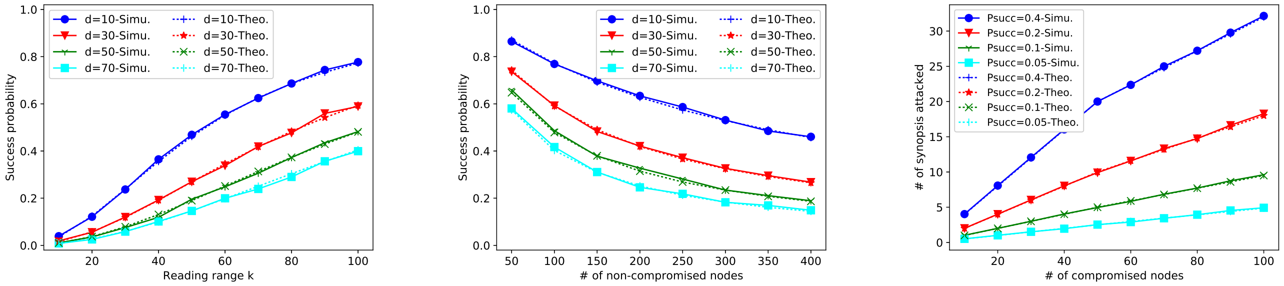
We observe that the enumeration attack is possible because compromised nodes know the nonce used for generating synopsis before choosing its reading. An effective way to defend against enumeration attack is to require every sensor node to commit to its reading before knowing the nonce, so that there is no opportunity for compromised sensor nodes to enumerate all possible readings. Our countermeasure requires each node to commit to its reading and forward the commitment to selected witnesses in its neighborhood, which allows the base station to verify whether the synopsis is generated before the sensor node knowing the random seed. In what follows, we detail the operations.

During network initialization, every node i learns the IDs of all the nodes in its h-hop neighborhood, denoted by $\mathcal{N}^h(i)$, and the base station learns the complete topology of the network. To initiate a data aggregation process, the base station broadcasts a command with a random nonce s_1. On receiving the command, each sensor node i with reading d_i computes a commitment as

$$\mathsf{Commit}_i = \langle ID_i, d_i, MAC(ID_i || s_1 || d_i) \rangle,$$

where $MAC(\cdot)$ denotes message authentication code computed using the secret key shared between node i and the base station, and $||$ denotes concatenation. It selects λ nodes from $\mathcal{N}^h(i)$ to serve as its witnesses using a deterministic random number generator seeded by the nonce s_1 and its node ID, where $\lambda \geq 1$ is a system parameter. Node i then forwards Commit_i to each of the λ witnesses.

Every node then follows VMAT to generate m synopsis and participates in m instances of secure MIN aggregation. In particular, the base station broadcasts another nonce s_2. At the end of the aggregation phase, the base station obtains $s_1^{\min}, s_2^{\min}, \ldots, s_m^{\min}$, i.e., m minimal synopsis across all n sensor nodes. For every $s_j^{\min} (1 \leq j \leq m)$, the base station determines the ID of the node that generated this synopsis and verifies that $s_j^{\min}$ is indeed generated from a valid reading as in VMAT. Consider $s_j^{\min}$ as an example. Assume that node i with reading d_i generated $s_j^{\min}$. During the confirmation phase, the base station uses authenticated broadcast to announce $\langle ID_i, d_i, s_j^{\min} \rangle$ to all the nodes. Every witness of node i, say node w, then sends a message $\langle ID_w, \mathsf{Commit}_i, MAC(ID_w || \mathsf{Commit}_i) \rangle$ to the base station.

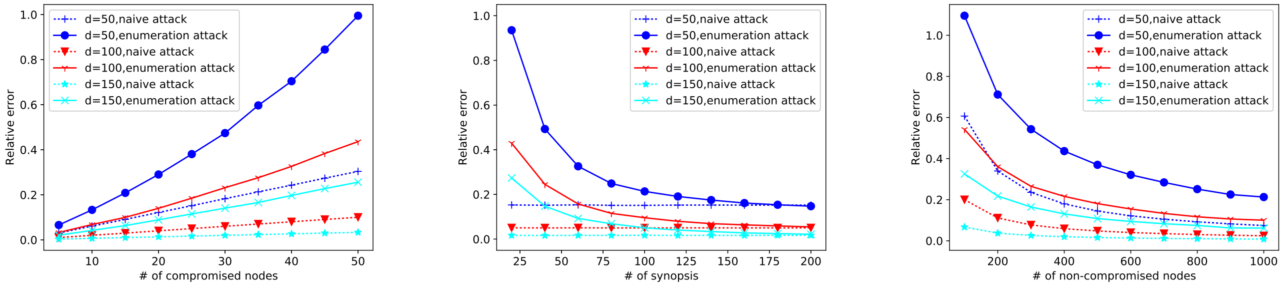


(a) P_{succ} vs. reading range

(b) P_{succ} vs. # of non-compromised nodes

(c) $\hat{m}$ vs. # of compromised nodes

Fig. 1. Success probability of enumeration attack, where $k = 100$, $n = 100$ and $m = 50$



(a) Aggregation error vs. # of compromised nodes

(b) Aggregation error vs. # of synopsis

(c) Aggregation error vs. # of non-compromised nodes

Fig. 2. Comparison of enumeration attack and naive attack in estimation error, where $k = 200$, $n = 500$, $c = 25$, and $m = 50$

On receiving the message, the base station first verifies whether node w is a valid witness for node i. If so, the base station verifies the MACs in the message and Commit_i. If the verification succeeds, the base station knows that node i's reading d_i was committed before knowing the nonce s_2.

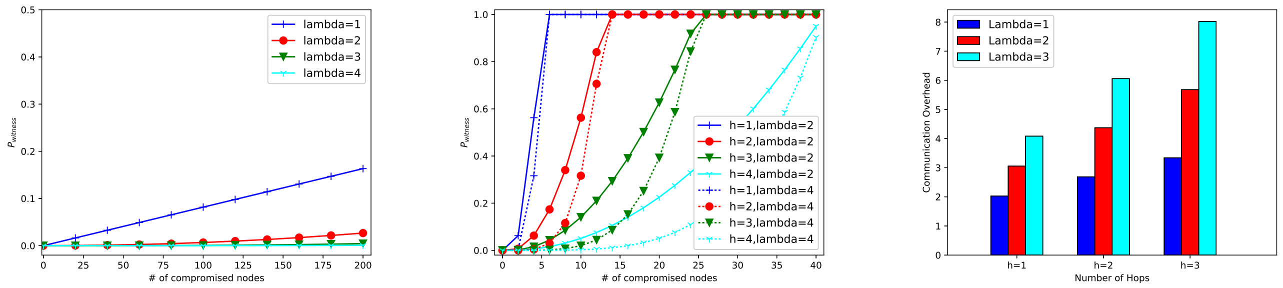
B. Simulation Results

We also use simulation studies to evaluate the performance of our countermeasure. We consider a 35×35 grid sensor network with $n = 1225$ sensor nodes, where the base station is located at one of the corners. Every sensor node (except the ones near the boundary) has 4 one-hop neighbors, 12 two-hop neighbors, 24 three-hop neighbors, and 40 four-hop neighbors. We measure the communication overhead incurred by our countermeasure as the average number of extra message transmissions per node and per synopsis.

Fig. 3(a) shows the impact of the number of compromised nodes on P_{witness}, the probability of all witnesses being compromised under the assumption that compromised nodes are distributed uniformly at random. As we can see, the larger the λ, the smaller P_{witness}, and vice versa. This is expected, as P_{witness} is approximately $(\frac{c}{n})^{\lambda}$. For example, when 10% of the nodes are compromised, the probability that all witnesses are compromised is 0.01 if $\lambda = 2$. A compromised sensor node can successfully launch enumeration attack on one selected synopsis if it can find a reading that leads to the minimal synopsis and all λ witnesses are also compromised.

The attacker may choose to compromise one selected sensor node and then the nodes within its h-hop neighborhood. Fig. 3(b) shows P_{witness} varying with the number of compromised nodes under different h. As we can see, the more compromised nodes, the smaller h, the higher P_{witness}, and vice versa. This is expected, as the λ witnesses are chosen uniformly at random from all the nodes within the h-hop neighborhood. When the number of compromised nodes exceeds the number of nodes in the h hop neighborhood, P_{witness} becomes one. In this case, the success probability is reduced to the probability that the compromised node can successfully find a reading that leads to the minimal synopsis among all sensor nodes.

Fig. 3(c) shows the impact of λ, the number of witnesses that store the commitment, on the extra communication overhead incurred by the proposed countermeasure. It is not surprising to see that the larger the λ, the more message transmissions incurred by the proposed countermeasure. In addition, the number of message transmissions also increases as h increases for the same λ. The reason is that the larger h, the larger the average distance between a node and its witnesses. Overall, our countermeasure incurs a small number of extra message transmissions. For example, when $h = 3$ and $\lambda = 3$, the proposed countermeasure incurs approximately 8 extra message transmissions over VMAT.



(a) P_{witness} vs. # of compromised nodes

(b) P_{witness} vs. # of compromised nodes

(c) # of hops vs. communication overhead

Fig. 3. Performance of the countermeasure, where $n = 1225$.

VII. Conclusion

In this paper, we have introduced a novel enumeration attack against VMAT to highlight the security vulnerability of a sensor node reporting arbitrary readings. In comparison with the naive attack, the enumeration attack allows a single compromised sensor node to cause significantly higher estimation error at the base station without being detected. We have also introduced an effective countermeasure against the enumeration attack. Theoretical analysis and simulation studies have confirmed the severe impact of the enumeration attack and the effectiveness of the proposed countermeasure.

Acknowledgment

We would like to thank the anonymous reviewers for their insightful comments that help improve the quality of this paper. This work was supported in part by the US National Science Foundation under grants CNS-1700039, CNS1651954 (CAREER), and CNS-1718078.

References

[1] B. Chen and H. Yu, “Secure aggregation with malicious node revocation in sensor networks, in *IEEE ICDCS’11*, Minneapolis, MN, 2011, pp. 581–592.

[2] D. P. Agrawal, “Applications of Sensor Networks,” *Embedded Sensor Systems*, Springer, 2017, pp. 35–63.

[3] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “Tag: A tiny aggregation service for ad-hoc sensor networks,” *SIGOPS Oper. Syst. Rev.*, vol. 36, no. SI, pp. 131–146, Dec. 2002.

[4] J. Zhao, R. Govindan, and D. Estrin, “Computing aggregates for monitoring wireless sensor networks,” in *IEEE International Workshop on Sensor Network Protocols and Applications*, May 2003, pp. 139–148.

[5] D. Wagner, “Resilient aggregation in sensor networks,” in *SASN’04*, Washington DC, 2004, pp. 78–87.

[6] H. Chan, A. Perrig, and D. Song, “Secure hierarchical in-network aggregation in sensor networks,” in *ACM CCS’06*, Alexandria, Virginia, USA, Oct. 2006, pp. 278–287.

[7] Y. Yang, X. Wang, S. Zhu, and G. Cao, “SDAP: a secure hop-by-hop data aggregation protocol for sensor networks,” in *ACM MobiHoc’06*, Florence, Italy, May 2006, pp. 356–367.

[8] K. B. Frikken and J. A. Dougherty, IV, “An efficient integrity-preserving scheme for hierarchical sensor aggregation,” in *WiSec’08*, Alexandria, VA, 2008, pp. 68–76.

[9] S. Papadopoulos, A. Kiayias, and D. Papadias, “Secure and efficient in-network processing of exact sum queries,” in *IEEE ICDE’11*, April 2011, pp. 517–528.

[10] B. Przydatek, D. Song, and A. Perrig, “SIA: Secure information aggregation in sensor networks,” in *ACM SenSys’03*, Los Angeles, CA, Nov. 2003, pp. 255–265.

[11] S. Roy, M. Conti, S. Setia, and S. Jajodia, “Securely computing an approximate median in wireless sensor networks,” in *SecureComm’08*, Istanbul, Turkey, 2008, pp. 6:1–6:10.

[12] S. Nath, P. B. Gibbons, S. Seshan, and Z. Anderson, “Synopsis diffusion for robust aggregation in sensor networks,” *ACM Trans. Sen. Netw.*, vol. 4, no. 2, pp. 7:1–7:40, Apr. 2008.

[13] H. Yu, “Secure and highly-available aggregation queries in large-scale sensor networks via set sampling,” in *IPSN’09*, San Francisc, CA, April 2009, pp. 1–12.

[14] S. Roy, M. Conti, S. Setia, and S. Jajodia, “Secure data aggregation in wireless sensor networks: Filtering out the attacker’s impact,” *IEEE Trans. Inf. Forensic Secur.*, vol. 9, no. 4, pp. 681–694, April 2014.

[15] ——, “Secure data aggregation in wireless sensor networks,” *IEEE Trans. Inf. Forensic Secur.*, vol. 7, no. 3, pp. 1040–1052, 2012.

[16] L. Hu and D. Evans, “Secure aggregation for wireless networks,” in *SAINT-W’03*, Washington, DC, USA, 2003.

[17] P. Haghani, P. Papadimitratos, M. Poturalski, K. Aberer, and J.-P. Hubaux, “Efficient and robust secure aggregation for sensor networks,” in *NPSEC’07*, Washington, DC, USA, 2007, pp. 1–6.

[18] G. Taban and V. D. Gligor, “Efficient handling of adversary attacks in aggregation applications,” in *ESORICS’08*, Berlin, Heidelberg, 2008, pp. 66–81.

[19] X. Xu, Q. Wang, J. Cao, P.-J. Wan, K. Ren, and Y. Chen, “Locating malicious nodes for data aggregation in wireless networks,” in *IEEE INFOCOM’12*, Orlando, FL, March 2012, pp. 3056–3060.

[20] S. Choi, G. Ghiniţă, and E. Bertino, “Secure sensor network sum aggregation with detection of malicious nodes,” in *IEEE LCN’12*, Clearwater, FL, Oct 2012, pp. 19–27.

[21] H. Li, K. Li, W. Qu, and I. Stojmenovic, “Secure and energy-efficient data aggregation with malicious aggregator identification in wireless sensor networks,” *Future Generation Computer Systems*, vol. 37, pp. 108 – 116, 2014.

[22] M. Li, I. Koutsopoulos, and R. Poovendran, “Optimal jamming attacks and network defense policies in wireless sensor networks,” in *IEEE INFOCOM’07*, May 2007, pp. 1307–1315.

[23] R. Zhang, Y. Zhang, and K. Ren, “DP^{2}AC: Distributed privacy-preserving access control in sensor networks,” in *IEEE INFOCOM’09*, Rio de Janeiro, Brazil, Apr. 2009.

[24] R. Zhang, J. Shi, and Y. Zhang, “Secure multidimensional range queries in sensor networks,” in *ACM MobiHoc’09*, New Orleans, LA, May 2009, pp. 197–206.

[25] R. Zhang and Y. Zhang, “LR-Seluge: Loss-resilient and secure code dissemination in wireless sensor networks,” in *IEEE ICDCS’11*, Minneapolis, Minnesota, June 2011.

[26] R. Zhang, J. Shi, Y. Zhang, and J. Sun, “Secure cooperative data storage and query processing in unattended tiered sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 433–441, Feb. 2012.

[27] R. Zhang, J. Shi, Y. Zhang and X. Huang, ”Secure top-k query processing in unattended tiered sensor networks,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4681-4693, Nov. 2014.

[28] D. Mosk-Aoyama and D. Shah, “Computing separable functions via gossip,” in *ACM PODC’06*, Denver, CO, July 2006, pp. 113–122.