
Yu Zhang

School of Computer Science, Northwestern Polytechnical University, Xi’an, China
Email: {zhangyu, zhouxs}@nwpu.edu.cn

Xing She Zhou, Yee Wei Law and Marimuthu Palaniswami

School of Computer Science, Northwestern Polytechnical University, Xi’an, China
Email: zhouxs@nwpu.edu.cn
Email: {ywlaw, palani}@unimelb.edu.au

Abstract — Reprogramming protocols provide a convenient way to update program images via wireless communication. In hostile environments where there may be malicious attacks against wireless sensor networks, the process of reprogramming faces threats from potentially compromised nodes. While existing solutions can provide authentication services, they are insufficient for a new generation of network coding-based reprogramming protocols in wireless sensor networks. We present a security approach that is able to defend pollution attack against reprogramming protocols based on network coding. It employs a homomorphic hashing function and an identity-based aggregate signature to allow sensor nodes to check packets on-the-fly before they accept incoming encoded packets, and introduces an efficient mechanism to reduce the computation overhead at each node and to eliminate bad packets quickly. Castalia simulations show that when the 5% of the nodes in a network of 100 nodes are rogue, using our approach, the efficiency of the secure reprogramming protocol based on network coding improves almost ten-fold for a checking probability of 2%.

Index Terms—Wireless Sensor Network, Reprogramming, Secure Code Dissemination, Network Coding, Pollution Attacks, Homomorphic Hashing

I. INTRODUCTION

Reprogramming over wireless links has become crucial service for wireless sensor networks (WSNs) due to the need of fixing bugs or providing new functionalities after a WSN is deployed [1]. Several reprogramming protocols [2], [3], [4], [5] have been proposed for WSNs. Deluge [2] is currently the benchmark reprogramming protocol since it has been included in TinyOS distributions, making it easily accessible by the WSN community. In recent years, the technique of network coding has been applied to improve the performance of existing reprogramming protocols, resulting in such protocols as Rateless Deluge [6], AdapCode [7] and Synapse [8]. Like Deluge, these protocols still divide the new code image into pages, and divide a page into packets. But unlike Deluge, they rely on the use of rateless coding to eliminate the need to convey control messages about which packets require retransmission; with network coding, a node need only receive a sufficient number of distinct, encoded packets to recover a transmitted program. The advantage of network coding based protocols is that they achieve significant savings of energy and communication, especially when packet loss rate or network density increases. There is hence much incentive in using network coding-based protocols for reprogramming in WSNs.

However, the network coding-based reprogramming protocols have a potential problem in hostile environments, where an adversary may attack the sensor nodes by launching famous pollution attack [9], in which a malicious node can send bad encoded packets that consist of bogus data, which leads to erroneous decoding of a large part of the original data upon retrieval. For reprogramming protocols that are not based on network coding, several security approaches [10], [11], [12], [13] have been proposed. All these approaches are extensions to Deluge. The technique common to them is the use of a single digital signature and off-line hash chains [14]. This technique is not sufficient for network coding-based reprogramming protocols because each sensor node produces unique encoded packets which cannot be signed by a base station.

The goal of this paper is to propose a new security approach that is able to defend pollution attack against reprogramming protocols based on network coding. To achieve this goal, we employ a homomorphic hashing function and an identity-based aggregate signature to allow sensor nodes to check packets on-the-fly before they accept incoming encoded packets, and introduce an efficient mechanism to reduce the computation overhead at each node and to eliminate bad packets quickly.

This paper is organized as follows. Section II clarifies our assumptions and the threats which this proposal based. Section III describes our approach in detail. Section IV details efficient mechanism. Section V presents the

II. ASSUMPTIONS AND THREAT MODEL

A. Assumptions

We assume the source of the code images, i.e., the base station, is a powerful node (e.g., a laptop) with sufficient energy supply and the only trusted entity in the network. We assume each sensor node has unique identification information (e.g., MAC address or production number). We assume while each sensor node is resource constrained, it has sufficient memory to store the security mechanisms our approach adds. We assume the packet size is large enough to hold a signature and other information required by a signature packet. This can be satisfied on sensor platforms with IEEE 802.15.4 compliant radios [15], where the maximum payload size is 102 bytes.

B. Threat Model

We assume that individual sensor-nodes are untrusted. An adversary may introduce its own nodes in a network, or it may capture, compromise and re-introduce existing nodes in the network, as sensor nodes are generally not tamper-resistant. With those capabilities, the adversary attempts to launch pollution attacks to reduce the performance of the network and to consume the limited resources (e.g., battery power, memory) on sensor nodes.

III. PROPOSED APPROACH

We choose the homomorphic hashing function based on the hardness of discrete logarithm [16], the message space is \( \mathbb{F}_p^d \), where \( p \) is a prime number and \( p \geq 2^{160} \) for 80-bit security. To check incoming encoded packets on the fly, sensor node i computes a raw hash value \( h_i = H(pkt_i) \), where \( pkt_i \) is a raw data packet, and uses a method to sign \( h_i \), which will be described later, in a way that allows verifiers to verify the authenticity of \( h_i \). When a verifier receives an incoming packet \( pkt = \sum_{i=1}^{j} \alpha_i pkt_i \) along with \( j \) pairs of (raw hashing value, weight) \( (h_i, \alpha_i) \), it first determines whether the hashing values are valid, and then verifies the incoming packet’s integrity by checking if
\[
\prod_{i=1}^{j} h_i^{\alpha_i} = H(pkt)
\] (1)

Figure 1 graphically demonstrates the construction of the incoming encoded packet in our approach.

In the following subsection, we describe a method to authenticate \( h_i \).

A. Assumptions

To verify the authenticity of raw hash values at all intermediate sensor nodes in an energy-efficient manner, the best method providing such a property is an identity-based aggregate signature (IBAS), in which different raw hash values can be authenticated by one single aggregate signature.

As far as we know, there is an IBAS scheme [17] which is suitable for secure reprogramming in WSNs although it has one too strong assumption, i.e., all signers must use a same unique string when signing. However, this assumption can be satisfied in the WSN secure reprogramming service because a unique Imgid or every code image is known to all nodes.

Let \( G_1 \) and \( G_2 \) be two cyclic groups of some large prime order \( q \) that efficiently support a bilinear mapping
\[
\hat{\cdot} : G_1 \times G_1 \rightarrow G_2 : \hat{\cdot}(aQ, bR) = \hat{\cdot}(Q, R)^{ab},
\]
\( Q, R \in G_1, a, b \in \mathbb{Z} \). The IBAS scheme of our approach works as table I.

Any sensor node can verify the signature by checking if
\[
\hat{\cdot}(S, P) = \hat{\cdot}(T, P_{\text{Imgid}}) \hat{\cdot}(Q, \sum_{i=0}^{j} P_{i,0} + \sum_{i=0}^{j} c_i P_{i,1})
\] (2)
TABLE I. THE IBAS OF OUR APPROACH

<table>
<thead>
<tr>
<th>Setup: a private key generator (PKG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. generates groups $G_1$ and $G_2$ of prime order $q$ and an admissible pairing $e: G_1 \times G_2 \rightarrow G_2^\ast$.</td>
</tr>
<tr>
<td>2. chooses an arbitrary generator $P \in G_1$.</td>
</tr>
<tr>
<td>3. picks a random $s \in \mathbb{Z}/q\mathbb{Z}$ as the master key of PKG and sets $Q = sP$.</td>
</tr>
<tr>
<td>4. chooses three cryptographic hash functions $H_1$, $H_2: {0,1}^* \rightarrow G_1$ and $H_3: {0,1}^* \rightarrow \mathbb{Z}/q\mathbb{Z}$.</td>
</tr>
</tbody>
</table>

Private key generation:

| 1. node $i$ receives from the PKG the values of $sP_{\alpha}$ as its private key for $\alpha \in \{0,1\}$, where $P_{\alpha} = H_1(\alpha) \in G_1$. |

Signing: To sigh $h_i$, node $i$

| 1. computes $P_{\text{msgd}} = H_1(\text{Idmsgd}) \in G_1$. |
| 2. computes $cl = H_1(h_i, \text{Idmsgd}) \in \mathbb{Z}/q\mathbb{Z}$. |
| 3. generates random $ri \in \mathbb{Z}/q\mathbb{Z}$. |
| 4. computes signature $(S_i, T_i)$, where $S_i = riP_{\text{msgd}} + sP_{\alpha} + cisP_{\beta}$ and $T_i = riP$. |

Signature Aggregation:

| 1. signatures $(S_i, T_i)$ for $1 \leq i \leq j$ can be aggregated into $(S, T)$, where $S = \sum_{i=1}^{j} S_i$, and $T = \sum_{i=1}^{j} T_i$. |

III. EFFICIENT MECHANISM

To reduce the cryptographic work at each sensor node while preventing malicious packets from infecting large portions of the network, we introduce cooperative mechanism where nodes cooperate in checking malicious incoming packets. By having a number nodes checking at every point in time and making them cooperate, expensive homomorphic hashing can be applied less frequently, hence, improving the performance of the network. Next we describe the details of how such efficient mechanism works.

We assume that nodes check blocks with probability $P_e$. Encoded packets that pass the check are marked as safe, while encoded packets that have not yet been checked are kept in an insecure window. Encoded packets are checked in batches. The batch window is equal to the insecure window. Whenever a node verifies its insecure window, valid encoded packets are marked as safe and the insecure window is reset.

One possible solution to improve computation time is to verify encoded packets in batches, either probabilistically or periodically. In such solution nodes do not check every encoded packet, but they check a window of encoded packets all at once. Batching is possible thanks to the homomorphic property of the hashing functions. Let’s assume we have a window of $J$ encoded packets to verify all together. We can build a batched encoded packets $pk_{\alpha}$ as a linear combination of all $J$ encoded packets. Let $J = \langle\langle pk_{\alpha}^1, c_1\rangle, \ldots, \langle pk_{\alpha}^J, c_J\rangle\rangle$, and let the resulting batched packet $pk_{\alpha} = \sum_{i=1}^{J} p:\alpha_{i}$, with combined coefficient vector $c_{\alpha} = \sum_{i=1}^{J} c_i$. The advantages of using batching is that nodes only need to check one $pk_{\alpha}$ packet rather than $L$ packets. However, this batching scheme is exposed to a specific byzantine attack, which is called the pairwise byzantine attack [18]. To solve this problem, we create a batched encoded vector $pk_{\alpha} = \sum_{i=1}^{J} p:\alpha_{o_{i}}$, using a random set of coefficients, where $c_{\alpha} = \sum_{i=1}^{J} c_i$. Doing this, it’s difficult for an attacker to launch such pairwise byzantine attack.

In our efficient mechanism, sensor nodes do not rely on other nodes to mark encoded packets as safe. However, they actively cooperate with other nodes to detect malicious encoded packets. Whenever a sensor node detects a malicious incoming encoded packet, it sends an alert message to all its neighbors. To prevent sensor nodes that have not been infected from processing the alert message, a given node keeps an insecure-activity table with the ID of: i) those nodes that downloaded incoming packets encoded with insecure window encoded packets, and ii) those nodes that delivered the encoded packets inside the insecure window. Alert messages are propagated from one node to another until all infected nodes are informed. If the insecure window is empty, alert messages are not processed. Alert messages are processed as soon as they are received. However, alert messages are only propagated after the node is convinced that a malicious encoded packet exists. Duplicated alert messages can be received for the same malicious packet since overlays often contain loops. However, such duplicated messages will be discarded when i) the insecure window is empty, or ii) the duplicate message comes from a sensor node that is not in the insecure activity table. In addition to alerting its neighbors, a sensor node takes the following actions: i) it puts encoded packets in the insecure window in isolation to be checked and cleaned in the background, ii) it stops using encoded packets in the insecure window for network coding, and iii) it starts checking encoded packets with probability one until the insecure window is secured and cleaned, thus, preventing new malicious incoming packets from infecting the reprogramming system.

One disadvantage that arises during the quarantine period is that valid encoded packets in the insecure window are not part of the re-encoding process. However, that network coding ensures a high level of representation of packets in the network under such small temporal glitches, thus, maintaining a high level of efficiency in the reprogramming system. To ensure that the insecure window is quickly cleaned and valid encoded packets are back into the system as soon as possible, nodes can adopt a mechanism based on binary batching trees to efficiently identify the malicious encoded packets (motivated by...
Such batching trees work as follows. A node first verifies all incoming encoded packets in the insecure window using batching. If this test does not find any malicious packets, then the process is stopped and all encoded packets are marked as safe. If the batch verification fails, then the insecure window is divided in two halves, which are then checked independently using batching. If one half of the insecure window has not been corrupted, then, all its encoded packets are marked as safe and they are not checked any more. Corrupted parts are again subdivided in two parts until the individual corrupted packets are identified and discarded.

IV. PERFORMANCE EVALUATION

In this section we show that the performance improvement of efficient mechanism in detecting malicious encoded packets when network coding is used for reprogramming.

We have built a simulator by Castalia [19] that allows us to study the damage that malicious nodes can cause in reprogramming system in WSNs under different settings (e.g. new generation programming protocol based on network coding, classic reprogramming protocol, or non-cooperative and cooperative environments).

We let simulator of Castalia generate the overlay topology of the network, where each sensor node has a constant number of neighbors k and the population size was fixed to 50-100 nodes. We randomly choose a portion of malicious nodes that generate malicious encoded packets to attack the reprogramming system. The percentage of the malicious nodes varied between 5-20% in our experiments. We also varied the percentage of bad packets injected by a particular malicious sensor node.

The normal sensor nodes cooperate to disseminate a new code image that has a large number of packets. The reprogramming system uses either protocol based on network coding, or classic protocol (e.g. Deluge) without coding. The simulator uses a page-by-page dissemination strategy and in each page-round each sensor node can receive and decode at most one page packets. In each page-round, each sensor node with probability p verifies the validity of the encoded packets and, if one or more of them are corrupted, then the node removes them. If cooperative protection is used, then, sensor nodes behave as described in Section IV to alert other infected sensor nodes, so that they also check their content. We measure the number of transmissions of correct encoded packets and the number of corrupted encoded packets.

A. Impact of the Probability of Checking

Fig.1 and Fig.2 respectively shows that the percentage of valid encoded packets and the number of infected nodes as a function of the probability of checking. We consider a network of 100 nodes in which 10% of them are attackers and send bad packets at a rate of 10%.

Fig.1 shows that the benefits of collaboration are achieved even for a minimum probability of checking (e.g. 1% provides 89% efficiency). Larger probability of checking results in small increase in the efficiency, hence, not justifying the extra-computational effort.

Fig.2 shows that the average number of infected sensor nodes per round drops fast as we increase the probability of checking.

B. Comparison with Different Settings

Table II shows the relationship between bad encoded packets and the probability of checking for a network of 100 nodes in which 5% of them are attackers. We study the case of network coding, and no coding, with a cooperative and a non-cooperative system.

From table II we can see that the classic reprogramming system like Deluge that do not use encoding only suffer from minor damage in the network and is independent of the probability of checking. However, this is not the case for network coding since malicious packets get quickly re-encoded in the network.
From Table II we can see that with network coding and no cooperation, the damage of the system decreases linearly with the checking probability, which requires nodes to check almost every incoming encoded packet to have acceptable levels of efficiency. If cooperation is introduced, then the performance of network coding improves. Actually, a cooperative mechanism improves the efficiency of the secure reprogramming system almost ten-fold for a checking probability of 2% (see Table II). Thus, with cooperation the system is able to limit the propagation of bad packets. Moreover, observe that the percentage of bad packets is very close to the minimum, which in this simulation is 5% Table II shows the relationship between bad encoded packets and the probability of checking for a network of 100 nodes in which 5% of them are attackers. We study the case of network coding, and no coding, with a cooperative and a non-cooperative system.

C. Impact of the Rate of Bad Packets

Fig. 3 shows the efficiency of the system for a network of 50 sensor nodes where each node uses network coding and checks with 1% probability as a function of bad packets’ rate. We observe that the efficiency of the system largely depends on whether a cooperative mechanism is in place, dropping to less than 20% if this is not the case. We also observe that the efficiency of the system increases linearly as the rate of infection of a malicious node decreases, achieving 90% efficiency for a bad packet rate of 10%.

![Figure 4 Relationship between good packets and bad packets generated by an attacker](image)

V. RELATED WORK

In recent years, solutions for securing Deluge have been proposed and improved upon [10], [11], [12]. They are mainly distinguished through their structure, granularity and strength of hashing. Hyun et al. [12] derive a hash tree only for the first page. This improvement drastically reduces the amount of overhead. There exists a scheme that uses a different one-way key chain for each hop from the base station [20], with the disadvantage of transmission overhead increasing with hop count. Merkle’s one-time signature has been proposed as a ROM-friendly alternative to digital signature [21]. The trade-off is that the nodes need to be stateful.

The preceding solutions only consider the authentication of data packets. A μTESLA-like scheme has been proposed for authenticating the 'Inject', 'Reboot', 'Erase' commands [1]. Another scheme [22] has been proposed for authenticating advertisement (ADV) and request (REQ) messages. Despite the progress in secure reprogramming, none of the aforementioned approaches can be used for network coding based reprogramming protocols because in these approaches, the packets of a page are transmitted as is (i.e., not encoded). Alternative approaches are necessary.

The closest solution to the problem in the literature is from Law et al. [23], who propose Sreluge which is resistant to pollution attacks. Sreluge employs a neighbor classification system and a time series forecasting technique to isolate polluters, and a combinatorial technique to decode data packets in the presence of polluters before the isolation is complete. However, to detect the presence of polluted encoded packets in the page, this solution is only by failing to verify a decoded page using its corresponding hash. By the time a bad packet is detected, precious energy might have been wasted.

VI. CONCLUSION

In this paper, we presented an approach that is able to defend pollution attack against reprogramming protocols based on network coding. This approach employs a
homomorphic hashing function and an identity-based aggregate signature to allow sensor nodes to check packets on-the-fly before they accept incoming encoded packets, and introduces an efficient mechanism to reduce the computation overhead at each node and to eliminate bad packets quickly.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their valuable comments. This work was partly supported by the National Natural Science Foundation of China under Grant No.60573161 and National Key Technology R&D Program of China under Grant No.2007BAD79B02.

REFERENCES


[15] IEEE Std 802.15.4-2003, “IEEE standard for information technology C telecommunications and information exchange between systems C local and metropolitan area networks Cspecific requirements C part 15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs).”


Yu Zhang received his B.S., M.S. degree and Ph.D in computer science from the Northwestern Polytechnical University, China. He is assistant professor at the Department of Information Engineering for Northwestern Polytechnical University. His main research interests include wireless sensor networks security and distributed computing.

Prof. Xing She Zhou received his ME from the Northwestern Polytechnical University, China. He is Ph. D. supervisor for Northwestern Polytechnical University and serves as dean of the school of Computer Science and Technology for Northwestern Polytechnical University. Professor Zhou is a Standing Director of China Computer Foundation (CCF), a reviewer of National Natural Science Foundation of China, a Vice-chair member of Embedded System Technical Committee of CCF, the Director of the Shaanxi Key Laboratory of Embedded System Technology, and an editor of the Chinese Journal of Computers. His research interests include embedded software design and verification, pervasive computing, sensor networks, and embedded test and control system.

Yee Wei Law received the B.S. degree in computer engineering from the University of Southampton, UK, ME degree in electrical engineering from the Nanyang Technological University, Singapore and Ph. D from the University of Twente, Netherland. Presently, He is a Research Fellow of the Department of Electrical Engineering for the University of Melbourne. His research interests include wireless sensor network security, wireless communication and sensor network application.

Prof. Marimuthu Palaniswami received his ME from the Indian Institute of Science, India, MEngSc from the University of Melbourne and Ph.D from the University of Newcastle, Australia before joining the University of Melbourne. He served as a co-director of an active research centre, the Centre of Expertise on Networked Decision & Sensor Systems (2002-2005). Presently, he is running the largest funded ARC Research Network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) programme. He has published more than 370 refereed papers and a huge proportion of them appeared in prestigious IEEE Journals and Conferences. He was the associate editor of Journals/transactions including IEEE Transaction on Neural Networks. His research interests include SVMs, Sensors and Sensor Networks, Machine Learning, Neural Network, Pattern Recognition, Signal Processing and Control. He is the co-director of an active research centre, Centre of Expertise on Networked Decision & Sensor Systems, attracting grants from industry and defence agencies in Australia and USA.