Decision Support Systems 67 (2014) 40–52

Contents lists available at ScienceDirect

Decision Support Systems

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

An approach to finding the cost-effective immunization targets for information assurance

Guannan Liu [a], Jin Zhang [b,*], Guoqing Chen [a]

[a] Department of Management Science and Engineering, School of Economics and Management, Tsinghua University, Beijing 100084, China
[b] Department of Management Science and Engineering, School of Business, Renmin University of China, Beijing 100872, China

ARTICLE INFO

Article history:
Received 10 February 2013
Received in revised form 3 August 2014
Accepted 7 August 2014
Available online 19 August 2014

Keywords:
Information assurance
Network immunization
Savability
Cost-effective immunization targets (CEIT)

ABSTRACT

Information assurance is increasing in importance as threats abound in the highly connected world of e-business. For enterprises, the goal is to achieve a secure information environment in a cost-effective manner. This paper focuses on the issue of how to cost-effectively immunize an enterprise's network to prevent threats (e.g., virus, rumor) from invading and spreading. An approach, namely Cost-Effective Immunization Targets (CEIT) is proposed as a means to identify the cost-effective immunization targets and provide direct cost/benefit trade-off solutions for practitioners. In the approach, a novel concept, savability, is introduced as an extension of return on security investment (ROSI), with the reduced expected infection probability as mitigated risks through immunization. Meanwhile, a bond percolation process, which can be done in just a single graph traversal, is incorporated to simplify the estimation of expected infection probability in place of repeated diffusion simulations. Theoretical analysis proves that the proposed approach can approximate the optimal solutions within a definite lower bound. Finally, experiments on real-world information network datasets reveal that the algorithm CEIT outperforms other immunization strategies in both homogeneous and heterogeneous cost cases. Further, a case study indicates that the CEIT-identified immunization targets are more likely to 'save' the important nodes with high potential infection loss, avoiding redundant immunization.

1. Introduction

With the rapid development of information technology (IT) and Internet applications, many traditional stand-alone information systems (IS) have been replaced by or extended to Web-based systems. In today's e-business environment, massive data communication and business transactions flow across the Internet. Hence, most industries and business sectors are highly connected, ranging from manufacturing and financial services, to retail, healthcare and so on. The connections have been further strengthened with new features such as mobility, virtualization, personalization, social networks, and rich-media data.

In the context of this connected world, a system can be viewed and treated as either a physical or a virtual network. Particularly, from the perspective of a company, all transactions are executed and transferred online, along with staff communications via email or other web applications. The company can then be treated as a connected information network, therefore, understanding how information flows and how to work with it for managerial effectiveness in an information network has recently become a focal point. The viewpoints of the studies on information flow in information networks can be two-fold. One is from a positive-effect perspective in light of information diffusion. That is, how information spreads across the network, what the centric nodes are in influencing others (e.g., opinion leaders), and their impacts (e.g., on social business, communication, advertising) could be among the major issues of concern [1,2]. The other viewpoint is from a negative-effect perspective in light of information assurance. The major issues of concern include how a threat propagates within a network (e.g., virus infection, rumor spreading), what the most risky nodes are in causing severe harm, and how to limit their impacts (e.g., on business continuity) [3–5]. Notably, these two perspectives call for addressing a common issue that relates to the mechanism of information flow in a network, along with other issues relating to specific problems and contextual features.

As far as information assurance is concerned, while a network facilitates communications and enables online transactions as normal traffic, it also faces challenges in being attacked or infected by computer viruses, breaches, malware, etc. Meanwhile, rumors, gossip or unverified information can also propagate in the network through the informational routes. It has been reported that information security breaches increased nearly 50% [6], and according to a recent survey by PwC, 93% of large organizations have suffered from breaches in 2011 [7]. Considering breaches, one of the commonly encountered threats currently, they may cause incidents such as system breakdowns, denial of services

* Corresponding author.

http://dx.doi.org/10.1016/j.dss.2014.08.002
0167-9236/

(DoS), data/infrastructure inaccessibility etc., leading to a remarkable amount of organizational expenses. For example, viruses such as ILOVEYOU, Code Red, SQL slammer, Sasser, etc. gave rise to business costs on the order of billions of dollars [8]. ICSA Labs reported that more than 83% of companies had suffered from a business loss between 10,000 and 1,000,000 dollars [9]. With respect to rumors in the workplace, they can also potentially damage reputations and affect the productivity of an organization [10]. Generally, it is apparent that because those online business and Web 2.0-based companies are more connective in operational activities and communication channels, information spreads much faster in the networks than in traditional ways. As a result, the threats and risks they face are deemed ever greater.

Having become more and more aware of the high risks of negative-effect information, many companies are investing an increasing amount of money and resources in the hope of working with a safer network environment to guarantee normal operations. Among the many measures and strategies being implemented, immunization is a common and effective way to prevent malicious information from spreading. The IT/IS management of a company is usually responsible for the network and information assurance, thus it would generally patch the computers regularly or pre-install anti-virus software on some computers. Moreover, the IT/IS management also wants to target some working units to take measures immediately when fake/malicious information is likely to propagate within the network. However, companies are usually constrained by limited resources for information assurance. According to a survey of CIOs, the annual budget specifically for information security comprised only 10% of their total IT budgets, although this amount had increased in recent years [11]. Generally speaking, because immunization can be rather costly, it is difficult for managers to make effective decisions on reducing expected loss from threats and incidents based on limited resources [12]. Therefore, a strategy for identifying the cost-effective immunization targets is badly needed.

In recent years, research efforts have been made to cope with related problems. Some researchers discussed the problem from an economic view, attempting to ascertain the optimal decisions on how much to invest in information assurance from a macro perspective [13–16]. Others devised various immunization strategies, in which the static network structure was a focal point in choosing immunization targets [17–20]. In this paper, we take a combined view in our investigation. That is, both the network connectivity and the dynamic diffusion process, along with immunization cost and potential loss related to information assurance incidents, are incorporated in seeking the immunization targets. Therefore, we introduce a concept, namely, savability, to evaluate the economic efficiency of immunizing any network node, taking both homogeneous and heterogeneous cases into account.

In concrete terms, the main ideas and contributions of our proposed approach could be described as follows:

1. Finding the cost-effective immunization targets is a novel problem that has emerged from the information assurance practices in companies. In contrast to the previous literature that focuses on how much to invest in information assurance, the formulated problem aims to decide whether a single information unit is worth being immunized from a micro level, taking both the network structure and the dynamic diffusion process into consideration. The obtained immunization set can help information assurance practitioners allocate the limited budget more wisely.
2. To solve the cost-effective immunization problem, a greedy method (i.e., algorithm CEIT) is developed because the problem is NP-hard. A novel concept, savability, is introduced in the method to measure the return on investment of immunizing a node, with both immunization cost and infection loss considered. As for the core part of CEIT, i.e., estimating the expected infection probability, we innovatively apply a bond percolation process to replace repeated independent cascade diffusion simulations, so that the estimation of expected infection probability is simplified and the time complexity is greatly reduced.
3. Theoretical analysis has proved that the proposed algorithm is guaranteed to approximate the optimal solutions within a definite lower bound. Additionally, experiments on real-world information network datasets have demonstrated that the proposed algorithm achieves greater expected reduced infection loss than other immunization strategies.

In the rest of the paper, Section 2 reviews the literature related to our work, Section 3 formulates the problem and introduces the diffusion and cost models, Section 4 proposes the Cost-Effective Immunization Targets (CEIT) algorithm, Section 5 presents the experimental results, and Section 6 applies the proposed algorithm to a real information assurance practice. Finally, the study is concluded in Section 7.

2. Related work

Recent years have witnessed a substantial amount of research on computer virus and rumor spreading. Many of these studies model the phenomenon in the spirit of epidemiology for investigating the propagation patterns and prevalence threshold in various types of networks and contexts [21–25]. It has been found that, for a network, immunization is an effective way to lower the epidemic threshold [26]. Thus, immunization can help prevent computer virus and rumors from prevailing and lower the final infection size. Considerable effort had been devoted to identifying better immunization strategies. For instance, Target immunization aims to immunize the highly connected nodes, and it has been demonstrated that it outperformed other strategies on scale free networks [17]. Holme et al. [27] presented betweenness strategy in dealing with complex networks, which centered on the betweenness centrality of edges or nodes. Acquaintance immunization, in which random neighbors of a randomly selected node are immunized without requirements for global information of the network [18], is also an effective immunization strategy. Subsequently, further research has been conducted regarding improvements in immunization strategies, resulting in a variety of extensions [28,19,20]. In brief, these research attempts barely employ static network structural information to choose the immunization targets.

The immunization resource is usually limited for an organization [12]. Therefore, how to optimally allocate the limited resources within the network remains to be an important issue. There exist some studies on the economics of information security. Sonnenreich [13] presented a quantitative model to calculate return on security investment (ROSI), which can be used to measure the cost-efficiency of security investment. Anderson et al. discussed the economics of information security from a broad perspective [14], followed by a variety of concrete economic models concerning the cost and loss in information security investment. Gorden and Loeb employed a cost-benefit analysis to determine the optimal level of resources to devote to securing information [15]. Lee et al. [16] presented an economic model from the benefit of a customer, and they derived the optimal investment decisions for firms. Kleczkowski et al. minimized the total cost of treatment and prevention by performing simulations, and derived an optimal control size [29]. Although the economics of information assurance is discussed in these studies, these models are established generally without considering the specific network structure where the nodes are usually assumed to be homogeneous and treated equally. The problems they solved may only help practitioners to decide how much they should invest in information assurance from a global perspective, while the question of whether an information unit or a node within the network is worth being immunized remains unanswered. In contrast, our proposed approach combines the cost analysis with the connectivity of each node, as well as the dynamic diffusion process, to help companies make a cost-effective decision on choosing immunization targets.

Finally, it is worth mentioning that this research problem is also related to the influence maximization problem, which is oriented to seed