Journal of Network and Computer Applications 36 (2013) 1387–1394

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Propagation modeling of active P2P worms based on ternary matrix[☆]

Ting Chen, Xiao-song Zhang*, Hong-yuan Li, Dong Wang, Yue Wu

School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

ARTICLE INFO

Article history:
Received 13 September 2012
Received in revised form
1 January 2013
Accepted 18 February 2013
Available online 1 March 2013

Keywords:
Active P2P worms
Propagation modeling
Ternary matrix
Time lags
Expressiveness
Flexibility
Ease of derivation and deployment

ABSTRACT

Propagation modeling of worms has become an attractive research field in recent years since it facilitates worm prediction, detection, analysis and prevention etc. In this work, we propose a novel ternary-matrix-based model to describe the propagation trend of active P2P worms. Compared to existing logic-matrix-based models, our model takes the time lags into consideration by introducing new states and special matrix operations. Our model is easy of derivation and deployment because it confines derivation process to pure matrix operations. Moreover, two other advantages of our model are fully explored. One is expressiveness: (1) practical P2P topology can be modeled in the topology matrix; (2) the state of any node can be identified at any time; (3) and the attack path of any node can be backtracked in linear time. Flexibility is the other merit: our model can adapt to different scenes by changing the related parameters, particularly our model is general for different kinds of time lags and P2P topologies.

1. Introduction

P2P worms which are defined as a kind of special worm spreading by P2P topology (Khiat et al., 2006) have posed a severe threat to network facilities and common users because of the numerous vulnerable P2P software and the unawareness of network security of P2P users (Khiat et al., 2006; Zhou et al., 2005). To understand, predict, detect and contain P2P worms, it is meaningful to model the propagation trend of P2P worms.

In terms of spreading strategy, P2P worms are usually classified into two categories: active and passive (Khiat et al., 2006). To be more specific, active P2P worms propagate by attacking the nodes in the neighbor lists actively (Yu et al., 2008) while passive ones camouflage them as useful resources and trick other users to download them (Thommes and Coates, 2006). In this paper, we only research the propagation modeling of active P2P worms.

The research of the propagation modeling of active P2P worms has lasted for several years, so a number of models have been proposed. (As this work is not a survey, we did not review the models of non-active-P2P worms here.) According to modeling approaches, any model belongs to either continuous-time class or discrete-time class (Fan and Xiang, 2010a,b,c). Particularly, continuous-time models (Luo et al., 2009, 2010; Yang et al., 2011) consider worm propagation as a continuous process, so differential equation is a natural approach for derivation. On the contrary, discrete-time models (Zhang et al., 2010, 2009; Chen et al., 2011) treat worm propagation as a discrete process, so they are always derived through difference equations.

The logic-matrix-based models (Fan and Xiang, 2010a,b,c; Fan et al., 2011; Jia et al., 2010) which were firstly proposed in 2010 are most similar with our model, so we summarize the formers here. The core idea of the logic-matrix-based models is (Fan and Xiang, 2010a,b,c): P2P topology with n nodes is represented by a n by n square matrix T and its each element t_{ij} (in the i-th row and the j-th column) is a logic constant '*T*' or '*F*' to indicate whether there is a directed link from node i to node j or not. The states of all nodes are represented by a state logic vector S of length n and its each element s_i is a logic constant '*T*' or '*F*' to indicate whether node i is infectious or not. The state logic vector at time $g+1$ is derived by the equation $S_{g+1}=S_g+S_g \times T$ (assuming all nodes are vulnerable). Note that the operations '+' and '×' in the equation are defined as special matrix operations in reference Fan and Xiang (2010a,b,c).

As stated in Fan and Xiang (2010a,b,c), logic-matrix-based models neglect the time lags of worm propagation. Put it another way, the time lags from sending the worm packets, to receiving the worm packets, to having the recipient nodes infected, to the infected nodes becoming infectious are not considered (Fan and Xiang, 2010a,b,c). That is because any node which has been intruded by a worm but has not yet become infectious stays in a state cannot be denoted by a logic constant. In other words, the logic-matrix-models are not expressive enough to contain more

[☆] This work was supported by the Fundamental Research Funds for the Central Universities E022050205.

* Corresponding author. Tel.: +86 13438865431.
E-mail addresses: t.chen.ycg@gmail.com, chenting19870201@163.com (T. Chen).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.02.032

than two different states. However, common sense suggests that nodes cannot turn from intruded state to infectious state immediately except for a few 'toy' worms of minimal functions.

In this paper, we propose a ternary-matrix-based approach to model the propagation of active P2P worms which takes the time lags into consideration. To derive our model, new states and special matrix operations are introduced. Our model is easy of derivation and deployment because its derivation process is strictly confined to pure matrix operations. Our model has two other advantages: expressiveness and flexibility. Expressiveness represents that in our model (1) practical P2P topology can be modeled; (2) the state of any node at any time can be identified; (3) and the attack path of any node can be backtracked in linear time. Flexibility indicates that our model can adapt to various scenes through changing the related parameters. Particularly, our model is flexible to different kinds of time lags and P2P topologies which highlight our model even further.

This paper is organized as follows: Section 2 overviews the related work and then distinguishes our model from existing ones; Section 3 presents the derivation of our model; Section 4 shows the experiments and validates the advantages of our model; Section 5 concludes.

2. Related work

In recent year, research work for the propagation modeling of active P2P worms has been carried out thus a number of models has been proposed. In this section, we compare our model with existing ones in three major aspects.

2.1. Time lags

A great deal of research about the propagation modeling of active P2P worms (Yu et al., 2008; Luo et al., 2009, 2010; Yang et al., 2011; Zhang et al., 2010, 2009; Chen et al., 2011; Fan and Xiang, 2009, 2010a,b,c; Fan et al., 2011; Jia et al., 2010; Saadat et al., 2009; Yang et al., 2010; Liu, 2009; Zhang et al., 2008; Yu et al., 2005a,b; Feng et al., 2008a,b) neglects time lags, while mainstream research concerning the propagation modeling of passive P2P worms (Thommes and Coates, 2006, 2005; Feng et al., 2008a,b; Ramachandran and Sikdar, 2006) considers time lags. A potential reason for the difference lies in their different modeling approaches. Precisely speaking, mainstream propagation models of active P2P worms are derived by discrete difference equations, on the contrary, the propagation models of passive P2P worms are usually derived through continuous differential equations.

So far we have found only two articles (Hatahet et al., 2008; Yao et al., 2006) about the propagation modeling of active P2P worms consider the transmission lag of worm body. The two models assume the transmission lag of each node is the same and fixed. However, we believe the transmission lag of each node in practice has something to do with their software & hardware abilities as well as network environment.

Our model takes time lags into consideration which makes it obviously different from the models proposed in (Yu et al., 2008, 2005a,b; Luo et al., 2009, 2010; Yang et al., 2011; Zhang et al., 2010, 2009; Chen et al., 2011; Fan and Xiang, 2010a,b,c, 2009; Fan et al., 2011; Jia et al., 2010; Saadat et al., 2009; Yang et al., 2010; Liu, 2009; Zhang et al., 2008; Feng et al., 2008a,b). Moreover, our model is flexible to different time lags, i.e., the time lag of each node could be varied, so our model differs from the two models in (Hatahet et al., 2008; Yao et al., 2006).

2.2. Fine-grained state information & attack path

All current propagation models of active P2P worms (Yu et al., 2008; Luo et al., 2009, 2010; Yang et al., 2011; Zhang et al., 2010, 2009; Chen et al., 2011; Saadat et al., 2009; Fan and Xiang, 2009; Yang et al., 2010; Liu, 2009; Zhang et al., 2008; Yu et al., 2005a,b; Feng et al., 2008a,b; Hatahet et al., 2008; Yao et al., 2006) except the logic-matrix-based ones (Fan and Xiang, 2009, 2010a,b,c; Fan et al., 2011; Jia et al., 2010) can only provide coarse-grained state information in derivation process. In other words, without test-beds, real word experiments or simulations, most models can obtain the number of nodes in each state only, but they cannot identify the states of specific nodes.

Our model can provide fine-grained state information because the information is actually stored in a special state vector. Furthermore, our model can find the attack path of any node in linear time with the help of a special vector. We have to note that although the logic-matrix-based models have potential to obtain fine-grained state information and attack path, this topic is not discussed in current articles (Fan and Xiang, 2009, 2010a,b,c; Fan et al., 2011; Jia et al., 2010).

2.3. P2P topology

P2P topology is overly simplified in most models. For example, the models in Yu et al. (2008), Fan and Xiang (2009, 2010a,b,c), Zhang et al. (2010, 2009), Chen et al. (2011), Fan et al. (2011), Zhang et al. (2008), Feng et al. (2008a,b) and Yu et al. (2005a,b) simplify unstructured P2P topologies as any topologies with power-law property. Besides, in models (Yu et al., 2008, 2005a,b; Yang et al., 2011), structured P2P networks are simplified as any topologies with equal degree of each node. As a distinguishing feature, our model is able to model the true P2P topology in practice by introducing a topology matrix which stores topology information.

Our model can adapt to different kinds of P2P topologies by simply replacing the topology matrix. However, most existing models can apply to one specific P2P topology. For instance, the models in (Zhang et al., 2009, 2010; Chen et al., 2011; Fan et al., 2011; Zhang et al., 2008; Feng et al., 2008a,b) are specialized in unstructured P2P topology; the models in (Yang et al., 2011; Yu et al., 2005a,b) are applied to structured P2P topology only; research work in (Saadat et al., 2009) and (Hatahet et al., 2008) aim at hierarchical and BitTorrent topologies respectively. One point should be mentioned here is that the logic-matrix-based models can also adapt to different topologies, however, this property is not fully explored in related articles (Fan and Xiang, 2009, 2010a,b,c; Fan et al., 2011; Jia et al., 2010).

3. Ternary-matrix-based model

In this section, we describe our ternary-matrix-based modeling approach through a simple example. At first, some necessary definitions are presented; then we define several special matrix operations; followed is the derivation process; then we pay some attention to the stability state of worm propagation; finally we present the method to obtain attack path in linear time. Fig. 1 shows a P2P topology which is used as an example throughout this section.

3.1. Definitions

The number of nodes in the P2P network is denoted by an integer n. In the example, n equals 4. According to the directed graph theory, a P2P topology consisting of n nodes can be