Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model[☆]

Wenjuan Li[a], Weizhi Meng[b,*], Lam-For Kwok[a], Horace H.S. IP[a]

[a] *Department of Computer Science, City University of Hong Kong, Hong Kong SAR*
[b] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark*

ARTICLE INFO

Keywords:
Intrusion detection
Collaborative network
Intrusion sensitivity
Machine learning
Trust management model
Insider attacks

ABSTRACT

To defend against complex attacks, collaborative intrusion detection networks (CIDNs) have been developed to enhance the detection accuracy, which enable an IDS to collect information and learn experience from others. However, this kind of networks is vulnerable to malicious nodes which are utilized by insider attacks (e.g., betrayal attacks). In our previous research, we developed a notion of intrusion sensitivity and identified that it can help improve the detection of insider attacks, whereas it is still a challenge for these nodes to automatically assign the values. In this article, we therefore aim to design an intrusion sensitivity-based trust management model that allows each IDS to evaluate the trustworthiness of others by considering their detection sensitivities, and further develop a supervised approach, which employs machine learning techniques to automatically assign the values of intrusion sensitivity based on expert knowledge. In the evaluation, we compare the performance of three different supervised classifiers in assigning sensitivity values and investigate our trust model under different attack scenarios and in a real wireless sensor network. Experimental results indicate that our trust model can enhance the detection accuracy of malicious nodes and achieve better performance as compared with similar models.

1. Introduction

Nowadays, intrusion detection systems (IDSs) have been widely implemented in many networks aiming to defend against a variety of attacks (Ghosh et al., 1998; Vigna and Kemmerer, 1998), and they have already become an essential component for current defense infrastructure (Scarfone and Mell, 2007). However, network intrusions have become much more sophisticated and hard to detect (Vasilomanolakis et al., 2015). To resolve this issue, IDS collaboration is considered as an effective way to enhance the detection capability of a single IDS.

Motivated by this, collaborative intrusion detection networks (CIDNs) have been developed, with the purpose of strengthening a single IDS by collecting knowledge and learning experience from other IDS nodes. This CIDN is expected to enhance the overall detection accuracy of intrusion assessment and improve the possibility of identifying novel attacks (Wu et al., 2003). However, insider attackers can compromise some peers (or *IDS nodes*) within the CIDN and utilize these compromised nodes to invade or threaten the whole collaborative network. For example, these malicious peers can make use of some insider attacks, such as Sybil attacks, newcomer attacks and betrayal attacks, to degrade the effectiveness and efficiency of CIDNs by sending out false information and continuously compromising other honest IDS nodes. In these cases, designing a robust CIDN model (i.e., effectively evaluating the trustworthiness of each node within the network) becomes very crucial and essential to protect this kind of networks against insider attacks while maintaining the detection capability.

In our previous work (Li et al., 2013), we identified that each IDS has different levels of sensitivity in detecting particular intrusions and thus proposed a notion of *intrusion sensitivity* that can be described as below:

- *Intrusion sensitivity* describes different levels of detection capability (or accuracy) for IDS nodes in detecting particular kinds of attacks or anomalies. Let I_s denote the detection sensitivity of a node and t denote a time period. For two IDS nodes A and B, we can say $I_s^A > I_s^B$

[☆] A preliminary version of this paper appears in Proceedings of the 8th IFIP WG 11.11 International Conference on Trust Management (IFIP TM 2014), pp. 61–76, 2014 (Li et al., 2014).

[*] Corresponding author.
E-mail addresses: wenjuan.li@my.cityu.edu.hk (W. Li), yuxin.meng@my.cityu.edu.hk (W. Meng), cslfkwok@cityu.edu.hk (L.-F. Kwok), cship@cityu.edu.hk (H.H.S. IP).

http://dx.doi.org/10.1016/j.jnca.2016.09.014
Received 30 June 2016; Received in revised form 8 September 2016; Accepted 30 September 2016
Available online 13 October 2016

if A has a stronger detection capability than B within this time period.

The previous work also proves that intrusion sensitivity is feasible in CIDNs, but still many challenges remain. One of the challenges is how to automatically assign the values of *intrusion sensitivity*.

Contributions: In order to address the above mentioned challenge, in this article, we design a supervised intrusion sensitivity-based trust management model to improve the robustness of CIDNs and propose an approach of automatically assigning the values of *intrusion sensitivity* by means of supervised machine learning classifiers. "Supervised" here means that the values of intrusion sensitivity can be allocated using supervised machine learning. Our contributions of this work can be summarized as below:

- We adopt a proper CIDN framework from the literature (Fung et al., 2008, 2009) and revise it to fit our model. The revised framework includes five major components such as IDS nodes, a trust management component, a collaboration component, a communication component and a query component. We then introduce how to integrate the notion of *intrusion sensitivity* in our model, which measures the detection sensitivity of an IDS node.
- We design a *supervised intrusion sensitivity-based trust management model* for CIDNs and explain how to compute trust values of different nodes. In order to automatically allocate the values of *intrusion sensitivity*, we propose a supervised approach which can employ supervised machine learning to automatically assign the sensitivity level for each IDS node.
- In the evaluation, we compare the performance of three different supervised learning classifiers in assigning the values of sensitivity. Afterwards, we simulate a CIDN and launch certain attacks to investigate the performance of our proposed trust model under different attack scenarios. In addition, we further evaluate our model in a real wireless sensor network to explore its practical performance. Experimental results indicate that our proposed trust model is more efficient and sensitive in detecting malicious nodes as compared to other similar trust models.

The remaining of this article is organized as follows. In Section 2, we review some related research proposals regarding building trust models in collaborative networks. In Section 3, we describe the details of our revised CIDN framework, introduce how to compute trust values of nodes and how to automatically assign *intrusion sensitivity*. In Section 4, we describe experimental settings and analyze experimental results. We further provide a discussion in Section 5 and conclude our work in Section 6.

2. Related work

Traditionally, an isolated (or single) intrusion detection system has no information about the whole protected environment; thus, it is more likely to be bypassed by novel and complex intrusions. To resolve this issue, collaborative intrusion detection networks (CIDNs) (Wu et al., 2003) have been proposed and implemented, which enable an IDS node to achieve more accurate detection by collecting and learning useful information from other IDS nodes. However, insider attacks like betrayal attacks are a big challenge in real-world applications for such collaborative network.

Distributed IDS systems: There are many kinds of distributed IDS systems. Janakiraman and Zhang (2003) proposed *Indra*, a distributed scheme based on sharing information between trusted peers in a network to guard a peer-to-peer network as a whole against intrusion attempts. Li et al. (2006) identified that most distributed intrusion detection systems (DIDS) relied on centralized fusion, or distributed fusion with unscalable communication mechanisms, and then proposed a distributed system based on the emerging decentralized location and routing infrastructure. The experimental results showed that their methods could greatly outperform the traditional hierarchical approach when facing large amounts of diverse intrusion alerts. However, these approaches assume that all peers are trusted, which is vulnerable to insider attacks (i.e., some nodes become malicious). The distributed intrusion detection systems can be roughly classified as follows: (1) *Centralized/Hierarchical systems*: Emerald (Porras and Neumann, 1997) and DIDS (Snapp et al., 1991); (2) *Publish/subscribe systems*: COSSACK (Papadopoulos et al., 2003) and DOMINO (Yegneswaran et al., 2004); and (3) *P2P Querying based systems*: Netbait (Chun et al., 2003) and PIER (Huebsch et al., 2005).

Trust model development: To mitigate the impact of insider attacks, several trust models have been proposed in the literature. For example, Duma et al. (2006) proposed a P2P-based overlay for intrusion detection (Overlay IDS) that mitigated the insider threat by using a trust-aware engine for correlating alerts and an adaptive scheme for managing trust. The trust-aware correlation engine is capable of filtering out warnings sent by untrusted or low quality peers, while the adaptive trust management scheme uses past experiences of peers to predict their trustworthiness. However, a major issue is that the past experience of a peer has the same impact regardless of the age of its experience.

To resolve this problem, Fung et al. (2008) proposed a Host-based IDS (HIDS) collaboration framework that enables each HIDS to evaluate the trustworthiness of others based on its own experience by means of a forgetting factor. The forgetting factor can give more emphasis on the recent experience of the peer. Later, Fung et al. (2009) improved their proposed trust management model by using a Dirichlet-based model to measure the level of trustworthiness among IDS nodes according to their mutual experience. This model had strong scalability properties and was robust against common insider threats and the experimental results demonstrated that the new model could improve robustness and efficiency. As the mechanism of feedback aggregation is a key component in the above trust model, Fung et al. (2010) further applied a Bayesian approach to feedback aggregation to minimize the combined costs of missed detection and false alarm. Their experiments indicated that the Bayesian approach could make an improvement in the true positive detection rate and a reduction in the average cost.

In addition, Quercia et al. (2006) proposed a distributed trust-based framework that satisfied a broader range of properties, which evolved an expressive and tractable trust calculation based on Bayesian formalization, protected user anonymity and integrated a risk-aware decision module. Then, Li et al. (2008) proposed an objective trust management framework (*OTMF*) using a modified Bayesian approach where the trust in the provider of second-hand information is considered when evaluating trust. They further conducted a performance evaluation and security analysis on *OTMF*, and the results showed that the *OTMF* was more effective and robust as compared to similar frameworks.

Many theories have also been investigated to evaluate the trustworthiness of communication entities such as Information Theory, Game theory and Grey Theory. For example, Sun et al. (2006) presented an information theoretic framework to quantitatively measure trust and model trust propagation in Ad Hoc networks. In their framework, trust is a measure of uncertainty with its value represented by entropy. Similarly, Tuan (2006) used the game theory to model and analyze the processes of reporting and exclusion in a P2P network. They found that if a reputation system was not incentive compatible, the more numbers of peers in the system, the less likely that anyone will report about a malicious peer.

Recently, Andreolini et al. (2015) identified mobility-based evasion attacks, where an attacker splits a malicious payload in such a way that no part can be recognized by existing defensive mechanisms and proposed a cooperative framework for intrusion detection. Several other related studies on collaborative networks can be referred to Bao et al. (2012), Cai et al. (2009), Kantzavelou et al. (2013), Liu et al.