Computer Networks 91 (2015) 376–389

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

SPAIS: A novel Self-checking Pollution Attackers Identification Scheme in network coding-based wireless mesh networks

Donghai Zhu [a], Xinyu Yang [a,*], Wei Yu [b]

[a] Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China
[b] Towson University, Towson, MD 21252, USA

ARTICLE INFO

Article history:
Received 5 December 2014
Revised 17 March 2015
Accepted 27 August 2015
Available online 8 September 2015

Keywords:
Network coding
Pollution attack
Wireless mesh network
Identification

ABSTRACT

Pollution attacks refer to ones where attackers modify and inject corrupted data packets into the wireless network with network coding to disrupt the decoding process. In the context of network coding, the epidemic effect of pollution attacks can degrade network throughput significantly because of the mixing nature of network coding. To address this issue, a number of schemes to identify malicious nodes have been developed in the past. Nonetheless, these schemes have their limitations and cannot effectively deal with pollution attacks. In this paper, we propose a novel light-weight Self-checking Pollution Attackers Identification Scheme (SPAIS), which can identify the pollution attackers effectively and efficiently. By making full use of the broadcast nature of wireless media and the insight that a well-behaved node can monitor its downstream neighboring nodes locally by cooperating with other nodes, SPAIS hierarchically organizes the network as levels such that the nodes in the same level can monitor their downstream level nodes cooperatively. Through the combination of theoretical analysis and extensive simulations, our experimental data demonstrates that SPAIS can more effectively identify pollution attackers with a lower cost in comparison with the existing candidate schemes. For example, even if the quality of network connections is not in good condition and the malicious nodes send only one corrupted packet, the pollution attackers can be identified with a high probability.

1. Introduction

Wireless mesh networks (WMNs) have become a promising technology for providing easy wireless Internet access because of highly self-organized and self-configured features of WMNs [1]. Nonetheless, the wireless links between the nodes are lossy due to the temporal and spatial fading of wireless channels, leading to a distributed WMN infrastructure and relatively low performance. To this end, network coding [2,3], which is a technique that integrates the coding and routing schemes, has been introduced to improve the performance of WMNs. Generally speaking, in contrast to the traditional "store and forward" method, network coding allows relay nodes to mix the information content in packets before forwarding them, which can improve the network throughput significantly. Network coding has been applied to various applications, including the content distribution [4], P2P networks [5], etc. Existing research efforts have shown the great advantage of network coding in wireless networks (e.g., the COPE [6], the MORE [3], etc.)

Network coding offers an appealing feature to construct networks. Nonetheless, because of the mixing nature, network coding is vulnerable to pollution attacks. Through pollution attacks, malicious nodes can alter or forge some corrupted messages and inject them into the network. The impact of pollution attacks can be devastating because with

* Corresponding author. Tel.: +86 13571956503; fax: +86 2982668098.
E-mail addresses: Dr.zhudonghai@stu.xjtu.edu.cn (D. Zhu), yxyphd@mail.xjtu.edu.cn (X. Yang), wyu@towson.edu (W. Yu).

http://dx.doi.org/10.1016/j.comnet.2015.08.023
1389-1286/

an epidemic propagation, the corrupted packets can pollute the whole network in a short time and degrade the network throughput dramatically [7].

In recent years, a number of authentication schemes have been developed to address pollution attacks against network coding [8–14]. In addition, a number of identification schemes [15–19] have been proposed to identify and locate the pollution attackers. In comparison with the authentication schemes, identification schemes are more proactive in defending pollution attacks. Because authentication schemes are all receivers-based, the detection may be too late as the attack may already consume the bandwidth and significantly disrupt network performance. If attackers can be quickly identified and immediately excluded from the network, the attack consequence can be largely mitigated.

Nonetheless, the existing identification schemes have their limitations and cannot be used to effectively deal with pollution attacks in WMNs where network coding is used. For example, Watchdog scheme [15] relies on a large number of extra trusted nodes to serve as watchdogs, which incurs a high cost to large networks. In MIS [16], a trusted server is responsible for computing the checksum of suspicious packets. The checksum transmission incurs a significant communication overhead. In [18], Algebraic-Watchdog scheme develops the hypothesis testing method to make the decision of identifying a malicious node. Nonetheless, that scheme needs to collect a large number of detecting samples and its computation cost is high. Similar to MIS [16], SpaceMac [19] relies on a trusted controller to identify the malicious nodes. In addition, SpaceMac and PAI [17] do not consider the lossy wireless environment, which leads to the limited applicability.

In this paper, we propose a novel light-weight Self-checking Pollution Attackers Identification Scheme (SPAIS) to effectively and efficiently identify the pollution attackers against network coding based legacy WMNs where MIMO (Multiple-Input and Multiple-Output) systems and directed antennas are not in use. Our proposed scheme has several novelties. First, unlike existing research efforts, SPAIS does not need any extra nodes to serve as watchdogs or any centralized server or controller. Second, it is self-checking and fully distributed. Third, it is light-weight and incurs only a little bandwidth overhead on the system. Moreover, network topology can be dynamic. If the identification is triggered and the topology has changed during data transmission, the source can re-determine the network level and re-distribute the random seeds to the corresponding nodes to ensure the secrecy of the random seeds. Our main contributions are summarized as follows:

- *Novel scheme.* In our design of SPAIS, we make full use of the broadcast nature of wireless media and adopt the principle that a well-behaved node can monitor its downstream neighbor nodes locally by cooperating with other nodes. To this end, our SPAIS hierarchically organizes the network as levels and the nodes in the same level can cooperatively monitor the nodes in the downstream level. To further reduce the bandwidth overhead, we develop a homomorphic hash function for network coding.
- *Theoretical analysis.* We conduct theoretical analysis and prove that our scheme is correct and secure. For example, under typical system settings, our analytical data shows that SPAIS allows only 1 out of 65,536 polluted packets passing the verification of the upstream level of the attacker if the network is lossless.
- *Evaluation.* We evaluate the effectiveness of SPAIS in terms of false positive probability and false negative probability through simulation experiments. The simulation data shows that even if in a wireless network where the average link quality is only 0.7 and the attacker sends out only one corrupted packet to launch a pollution attack, the attacker can be located with a probability of 95.62%. In addition, even if there are 30% of malicious nodes in the network, the probability that the malicious nodes can disparage a well-behaved node is only 1.45%. We also compare the computation overhead and bandwidth overhead with the existing schemes. The experimental data confirms that our scheme is more efficient than existing schemes.

The remainder of the paper is organized as follows: we introduce the network and threat models in Section 2. We present our proposed scheme in Section 3. We conduct the security analysis of our scheme in Section 4. In Section 5, we show both analytical and experimental results to validate the effectiveness of our proposed scheme in comparison with existing schemes. We discuss the limitation and the potential future work in Section 6. We review the related work in Section 7 and conclude the paper in Section 8, respectively.

2. System and threat models

In this section, we first introduce the system model and then present the threat model.

2.1. System model

In this paper, we consider a typical multicast scenario in WMNs modeled as a directed acyclic graph $G = (V, E)$, in which a source $\mathcal{S}$ sends a multicast stream of packets to a set of receivers $\mathcal{R}$. We focus on the backbones of legacy WMNs where MIMO systems and directed antennas are not in use, which consists of stationary mesh routers and mesh gateways. For more information on WMNs, please refer to [1]. We assume that all nodes in the network perform the generations-based random linear network coding that is well adopted and described in [20].

$\mathcal{S}$ generates a stream of packets, and divides them into generations. Each generation consists of m packets $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_m$, and $\underline{\mathbf{x}}_i$ is represented as a vector of $(\underline{x}_{i,1}, \underline{x}_{i,2}, \ldots, \underline{x}_{i,n})$ of finite field $\mathbb{F}_q^n$. Here, q is a prime of a proper size, and all the arithmetic operations are conducted through $\mathbb{F}_q$. In addition, $\mathcal{S}$ generates an augmented packet $\mathbf{x}_i$ by prefixing $\underline{\mathbf{x}}_i$ with the i^{th} unit vector of dimension m and $\mathbf{x}_i$ is denoted as

$$\mathbf{x}_i = \left(\overbrace{\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0}^{m}, \underline{x}_{i,1}, \underline{x}_{i,2}, \ldots, \underline{x}_{i,n} \right). \quad (1)$$

For the random linear network coding, $\mathcal{S}$ sends the linear combination of packets $\mathbf{x}_i$ and the coefficients are randomly selected from the finite field $\mathbb{F}_q$. That is to say, for packets $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ and random coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$, an