International Conference on Computer Modeling and Simulation

Using simulation to characterize topology of Peer to Peer Botnets

Junfeng Yu, Zhitang Li, Jun Hu, Feng Liu
Network &computing center, college of computer science & technology
Huazhong University of Science and Technology
Wuhan, China
jfengyu@gmail.com

Lingyun Zhou
College of computer science & technology
South central university for nationalities
Wuhan, China

Abstract—**Future network intruders are constantly changing, improving and extending the capabilities of their botnets. Peer to peer bots are now under widespread development and are quickly evolving into a much tougher species to kill. However, partly due to the lack of understanding of the structural potential of command and control mechanism a botnets can have, countering peer to peer botnets has been ineffective. In this paper, we explore the characteristic of communication networks created by peer to peer botnets. We propose a simulation approach to characterize the structural properties and robustness of P2P botnets, corresponding to different paradigms of forming botnets. Through calculation and simulation for topology construction procedures, we show that extremely resilient peer to peer botnets can be formed to deliver attack code quickly. Such a theory would help predict botnets containment for a given topology and help develop strategies to improve defense against P2P botnets, because applying those strategies can always translate into some network topology transformation.**

Keywords-Peer to Peer Botnets; simulation; complex network; topology

I. INTRODUCTION

Over the past several years, botnets have turn out to be the most serious problem that ever confronted by the Internet. Many recent reports indicated that botnets are in a new state of steady evolution and have proceeded in sophistication (e.g., [1, 2, 3]). It is noted that botnets development at present quickly falls into two categories. One area of development attempts to provide more intricate attack functionalities. The new generation of botnets can send email spam, spread new malware, launch DDOS attack, attempt to harvest information and even upgrade their components. Another area of botnet development is focus on innovation in control (C&C) control mechanism [3]. Botmasters are looking for more delicate strategies than standard IRC that they can dominate their bots army without being caught. The peer to peer (P2P) architecture enables true distributed communication. It creates networks of computing resources which exhibit fault tolerance and very high variability that are perfect for this felonious intent. Due to the structural resilience of P2P botnets, it is very difficult to annihilate their entirety.

In order to respond to this threat, security researchers require an understanding of the structural potential of peer to peer command and control (C&C) mechanism and interplay of topological resilience that might shape their behaviors. In this paper, we focus on the characteristic of topology employed by P2P botnets. We formally model the P2P botnets topology with the help of complex systems.

The prevalence of P2P botnets have gained wide attention from the research community, there have been many deep studies into botnets moving to the use of P2P concepts. Several previous studies have proposed a few approaches [4, 5, 6] to detect and mitigate the existence of botnets in monitored networks. For example, BotHunter [4] is designed to detect bots using a predefined infection life cycle dialog model. However, P2P botnets are evolving and can be quite flexible. Developing a clear understanding of the structural properties is the key step towards design of efficient responses against the P2P botnets. A number of characterizations and analyses of different topology types employed by P2P botnets are available in the literature [7, 8]. In [7], Sam stover et al. gave a comparison of recently successful botnets that employ P2P concepts and discuss how topologies for command and control structure have changed over time. Dagon et al. [8] present a taxonomy of botnets based on topological structure and there corresponding metrics. Our research presented in the paper belongs to this category.

The rest of the paper is organized as follows: We present an overview of P2P botnets system components in Section Ⅱ. In SectionⅢ, we discuss in detail a general model utilized by P2P botnets to build C&C overlay networks. We describe the methodology and results of our simulation study in Section Ⅳ. Finally, we conclude in sectionⅤ.

II. SYSTEM COMPONENTS OF P2P BOTNETS

Jose Nazario proposed that any malware systems are composed of six components [9]. In our paper, we focus our attention on ways of maintaining P2P C&C channel and the topology of the P2P botnets, we simplified it into two functional components: attack functionality and P2P functionality.

- Attack functionality: This is the components that actually launch attacks. Since botnets comprise hundred thousands of Zombie Computers, It can

978-0-7695-3562-3/09 $25.00 © 2009 IEEE
DOI 10.1109/ICCMS.2009.61

perform any nefarious action that requires large numbers of computers. When P2P overlay command and control channel is well formed，botmasters can update new functionalities be crafted to any purpose at anytime by inject new code in the overlay.

- P2P functionality: This component maintains the C&C networks among the bots army. Any infected nodes can provide and retrieve information at the same time. This feature makes P2P botnets extremely robust against node failures [10]. Using P2P communications to create an overlay network for controlling compromised computers is not new, but the P2P functionality employed by botnets has steadily evolved. The continuous evolvement of P2P functionality enables the dynamic nature of distributed P2P systems.

III. General model of P2P functionality

A. *Common definitions*

In the recent years, botnets such as Slappe, Sinit, Phatbot, Peacomm, and Nugache have implemented different kinds of P2P control architectures [11]. For example, once the computer is infected, Peacomm initiates communications by contacting a small list of IP addresses, in order to receive coordination instructions from a community of peers within the larger botnets. It also receives a list of additional IP addresses of infected machines and adds them to its list of available peers, building up a distributed network to aid in the download of more malicious codes. Similarly, most of other P2P botnets build up a P2P C&C channels like this way, in the sense that every infected node constantly maintain a list of known nodes in the P2P network and endeavor to make this list synchronized with each other, to ensure that a robust and efficient overlay network be well formed. For a more formal definition, we assume the following common settings for P2P botnets functionality:

- Each bot is identified with an unique descriptor which include IP address and a profile. The profile contains those properties of the bots that are relevant for defining the topology, such as ID, network attribute, etc.
- Each bot has a descriptor table of fixed size C, namely partial view, where each bot keep information about neighbor bots.
- Each bot is categorized into two types according to their network attribute. The first type called servent bot [12] which are accessible from the globe internet. It acts as both client and server. Only servent bot are candidates for being in peer lists. The other type is named client bot which can not connected from the global internet and acts as a client.

B. *Topology construction procedures*

In this paper, we consider botnets establish an overlay network with a two phase process. In the first phase, when botmasters issue an infection command, each sevent bot build new copy of bot and propagate it through infection vectors. The parent bot randomly replace one item of descriptor table with its own descriptor and pass it to new victims. In this way, the newly infected bot inherits the descriptor table from parent bot. It provides an effective way to build P2P communication channels during propagation. The infection vectors used by bots are not relevant to the analysis and are not considered. We also ignore failures to propagate in our model. This topology building strategy be used in initiate phase and thus is called “Type Ⅰ” construction procedure throughout the current paper. In the second phrase, an advanced overlay construction should continue when botmaster issues a reconstruction command. This process effectively reorganizes the botnets that all bots will have uniform and balanced connections. We called this process as “Type II” construction procedures.

For simplicity, we present pseudo-code of these two processes in the following text. In the topology construction scenario, each bot executes the same protocol shown in Fig. 1. The protocol consists of two threads: an active thread infecting new victims and constructing peer list, and a passive thread waiting for incoming messages of descriptors from other peers. Furthermore, we present the Type II construction procedure in Fig.2.

```
do forever
  if(firstInfection AND serventBot) then
    myDescriptor <- (myAddress, myProfile)
    send myDescriptor to parent bot
    inheritView(view)
    firstInfection = FALSE
  endif
  wait( infection command)
  myDescriptor <- (myAddress,myProfile)
  infectNewVictim(buffer)
enddo
```

(a) active thread

```
do forever
  p <- waitMessage()
  buffer <- merge(rnd(view,c-1,p)
  view <- buffer
enddo
```

(b) passive thread

Figure 1 P2P botnets pseudocode for Type I topology construction

```
do forever
    wait(topology reconstruction command)
    buffer <- selectPeer(c)
    view <- buffer
enddo
```

Figure 2 P2P botnets pseudocode for Type II topology construction

The two key methods are selectView(c) and rnd.View(View,d). Method of selectView(c) returns some descriptors size of c being chosen uniformly at random from a globe view of peer bots. In a real P2P botnets, regardless of the specific structure of the botnet (unstructured and structured), members of the botnets are coordinated through the C&C channel. A Botmaster can issue a sevent bot register or subscription command periodically to maintain the global view and publish it in a dynamic server. In this way, method selectView(c) get the globe view of all peer bots and return descriptors size of c to each bot. To the best of our knowledge, this holds true for most of the existing botnets observed in the wild. Method rnd.View(View,d) return d descriptors which is subset of the View with an unbiased random selection.

As a final note about our abstractions and simulations, we stress that they omit a lot of issues, including the strongly diurnal behavior of botnets [13], the message delays in the network, the actual failures of propagating malicious codes, etc. However, these models help us understand the fundamental properties of various types of topology.

IV. SIMULATION STUDY

A. *Experimental methodology*

In what follows, we discuss the most important structural properties of the system as a whole. Instead of an analytical approach presented in [7], in our methodology, we switch to a graph theoretical framework, which provides richer possibilities of interpretation from the perspective of fundamental properties of topology and its robustness to responses.

To translate the problem into a graph theoretical language, we consider the communication topology or overlay topology defined by the set of bots and their partial views. In this paper, We represent the topology of the P2P botnets by an undirected graph: G=<V,E>, $|V|$ is the total number of bots in P2P botnets. The undirected edges E of the communication graph are defined as follows. If bot a stores the descriptor of bot b in its view then there is an undirected edge (a, b).

To characterize the dynamics of the C&C overlay networks of P2P botnets, we investigate: (1) what properties (such as diameter, degree of connectivity, and clustering) these types of network exhibit, and (2) what underlying factors contribute to the formation and properties of such networks. We simulated the pseudo-code of simulation model presented in section Ⅲ. As pointed out in [13], botnets in recent years have dropped their sizes to an average of 20,000, even though the potential vulnerable population is much larger. In addition, each bot connects with there neighbor nodes to maintain the overlay networks, however, there is a tradeoff between robustness and disclosure. If c is too large, the overlay networks have strong robustness even though with high disclosure to the defenders. Accordingly, we chose to the network size to N = 20k, and the partial view size to c = 20 in our current simulation.

B. *Properties of degree distribution*

There are several ways of measuring the functionality of overlay networks [14,15,16]. Degree distributions of graph have been identified as a key property because of its relationship to robustness of a graph, its effects on patterns of epidemic spread, and its importance in the distribution of resource usage of nodes. Since the topology construction procedures directly affect the overall properties of the overlay network created in P2P botnets, It is important to exam the distribution of degrees between two different topology construction procedures. This will help us find out how the construction procedures affect the properties of the overlay.

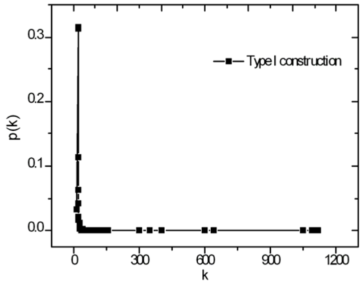

Figure 3 degree distribution of servent bots created by Type I construction procedure

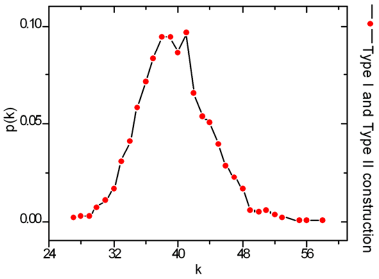

Figure 4 degree distribution of servent bots created by Type I and Type II construction procedures

Fig. 3 and Fig. 4 show a comparison of the degree distribution for servent bots when a botnet uses different construction strategies. Fig. 3 indicates that the degree distribution of the P2P botnets overly networks with Type I construction procedure shows a very inhomogeneous distribution. As we can see that connections to servent bots are extremely unbalanced: 80% of servent bots have degrees less than 30, while there are some bots have a degree between 600 and 1000. The emergence of these inhomogeneous degree distributions in the network implies that the first construction phrase can’t build a robust overlay network. Fig. 4 shows that the degree distribution is homogeneous (all the values are close to the average, like in passion distributions) with both Type I and Type II construction procedures, our simulation results indicated that topology reconstruction procedure is necessary for forming a uniformly connected botnets.

C. Average path length and clustering coefficient

Degree distribution is an important property of graphs. However, there are other equally important characteristics of networks that are independent of degree distribution [17]. In this section we explore the average path length L and the clustering coefficient C as two such characteristics in our simulation study.

We conducted a simulation experiment in which we add 100 new bots at every cycle until a size of 20k P2P botnets had been evolved. In each cycle we measure the average path length and clustering coefficient and obtain the results shown in Fig. 5. We can find in Fig. 5 that the average path length of networks created by different construction strategies increases approximately logarithmically with the size of the network. The C&C overlay network in P2P botnets with both types of construction procedures has a systematically shorter average path length than that one with only Type I construction procedures.

Fig. 6 shows the clustering coefficient of the networks with different building procedures on the log-log scale. We find that the clustering coefficient of network with both types of construction procedures is a little higher than of networks created without Type II construction phase. However, the clustering coefficient of these two model decrease with the network size following approximate a power law $C \sim N^{-0.68}$ and $C \sim N^{-0.43}$. This behavior is distinct from the behavior of small world networks where clustering is independent of system size.

In summary, we report statistical measures showing that P2P botnets employed primary topology construction procedures form a scale free network with small-world properties. While this type of networks can be reconstructed under an evolutionary process by iteratively adjusting its pattern of connections. With more advanced topology reconstruction process, all current bots will have uniform and balanced connections that exhibit stronger robustness to adversary response.

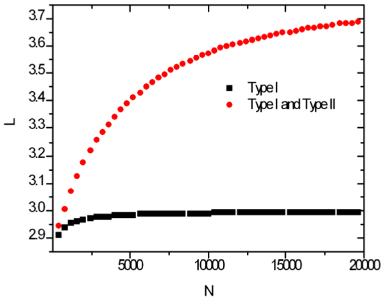

Figure 5 comparison of the average path length with two different types of construction procedures

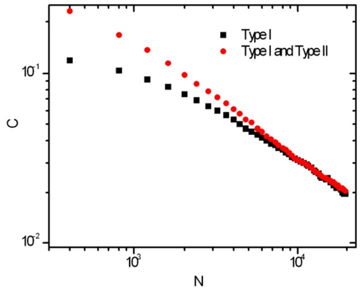

Figure 6 comparison of the clustering coefficient on the log-log scale with two different types of construction procedures

D. Robustness to bots removal

We have designed a simulation to study the resiliency of P2P botnets to different types of responses by defenders, aiming to answer two questions: (1) For a given P2P botnets, if fraction of f nodes are disinfected, what percentage of remaining nodes will remain connected? (2) What is the impact of a botnets size on its resiliency? Here, we assume that a bot does not attempt to reconnect any bots after the disinfection. Also, since disinfection is often about cleaning up bots, not connections between bots, our resiliency study will focus on dropping nodes, not links, from a botnets. We consider a defender who will try to disable a P2P botnets using two different strategies. First, a defender may shutdown high-degree bots. This is an effort to reduce P2P botnets connectivity more quickly than simply annihilating bots at random. We call this response as degree based response. The second response strategy is random response, since a defender may annihilate botnets randomly upon discovery.

David dagon et al. proposed two metrics: size of gaint component, inverse average path lenth, to measure the robustness of botnets [8]. In this paper, we denote the size of the giant component as S, which will be used together with inverse average path length l^{-1} to study the robustness of

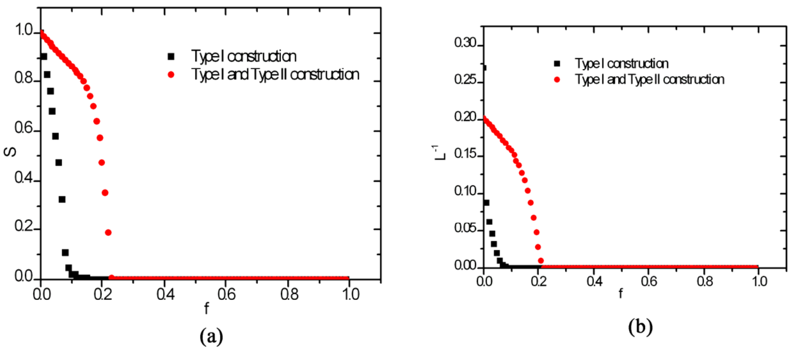

Figure 7 robustness against degree based response is measured by S and l^{-1} as a function of the percentage f of bots removed from the system

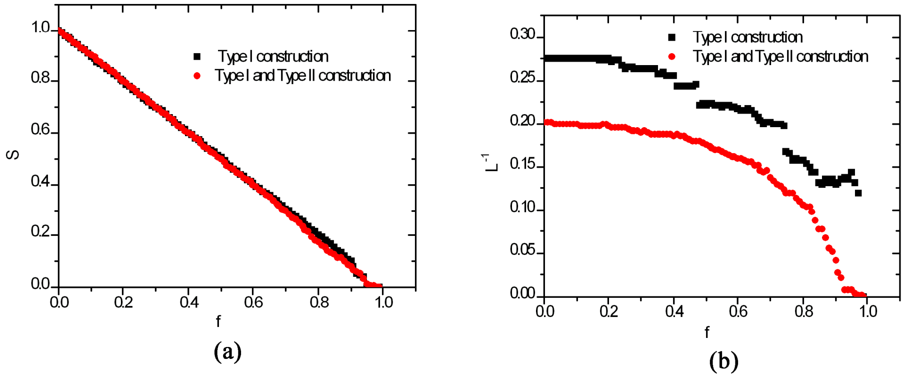

Figure 8 robustness against random response is measured by S and l^{-1} as a function of the percentage f of bots removed from the system

P2P botnets against a defender's countermeasures. Here we explore the effects of two types of response and compare them with both overlay networks created with type I construction model and type II construction model. For each experiment, we investigated the relative size of the largest component that remains connected, S, and the average inverse geodesic length l^{-1}, in function of the fraction f of the nodes removed from the system. We expect that the size of S and l^{-1} decreases as an increasing number of nodes are removed from the network. In all simulations, we create overlay network with a total number of bots N=20k that are interconnected randomly to each other with each node maintaining a number of neighbors C=20.

We plot the Fig.7 for degree based response against P2P botnets (both with N=20k nodes and K=400k edges) as functions of the percentage f of removed nodes. We compare Type I construction procedure with both Type I and Type II construction procedures. The Type I construction model shows highly different behavior with respect to responses: if we remove 5% of nodes in a targeted way, the size of S and l^{-1} is reduced to zero, that destroy completely the botnets system; instead, when we remove nodes in network created with both types of construction procedures, the network shows a slow descendent curve and also for the high value of f = 20% the system maintains a considerable connected.

In Fig. 8 we plot again the largest connected component and average inverse geodesic length for random response against P2P botnets as a function of f. As show in figure 8 the networks with two different creation procedures behave quite similarly and decay linearly under random response, in contrast to degree based response where l^{-1} and S decay exponentially.

Both in two different types of topology construction procedures model the giant component and inverse geodesic length persists for high rates of random response, while, if the nodes are removed in the degree based response mode, the size of the fragments that break off increases rapidly. The main conclusion of the simulation study presented in this section is that P2P botnets with Type I and Type II construction procedures, due to their homogeneity, exhibit a stronger robustness with respect to degree based response, while P2P botnets with only Type I construction procedure, because of their heterogeneity, are fairly robust to random response, though very vulnerable to degree based response.

V. Conclusion

In this paper, we first note that P2P botnets system components can be divided into two parts: attack functionality and P2P functionality, in the perspective of botnets developer. Then, we present a general P2P functionality model by considering C&C mechanism, such as common settings of each bot and topology construction processes. To understand how the structural potential of P2P command and control mechanism and interplay of topological resilience that might shape their behaviors, we propose a simulation approach to characterize the structural properties and robustness of P2P botnets, corresponding to different paradigms of forming botnets.

We have shown that naive defenses don't work. Simply randomly remove infected computers does not slow down the attacker much, regardless of whether the connectivity between peer bots follows a random or scale-free pattern. We found that the degree based responses on P2P botnets offer best approach. As we have shown, complex network methods may reveal some invariant properties of the P2P botnets and open an insight about functional aspects of their topology in terms of robustness and communication efficiency. Such a theory would help predict botnets containment for a given topology and help develop strategies to improve defense against P2P botnets, because applying those strategies can always translate into some network topology transformation.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No. 60573120, and by the National High Technology Research and Development Program of China (863 Program) under Grant No. 2007AA01Z420.

References

[1] Vogt R, Aycock J, Jacobson MJ. "Army of botnets, " In: Proc. of the 14th Annual Network & Distributed System Security Conf. (NDSS). 2007, pp. 111-123.

[2] Dittrich, David; Dietrich, Sven, "P2P as botnet command and control: A deeper insight," Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on , vol., no., pp.41-48, 7-8 Oct. 2008

[3] Ping Wang, Sherri Sparks, and Cliff C. Zou "An Advanced Hybrid Peer-to-Peer Botnet,"In Proceedings of the USENIX 1st Workshop on Hot Topics in Understanding Botnets (HotBots '07), Cambridge, MA, April, 2007,pp.2-2.

[4] Guofei Gu, Phillip Porras, Vinod Yegneswaran, Martin Fong, Wenke Lee. "BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation.," In Proceedings of The 16th USENIX Security Symposium (Security'07), Boston, MA, August 2007, pp. 182, 167.

[5] J. R. Binkley and S. Singh, "An algorithm for anomaly-based botnet detection, " in USENIX 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI 06),, June 2006, pp. 43--48.

[6] Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. "BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection.,"In Proceedings of The 17th USENIX Security Symposium (Security'08), San Jose, CA, July 2008.

[7] Sam Stover, Dave Dittrich, John Hernandez, and Sven Dietrich." Analysis of the Storm and Nugache Trojans: P2P is here". In USENIX, vol. 32, no. 6, December 2007.

[8] David Dagon, Guofei Gu, Chris Lee and Wenke Lee." A Taxonomy of Botnet Structures,"In Proceedings of The 23rd Annual Computer Security Applications Conference (ACSAC 2007), Miami Beach, FL, December 2007, pp. 325-339..

[9] Jose Nazario, Defense and Detection Strategies Against Internet Worms, Artech house Publishers, Norwood, MA, 2004.

[10] M. Jovanovic, F. Annexstein, and K. Berman, "Modeling peer-to-peer network topologies through small-world models and power laws," in Telecommunications Forum, November 2001.

[11] Lidong Zhou, Lintao Zhang, Frank McSherry, Nicole Immorlica, Manuel Costa, and Steve Chien. "A First Look at Peer-to-Peer Worms: Threats and Defenses," Proceedings of the 4th International Workshop on Peer-To-Peer Systems (IPTPS 2005), vol. 3640, 2005, pp. 24.

[12] "Servent," http://en.wikipedia.org/wiki/Servent.

[13] D. Dagon, C. Zou, and W. Lee, "Modeling botnet propagation using time zones," in Proceedings of 13th Annual Network and Distributed System Security Symposium (NDSS), Feburary 2006, pp. 235–249.

[14] Li J, Ehrenkranz T, Kuenning G, Reiher P. "Simulation and analysis on the resiliency and efficiency of malnets," In: Proc. of the IEEE Symp. on Measurement, Modeling, and Simulation of Malware (MMSM 2005). Monterey: IEEE Computer Society Press, 2005. pp.262-269.

[15] Petter Holme, Beom Jun Kim "Edge overload breakdown in evolving networks," Phys. Rev. E 66, 036119 (2002).

[16] M. Kim, "Robustness in large-scale random networks," S.M. Thesis, Massachusetts Institute of Technology, 2003.

[17] M. Newman, S. Strogatz, and D. Watts, "Random graphs with arbitrary degree distributions and their applications," Phys. Rev. E., vol. 64, 2001.