Modeling Social Engineering Botnet Dynamics across Multiple Social Networks*

Shuhao Li[1,2,3], Xiaochun Yun[1,2,**], Zhiyu Hao[1], Yongzheng Zhang[1], Xiang Cui[2,3], and Yipeng Wang[1,3]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
yunxiaochun@iie.ac.cn
[3] Graduate University of Chinese Academy of Sciences, Beijing, China

Abstract. In recent years, widely spreading botnets in social networks are becoming a major security threat to both social networking services and the privacy of their users. In order to have a better understanding of the dynamics of these botnets, defenders should model the process of their propagation. However, previous studies on botnet propagation model have tended to focus solely on characterizing the vulnerability propagation on one infection domain, and left two key properties (cross-domain mobility and user dynamics) untouched. In this paper, we formalize a new propagation model to reveal the general infection process of social engineering botnets in multiple social networks. This proposed model is based on stochastic process, and investigates two important factors involved in botnet propagation: (i)bot spreading across multiple domains, and (ii)user behaviors in social networks. Furthermore, with statistical data obtained from four real-world social networks, a botnet simulation platform is built based on OMNeT++ to test the validity of our model. The experimental results indicate that our model can accurately predict the infection process of these new advanced botnets with less than 5% deviation.

Keywords: network security, social network, social engineering attack, botnet, propagation model.

1 Introduction

Botnets, which are considered to be the platforms of cyberattacks, have the ability to send spams, launch DDoS attacks and steal sensitive information. They have become a growing risk with the potential to severely impact important infrastructure components like communication systems. In recent years specifically, the rapid development of social networking services (SNS) [1] and the diversification of social engineering attacks (SEA) [2] demonstrate that new widely spreading botnets have become an emerging threat to social networking services

* This work was supported by "The National High Technology Research and Development Program of China"(863 programs: No.2007AA010501) and "The National Natural Science Foundation of China" (No.60703021, No.61003261 and No.61070185).
** Corresponding author.

D. Gritzalis, S. Furnell, and M. Theoharidou (Eds.): SEC 2012, IFIP AICT 376, pp. 261–272, 2012.

and personal privacy in social networks (SN). Comparing to traditional botnets, the evolved variety have the ability to not only steal and abuse social network users' digital identities, but also to disseminate a large number of advertising messages in social networks. In addition, they can also employ SEA to spread bots across multiple social networks. Therefore, the infection scope is significantly expanded by these new botnets. For example, the Koobface botnet [3], which was initially launched in the scope of Facebook and MySapce, is one of the most popular botnets in social networks. Up to now, it has generated 56 variants and infected more social networks than originally anticipated ones.

We believe that new advanced SEA botnets (similar to Koobface, but more destructive) will emerge in large numbers, resulting in wide-scale damage to social networks. This belief is based on the following observations:

1) These botnets have the ability to capture bots by using SEA, which is much easier than other intrusion means. SEA greatly reduces attacking difficulty for hackers, although it requires the participation of users to complete the injection of bot programs.

2) With the characteristic of cross-domain propagation, the botnets are able to disseminate bot programs from one social network to another, which makes the infection rate of the bot programs much wider and faster.

In order to know more about the evolution and dynamics of botnets and how to mitigate them effectively, defenders should first be able to measure and predict the size of botnets. Therefore, modeling botnet propagation has become an important research direction. In this paper, we propose a propagation model based on stochastic process, to describe the complex infection process of new botnets in social networks. In order to develop a more accurate description than those in previous attempts, both the properties of botnet cross-domain spreading and user dynamics (the variation of active users and users' responses) in social networks are investigated in our model. We believe the proposed model in this paper could be a practical and effective tool for social network administrators and botnet defenders, who are currently struggling against such botnets (*e.g.*, Koobface).

Our contributions are summarized as follows:

1) We have formalized a new propagation model to predict the general infection process of new social engineering botnets in social networks based on stochastic process. Two factors (the cross-domain spreading and user dynamics) found to significantly influence the infection of botnets, are well investigated in our proposed model. It was able to predict the infection process of these new botnets with a small deviation.

2) We have provided a botnet simulation platform based on OMNeT++ [15] with the statistical data obtained from real-world social networks. It was applied to simulate and evaluate the propagation process of these new botnets. In the proposed platform, multiple behaviors of social network users are simulated with a guarantee of efficacy, which is considered to be superior to other botnet simulators [5] in the ability of generalization.

The remainder of this paper is organized as follows: Some related work is discussed in Section 2; and Section 3 is dedicated to the background on these new botnets; the proposed propagation model is introduced in Section 4; the simulation evaluation and experimental results are provided in Section 5; several limitations of our work are discussed in Section 6. Finally, we conclude the paper while highlighting the scope of future work.

2 Related Work

Several propositions in the literature exist for modeling malware propagation. In [11], Zou *et al.* presented an email worm model, which revealed the significance of user behaviors by considering email checking time and the probability of opening email attachments. In [12], Yan *et al.* identified some parameters that are related to malware propagation in social networks. In [13], Cheng *et al.* proposed an equation-based model to analyze the mixed behaviors of delocalized and ripple based propagation, which is related to hybrid malware in generalized social networks. According to our knowledge, most of previous work introduced the propagation models based on a single infection domain, leaving cross-domain spreading and user dynamics unconsidered. However, there are a few exceptions. For example, Dagon *et al.* [14] used time zones to model the propagation process of some botnets, arguing that victims turn their computers off at night, which leads to diurnal properties in botnet activity. But their study represents only one factor involved in user dynamics.

3 Scene Analysis

3.1 Background

An SNS is an online service, platform, or site which focuses on the building and reflecting of social networks or social relations among people. There are many popular social networks (*e.g.*, Facebook, Twitter, Gmail and Tencent QQ) with hundreds of millions of registered users connected through friendship links. The friendship link is generally a binary relationship, representing two users as friends. Usually, social network users connect, express themselves, or share information with social network messages. These messages may have different forms in different social networks, but typically have a similar function. Meanwhile, social network messages can be also exploited for the purpose of spreading malwares by attackers, especially social engineering hackers. These malwares can launch SEA, which relies on social disguises, cultural ploys and psychological tricks to persuade computer users in assisting hackers with their illegal intrusion or use of computer systems and networks [2].

To demonstrate the efficacy of our proposed model, we deployed several SNS crawlers to capture the statistical data of social network users' behaviors, which includes access time, online duration, and the interval time prior to users' responses. Finally, based on the social network generation algorithm in [4], we built a simulation platform based on multiple virtual social network domains, which allows us to evaluate our propagation model in a controlled environment.

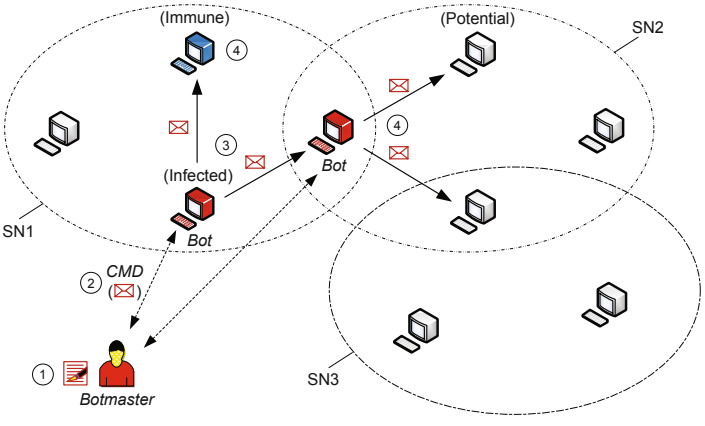

Fig. 1. The general process of SS-botnets' propagation

In our experiment, we designed the user model according to real statistics obtained from our SNS crawlers, and then simulated the whole infection process of these new botnets. Our experimental results show that our model can yield an accurate prediction on the size of the new botnets, on the condition of the topology information obtained from related social networks and the information of the initial infected nodes detected in the social networks.

3.2 The SS-Botnet Propagation Process

To more clearly describe this issue, we first provide two key definitions:

Definition 1. *A* ***trap message*** *is a social network message which is constructed with either bot programs already embedded within, or with a means of accessing them (*e.g.*, download URL). This message is forged with a social engineering approach, and designed to appear attractive to victims.*

Definition 2. *The botnet making use of trap messages for the purpose of spreading bots in one or more social networks is referred to as a* ***Social network & Social engineering botnet*** *(****SS-botnet*** *for short).*

It can be seen that SS-botnets represent a class of advanced widely spreading botnets. Generally, the propagation process employed by SS-botnets can be divided into four steps. Fig. 1 illustrates the infection process of an SS-botnet across three social networks.

Step 1. Botmasters fabricate a trap message, which is similar to normal social network messages and attractive to receivers.

Step 2. Botmasters construct the command (marked as *CMD* in Fig. 1) with the trap message, and send CMD to the bots under the control of them.

Step 3. The target bot extracts the trap message from CMD, and then impersonates the victim user to send the trap message to all the friends of the user. We set this time to t_s.

Step 4. The victim's friends start to deal with the trap message after a uncertain time interval (assume this time is t_r). Generally, they have two choices: believing the trap message (by activating the bot program within the trap message), or avoiding it (by recognizing it as a threat). In the former case, if the antivirus software can't detect this threat, their computers will be infected and become bots, attacking other computers (Step 3). In the latter case, the users' computers become immune to the botnet infection due to the users' security awareness and exercise in judgement.

For the initial stage of SS-botnets' propagation, botmasters can anonymously register several accounts in one or more social networks to disseminate trap messages and capture the first group of initial bots. In addition, if the propagation is blocked by an effective defense, botmasters have the ability to design new bot programs and trap messages for the purpose of bypassing this defense. As such, the SS-botnet can continue with high-speed infection, and maintain their ability to attack.

3.3 Affecting Factors

We believe that there are two key factors affecting the propagation of SS-botnets as follows.

1) Cross-Domain Spreading

A *cross-domain node* is a social network node with the ability to allow its user to use more than one social networking service. Since the users of cross-domain nodes have access to multiple social networking services, they typically have different friends in different social networks. To illustrate the threat involved with this, envision the following scenario: botmasters design the bot program to disseminate the trap message from one social network to another. This will make the infection radius of SS-botnets wider and the rate faster. Therefore, cross-domain spreading should be taken into full account while developing the SS-botnet propagation model.

2) User Dynamics

The Variation of Active Users. Intuitively, if a social network node is not active (*i.e.*, its user isn't online in any social network at this time), it is impossible to infect others or to be infected by others. Therefore, the time of users' visiting social networks, and how long they remain active on each network, have a great impact on the infection process of SS-botnets.

Users' Responses. The general process of the SS-botnet propagation requires users' responses to the trap messages propagated by bot programs, although the response behaviors are often unintentional. It would seem that when the receivers deal with the trap message and the probability that users are deceived by trap messages also have a great effect on the infection process of SS-botnets.

4 Modeling Methodology

4.1 Related Notations

The related notations of our propagation model are provided in Table 1.

Table 1. The notations in the proposed model

Notation	Explanation
n_i	The node with index i
V	The set of all nodes under considersion
K	The number of social network domains
D_k	The social network domain with index k
$e_{ij,k}$	The variable indicating the relationship between n_i and n_j in D_k
$p_{i,t}$	The probability that n_i is infected at time t
I_t	The set of bots at time t
tw_i	The time interval before n_i's response
$\alpha_{ij,t}$	The connectivity between n_i and n_j at time t
$Act(i,k,t)$	The function indicating whether n_i is active in D_k at time t
β_i	The immunity to bots for n_i
S_0	The set of the original infected nodes

4.2 Theoretical Description

We assume an SS-botnet can spread bots in K social network domains and the topologies of these domains are known. The set V is given by

$$V = \bigcup_{k=1}^{K} D_k \tag{1}$$

There are cross-domain nodes in the set of V, which belong to more than one domain. We can formulate this case as follows:

$$\exists k_1, k_2 \in [1, K] \rightarrow D_{k_1} \cap D_{k_2} \neq \emptyset \tag{2}$$

A cross-domain node has the ability to be active in more than one domain simultaneously, a single domain, or no domain at any time. In the propagation process of SS-botnets, it is possible that a node is in the offline state at a given time (because the user may not be online in any social network). Obviously, the infection can't work for the offline nodes. Without loss of generality, we assume D_0 is a special domain where offline nodes reside.

We assume the time interval from sending the trap message (t_s) to the receiver's response (t_r) is tw. And $tw = t_r - t_s$. From a statistical point of view, it can be found that tw follows a heavy-tailed distribution according to [6]

$$Pr(tw) = \frac{\tau}{(tw)^{\tau+1}} \tag{3}$$

Where $tw > 1$ and $\tau > 0$. In addition, τ determines the mean of tw, which have different values for different social networks.

We employ the discrete-time stochastic process to model the propagation process of SS-botnets. The justification for employing the stochastic process is as follows: (i)The cycle time of sending a trap message, which has been fixed in the design of bot programs, can be discretized. (ii)The parameter tw can be discretized, which won't alter the characteristics of the tw distribution.

n_i is not infected at time t if and only if it was not infected at time $(t-1)$, and effectively blocked or ignored all trap messages sent by its friends in the same social network domain at time $(t-tw_i)$. These events are independent, therefore $p_{i,t}$ can be expressed as

$$1 - p_{i,t} = (1 - p_{i,t-1}) \cdot \prod_{j \neq i} (1 - \beta_i \cdot \alpha_{ij,t} \cdot p_{j,t-tw_i}) \tag{4}$$

We use the approximation equation $(1-x)(1-y) \approx 1-x-y$ (when $x \ll 1$, $y \ll 1$) to simplify Equation 4. Since $p_{i,t-1} \ll 1$ and $p_{j,t-tw_i} \ll 1$, we can get

$$p_{i,t} = p_{i,t-1} + \beta_i \cdot \sum_{j \neq i} (\alpha_{ij,t} \cdot p_{j,t-tw_i}) \tag{5}$$

From Equation 5, it is seen that the current state of n_i depends on both the state of n_i at time $(t-1)$ and the states of its friends at time $(t-tw_i)$. In addition to this, the parameter $\alpha_{ij,t}$ indicates the connectivity between n_i and n_j at time t, dictating the probabilistic behavior of the stochastic process; while the parameter β_i indicates the immunity to bots for n_i, affecting the result of the infection process.

First, we analyze $\alpha_{ij,t}$, which can be stated as

$$\alpha_{ij,t} = \sum_{k=1}^{K} Act(i,k,t) \cdot Act(j,k,t-tw_i) \cdot e_{ij,k} \tag{6}$$

Then the function $Act()$ is given by

$$Act(i,k,t) = \begin{cases} 1 & \text{if } n_i \text{ is active in } D_k \text{ at } t\ , \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Additionally, the value of $e_{ij,k}$ in Equation 6 can be determined according to the given topology of relationships in social network domains. From Equation 6, we can see that if n_i and n_j are friends in one or more domains, and n_i is active at t while n_j is active at $(t-tw_i)$, $\alpha_{ij,t} \geqslant 1$; 0 otherwise. And $\alpha_{ij,t} \in [0, K]$.

Next, we describe β_i, which expresses the ability of a user to assess and judge the trap message, along with the anti-bot capability of n_i. For simplifying the calculation, we set $\beta_i \in (0, 1]$. The smaller the value of β_i is, the larger the probability that the user remains uninfected will be. When β_i approaches to zero, n_i is hardly infected by bot programs, and if $\beta_i = 1$, n_i will be captured as long as it is attacked by other infected nodes with the trap message.

For the propagation process of an SS-botnet, we are concerned about the total number of bots at a given time (*i.e.*, the future size of an SS-botnet). Let I_t denote the set of infected nodes at time t, and the expectation of $|I_t|$ is given by

$$E(|I_t|) = \sum_{i=1}^{|V|} p_{i,t} \tag{8}$$

The propagation process is initiated by infecting a given number of nodes, which belong to the set S_0. And the initial condition can be stated as

$$p_{i,0} = \frac{|S_0|}{|V|} \tag{9}$$

Where $|S_0| \geqslant 1$.

5 Evaluation

In this section, we firstly introduce the statistical investigation for determining $\alpha_{ij,t}$. Subsequently, we examine the ability of our proposed model to predict the propagation process of an SS-botnet, with a goal of achieving early warning. Finally, we analyze the effects of the cross-domain spreading and user dynamics.

5.1 Statistical Investigation

In order to apply the proposed model to the prediction of the real-world SS-botnet dynamics, we should make the value of the parameter $\alpha_{ij,t}$ in accordance with realistic situations. For the topological information of social networks, there are two factors affecting the value of $\alpha_{ij,t}$: (i)the visiting time and (ii)the duration of active users. The former can be determined by the number of active users in target social networks at given time point, while the latter can be modeled by a heavy-tailed distribution [6] [7]. Assume that the total number of users in a social network is more or less constant over a short period of time, and let Nt_k represent the total number of users in D_k. Furthermore, we use $Na_{k,t}$ to represent the number of active users in D_k at time t.

In our investigation, we design some crawlers for fetching the number of active users in four popular social networks. The whole operation process lasted about three months with a sampling cycle of one hour. The studied social networks are: the BBS of Graduate University of Chinese Academy of Sciences (KYXK BBS) [8], the BBS of Tsinghua University (SMTH BBS) [9], QQ Game and QQ Instant Message (QQ IM) [10]. Fig. 2 shows the statistical results. From Fig. 2 it can be found that the number of active users is typically at a low level from midnight to 8:00, and at a higher level from 13:00 to 22:00. The minimum value (at 5:00) is about 1/3 of the maximum. Furthermore, we find that the diurnal variation trends of active users in these four social networks are very similar. Therefore, our equation assessing the visitation time and the duration of users is as follows:

$$\frac{Na_{1,t}}{Nt_1} \approx \frac{Na_{2,t}}{Nt_2} \approx ... \approx \frac{Na_{K,t}}{Nt_K} = R_t \tag{10}$$

Where R_t denotes the ratio of $Na_{k,t}$ and Nt_k. Thus $\alpha_{ij,t}$ can be calculated as follows:

$$\alpha_{ij,t} = (R_t) \cdot (R_{t-tw_i}) \cdot (\sum_{k}^{D_k \in Dc_{ij}} e_{ij,k}) \tag{11}$$

Where Dc_{ij} indicates the common social network domains of n_i and n_j, and the value can be determined with the given topological information of the target social networks.

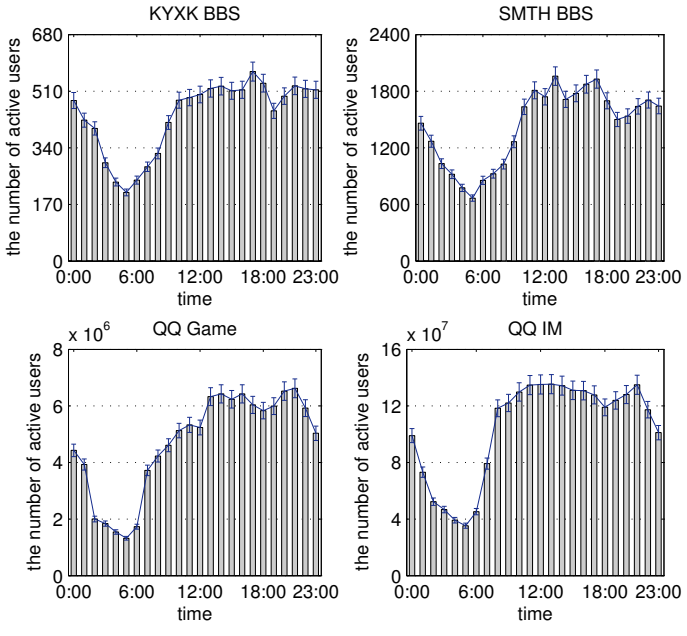

Fig. 2. The diurnal variation of active users in social networks

5.2 Simulation Experiment

OMNeT++ is an extensible, modular, component-based C++ simulation framework used for building network simulators. We've built a simulation platform based on OMNeT++ for the purpose of evaluating our propagation model. The simulation is made comparable to real-world situations by adjusting the values of the property parameters (*e.g.*, the out-degree and in-degree of nodes and the small average shortest path length). Thus, we can claim the three following characteristics of our simulation platform:

1) High Extensibility: this platform can simulate several virtual social networks simultaneously, and every virtual social network is generated according to the algorithm in [4].

2) Fine-Grained: the platform can simulate multiple behaviors of nodes, such as when to deal with the trap message.

3) High Flexibility: in our platform, the interval between discrete time points can be adjusted to simulate the infection processes of different SS-bontets.

In our platform, we designed a virtual SS-botnet that can spread across two social network domains, and then simulate the propagation process of the SS-botnet. Additionally, we assume β_i follows the uniform (0, 1) distribution. The values of related parameters are shown in Table 2.

Fig. 3 shows the comparison between our model and the results of two simulations for the propagation process of the SS-botnet in one and two domains.

Table 2. The values of the relative parameters in the simulation

Parameter	Value
Nt_k	10000
$\|V\|$	10000 and 20000
K	1 and 2
$\|S_0\|$	1 and 100
τ in the distribution of tw	1 and 1.5
the max out-degree in D_k	100
the mean of β_i	0.7
the starting time	5:00

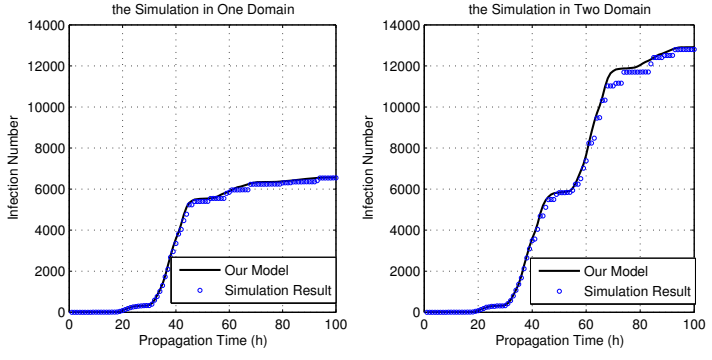

Fig. 3. The comparison between our model and the simulation results

The comparison indicates that our model can accurately predict the size of an SS-botnet with less than 5% deviation.

Furthermore, we experiment with MATLAB for evaluating the influence of different factors on the infection of the SS-botnet according to our model, which yielded some intriguing results (as shown in Fig. 4).

Fig. 4(a) illustrates the influence of the diurnal variation of active users on the propagation process of SS-botnets. Since the starting time of our simulation is 5:00, it can be found that the infection speed in the daytime is much faster than in the night, presumably because more users are offline at night. Fig. 4(b) illustrates the influence of the factor tw. If tw is not taken into consideration, the infection rate will be faster but not realistic. Fig. 4(c) presents the propagation processes of different numbers of cross-domain nodes (5, 10, 100). It can be found that the proportion of cross-domain nodes can also affect the infection rate of the SS-botnet. In other words, the more cross-domain nodes available for infection are, the faster the SS-botnet will spread. Fig. 4(d) illustrates the influence of both the degree and the number of initial infected nodes. In this figure, a *popular node* is the node with a high in-degree, which is different from a normal node. It's shown that choosing popular nodes as initial nodes or increasing initial nodes can accelerate the infection rate.

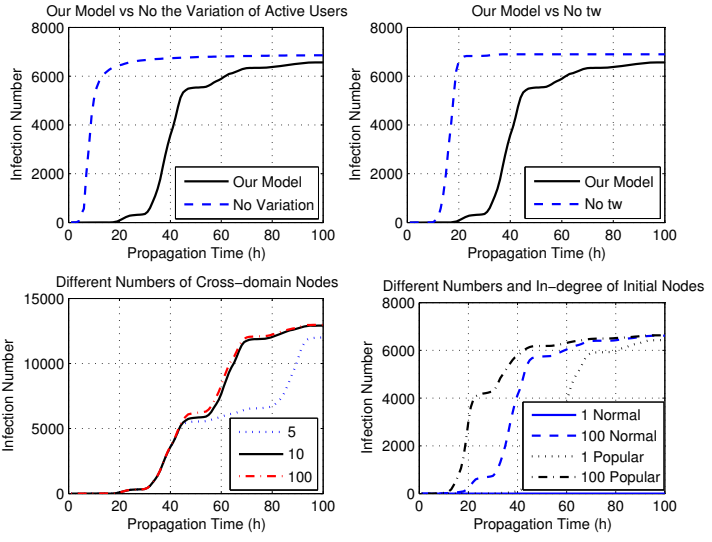

Fig. 4. The influence of different factors

6 Limitations and Discussion

In this section, we discuss some limitations of our propagation model. Firstly, in the proposed model we assume that social networks are static when considering their topological characteristics and the initial set of infected nodes. In reality, however, the number of registered users and the relationships in real-world social networks are always changing, which makes the propagation process of SS-botnets more complex. We believe these real-life variations are very important factors which should be considered in future work. Secondly, the parameter β_i follows a uniform (0, 1) distribution in our model, while in reality it may follow other distributions (such as a beta distribution). In future work, we plan to investigate the distribution of β_i to optimize our model. Finally, since SS-botnets are emerging and our work is a prospective study, there is no real attack data available for the validation of our model. In addition, fetching the information of an SS-botnet's bots is more related to the detection technologies, which is beyond the scope of this paper.

7 Conclusion

In this paper, we present a new propagation model for SS-botnets which represent an new type of emerging botnets and have a growing risk to social networking services. By exploiting the general infection process of SS-botnets, we reveal two key factors that affect the infection of SS-botnets: (i)the cross-domain spreading of bots and (ii)the dynamics of social network users. Next, we propose

the theoretical description of the proposed model based on stochastic process, and determine the values of some important parameters in our model with statistical data obtained from real-world social networks. Furthermore, we built a platform based on OMNeT++ for simulating the propagation process of SS-botnets to evaluate our model. The experimental results indicate that our model has the ability to accurately predict the size of SS-botnets with a small deviation. In future, we would like to make our simulation platform more comprehensive to support the further study of the SS-botnet dynamics. And then, we will focus more research on the ways to fight against SS-botnets. Our ultimate goal is to utilize the proposed model and our simulation platform to help defenders design effective containment mechanisms of SS-botnets.

References

1. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: 19th International Conference on World Wide Web, pp. 591–600 (2010)
2. Abraham, S., Chengalur-Smith, I.S.: An overview of social engineering malware: Trends, tactics, and implications. Technology in Society (2010)
3. Thomas, K., Nicol, D.M.: The Koobface botnet and the rise of social malware. In: 5th International Conference on Malicious and Unwanted Software (MALWARE), pp. 63–70 (2010)
4. Toivonen, R., Onnela, J.P., Saramáki, J., Hyvónen, J., Kaski, K.: A model for social networks. Physica A: Statistical and Theoretical Physic. 371(2), 851–860 (2006)
5. Ruitenbeek, E.V., Sanders, W.H.: Modeling peer-to-peer botnets. In: 5th International Conference on Quantitative Evaluation of Systems, pp. 307–316. IEEE Computer Society (2008)
6. Barabási, A.L.: The origin of bursts and heavy tails in human dynamics. Nature 435(7039), 207–211 (2005)
7. Sarat, S., Terzis, A.: On using mobility to propagate malware. In: 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops (WiOpt 2007), pp. 1–8. IEEE (2007)
8. KYXK BBS, `http://kyxk.net`
9. SMTH BBS, `http://www.smth.edu.cn`
10. Tencent QQ, `http://www.qq.com`
11. Zou, C.C., Towsley, D., Gong, W.: Email worm modeling and defense. In: 13th International Conference on Computer Communications and Networks (ICCCN 2004), pp. 409–414 (2004)
12. Yan, G., Chen, G., Eidenbenz, S., Li, N.: Malware propagation in online social networks: nature, dynamics, and defense implications. In: 6th ACM Symposium on Information, Computer and Communications Security, pp. 196–206 (2011)
13. Cheng, S.M., Ao, W.C., Chen, P.Y., Chen, K.C.: On Modeling Malware Propagation in Generalized Social Networks. IEEE Communications Letters 15(1), 25–27 (2011)
14. Dagon, D., Zou, C., Lee, W.: Modeling botnet propagation using time zones. In: 13th Annual Network and Distributed System Security Symposium (NDSS 2006) (2006)
15. Varga, A., Hornig, R.: An overview of the OMNeT++ simulation environment. In: 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops, pp. 1–10 (2008)