International Journal of Security and Its Applications
Vol.9, No.9 (2015), pp.123-130
http://dx.doi.org/10.14257/ijsia.2015.9.9.11

Detection and Prevention to Network Worm Virus Based on Segmentation Algorithm

Yuqi Tang

Tangshan Vocational & Technical College, Hebei, China
Yuqi2201@126.com

Abstract

This paper Through an analysis of the worm propagation behavior, found that there are certain constraints relationship of network size to the propagation speed of the worm . On this basis, network segmentation algorithm based on DFS is proposed to reduce the size of the network, in order to gain subnet boundaries , in order to further suppress the spread of worms provide a theoretical basis , provide reference and basis for the erection of a network worm isolation system. the method of using DFS network segmentation algorithm based on network topology was proposed to solve large-scale network analysis. It provides a strong theoretical support for fast access network backbone node. The result of experiment shows that the proposed method is effective.

Keywords: *sub-graphs; DFS; network segmentation algorithm; Worm detection*

1. Introduction

Social networks worms (hereinafter called "social network worm “) is use of social engineering to entice users to click various ways to spread a worm, it has the characteristics of hidden, long life cycle and difficult to eradicate, and so on.it is difficult to spread through released patches and other technical means for effective control, and so the potential hazard is more serious [1]. At the same time, the actual environment network administrators and network user lack of security awareness, so it provide a breeding ground for social network worm propagation and survival. With the rapid development of the growing number of Internet users and various forms of virtual social network, spread through the network user’s social network worm has become one of the major threats to network security risks [2].

In response to the threat of worms, as well as potential large-scale network anomalies, there is a big shortage in methods and strategies used by people currently. These methods , some still in the theoretical stage , such as benign worms; others due to too much impact of engineering factors is not practical, and practical application is still far , such as auto repair vulnerabilities . Endless variety of worms, and once a major outbreak, it caused huge losses [3]. The traditional host-based protection, including virus prevention technology and virus firewall technology, can only be the point guard; and LAN-based worms isolation, the same cannot cope with the large-scale network worm outbreaks.

In this paper, we apply the method of using DFS network segmentation algorithm based on network topology by analyzing the study of graph depth-first search (Depth First Search, referred to as DFS) segmentation algorithm, and thus obtain a description of the network connectivity of undirected connected graph G (V, E).this provide a strong theoretical support for fast access network backbone node.

2. Effect of Network Size on Worm Propagation

Since the Internet itself is an open and complex giant system with multi-variable, the size of the dynamic change constantly, but taking into account the worm outbreak time is

ISSN: 1738-9976 IJSIA

short, and in general the target node number is relatively large [4,5], in a certain period of time the number of N can be regarded as a constant. Social networks, social networks derived from social networks, it starts form the email, BBS, instant messaging and P2P content sharing promoted the development of social networks, with the application of WEB2.0 technology, the appearance of online social networking sites further enrich the social network in the form of Internet. Mainstream social networks of currently existing including: e-mail network, P2P content sharing network, instant chat network and online social network sites and so on[6,7]. In order to define the scope of the study and social network worms and other types of Internet worm phase difference, the social network worms is defined as: social network worm is a section of malicious programs which does not depend on the specific vulnerabilities, Take advantage of social engineering techniques to deceive Internet users click to execute and infect computer systems. Worms spread through social networks released immune system patches and other technical means, only by way of enhanced user safety awareness to prevent it [8].

2.1 Worm Propagation Stage

Worm propagation process can be divided into of the three stages of slow start, fast propagation and slow end. As shown in Figure 1.

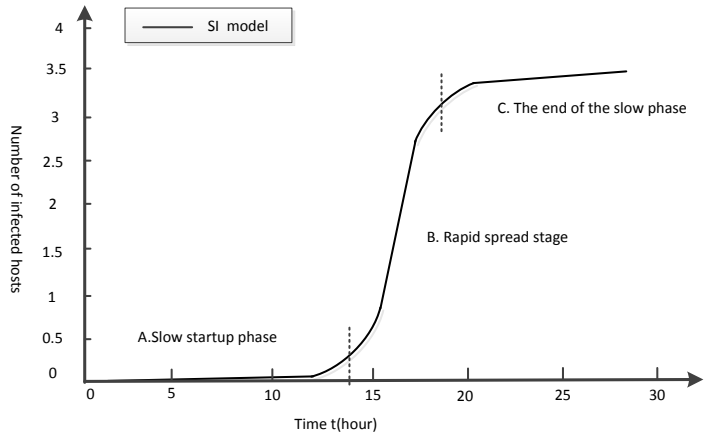

Figure 1. Propagation Phases

As can be seen from the figure, if the worm is in fast propagation and slow end stage, large-scale worm propagation will lead to network congestion, and result in a decline in the performance of network routers, and worms have caused enormous damage, thus inhibiting study in this paper is the first stage in the spread of the worm, which would be truly effective in inhibiting the spread of worms.

2.2 Network Size and Worms Coverage

For random scanning of worm propagation, assuming a slow start in each phase of the network nodes, such as the probability of being susceptible to infection, the worm's infection rate β is constant interference, regardless of network performance and other factors.

Then for random scanning, the infection rate can be defined as:

$$\beta = KN \quad (1)$$

Among them, k is a constant, determined by the scan efficiency and success rate and frequency .At time t network, the number of the susceptible host is s (t), the number of the infected host is I (t), then the classical SI models. In this way, combined with (1) and (2), set the initial number of infected nodes I (0) =I0, then:

$$\frac{dI_t}{dt} = \mathrm{KN} I_t(\mathrm{N} - I_t) \tag{2}$$

Due to the slow starts of the worm stage, N-It $\approx$ n, so you can approximate that the solution of equation (3) is:

$$I_t = I_0 e^{\mathrm{KN}^2 t} \tag{3}$$

By equation (1) can be seen, rates of random scanning directly depends on the size of the network, so as to affect the speed of the worm's spread. If the network of n nodes, and I nodes infected with worms, now defines I/N as the worms cover p, then p=I/N. Substituted into equation (2), you can get worms cover equation for:

$$\frac{dP_t}{dt} = \mathrm{KN}^2 P_t(1 - P_t) \tag{4}$$

Worm cover equation (4) reflects the impact of network worm propagation.

2.2 The Simulation Results

On the basis of random scanning and SI model, the analysis of worm propagation is not affected by other means of communication network, segmentation of worm inhibitory effect. IfK=2^{-64}， I(0)=10 unchanged, the spread of the worm was simulated under the different network scale:: a) worms spread throughout the IPV4 address space, the network size N=2^{32};b) the size of the network N=2^{32}/3; c) the size of the network N=2^{32}/10. The simulation results are shown in figure 2.

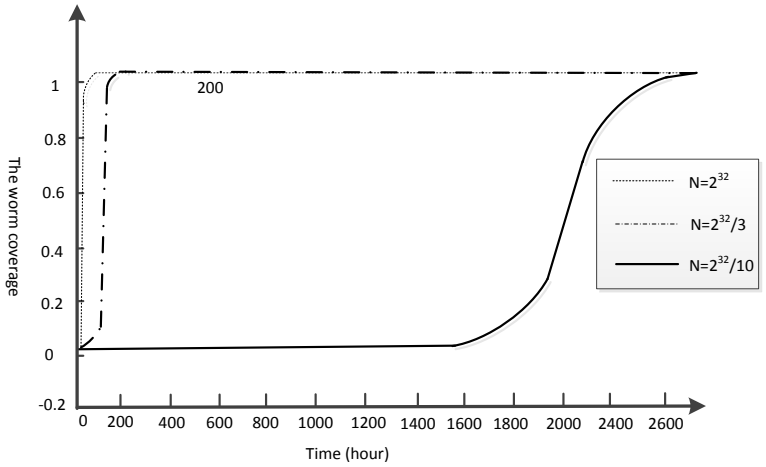

Figure 2. Worm Coverage and Network Scale Simulation Diagram

As can be seen from the figure, the number of infected nodes and infected nodes in the percentage of the entire network, when the network is small, the spread of the worm have been a much longer time before they start to rise significantly.

3. DFS Partition Algorithm Thought and Realization of Graphs

Based on the detection of worms, it should divide the network to reduce the scale of the network, detect and isolate the worms on the demarcation of the border, only in this way can effectively curb the worm in large-scale network wandering and impact, curb the spread of worms.

Method of providing or using the Internet as measured by network management institutions at all levels [8-10], it is possible to obtain a more accurate network topology information and data. Through the analysis of network topology, you can get a description of the network connection undirected connected graph G (V, E).

Definitions 2.1 The collection V and E consists of Graph G, by denoted G(V， E). Where: V is the set of vertices, E is a finite set of vertices in V. ,generally, the graph G, the vertex

set and edge set denoted by V (G) and E (G). E (G) can be the empty set, if E(G) is empty, then the G graph only vertices and no edges.
The basic idea of DFS algorithm of the graph is:
Assume that the initial state is the graph of all vertices not been accessed, for a vertex Vi from the figure, the access to this vertex , followed by Vi from the neighbors have not been accessed in a depth -first traversal until all of the figure and there is a path Vi vertices are visited until ; this case, if there are vertex graph is not accessible, the vertices of the alternative access is not the starting point and the process repeats until all graph vertices have been visited so far.
Known figure G (V, E), when each edge of the figure search , according to the following steps:

1) When E (G) of all sides without fully search, either take a vertex $v_i \in V(G)$, to v_i and to mark the stack ;
2) when search v_i points associated with the edge , if the other side to mark the endpoint is not present, then the other end to make [as v_i, give signs, and stack, sub (2) , otherwise transfer (3) ;
3) when all the search associated with vi edge is completed (Vi when that does not exist in order to search for the endpoint and without side), then v_i exit point from the top of the stack , the stack is not empty, as if to let [v_i removed after the top element as v_i, turn (2)] , otherwise transfer (4) ;
4) If the stack is empty , but there are still signs of the vertices not given , take any of the vertices as vi, turn (2) , if all the vertices have signs , the algorithm ends.

Definitions 2.2 When graph G (V, E) carried DFS search ,the first vertex is called the root vertex first began ; searching from the point the v along (v, w) side , nodes v called the father points w (using father (w) indicates the father of the node w), w called son node of v; when the w has nosign, (v, w) side called branches, when w has sign, (V, w) is called the back side edge.

Definition 2.3 TREE,BACK. TREE is defined as the set of branches side, BACK is defined as the set of back edges.

Definitions 2.4mark (v). mark (v) = 0 means that v point has not been searched , mark (v) = 1 means that v point has been searched.

Definitions 2.5 num (v). num (v) represents DFS number of point v ,that is, point v in the search process to be accessed in the order . If the edge (v_i,v_j) is the branch side, there $num(v_i)<num(v_j)$; if (v_i,v_j) is back, with $num(v_i)>num(v_j)$.
DFS algorithm steps are as follows :
(1) TREE ← ø, BACK ← ø, i ← 1,
v ∈ V (G) for
[father (v) ← 0, mark (v) ← 0] ;
(2) Choose one point r optionally satisfies the condition mark (r) = 0, for
[v ← r, mark (v) ← 1, num (v) ←i] ;
(3) If all edges associated with v points have signs , then turn (5) ; otherwise choose one side not sign (v, w) switch (4) ;
(4) to (v,w) side with direction from v to w, and give the sign* to show through checking;
If the mark (w) = 0, then for
[i ← i +1, num (w) ← i, TREE ← TREE ∪ {(v, w)}, mark (w) ← 1, father (w) ← v, v ← w, turn (3)] ;
If the mark (w) = 1, then for
[BACK ← BACK ∪ {(v, w)}, transfer (3)] ;
(5) If the father (v) ≠ 0, then for

[v ← father (v), turn (3)] ;
Otherwise, as,
[If v ∈ V (G) always have mark (v) = 1, then the algorithm ends ; otherwise for i ← i +1, turn (2)]
.

4. Network Segmentation Algorithm based on DFS

Large-scale network segmentation is the Division of a graph, traversing through the nodes on the network,the network can be divided to obtain sufficient to control the spread of the worm node.

Definitions 3.1 segmentation point. If a network-connected graph G(V，E) nodes in v (v ∈ v (G)), and v all associated edges deleted from the G,increase the connectivity number of blocks, then v is the dividing point, also called the backbone node.

Define 3.2 leaves. If v has one and only one side connected to the other vertex, v is said to be graph leaves. It is obvious that the personal computer is a leaf node in the network.

Define 3.3 graphpruning. Due to the leaf node is not backbone nodes in the network, so before network segmentation, you can remove the leaves, If the parent node of leaf nodes in figure of leaf cutting, degenerate into leaves, you can also delete it until there are no leaves can be removed so far, this process is called pruning, referred to as pruning. As shown in figure 3.

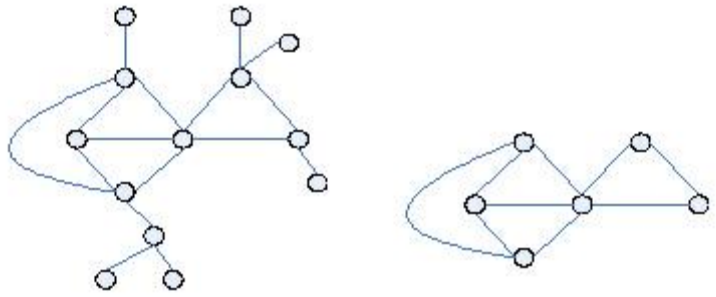

Figure 3. Pruning of Diagram

Definition 3.4 T (v). T (v) is reached by the v start up the DFS tree T to v "seed" node u, up through a back edge (u, w) has reached the point where w collection. v∈T(v).

Definition 3.5 low (v). Low (v) is the minimum number of DFS T (v) in the set of vertices.
That is $low(v) = \min_{w \in T(v)} \{num(w)\}$
The calculation steps of low (v)are as follows:
(1) When you first visit v point , so low (v) ← num (v);
(2) Through to the reverse side (v, w) , so
low (v) ← min {low (v), num (w)};
(3) The son node w of v search the exit from the stack , returning to the v -point , so
low (v) ← min {low (v), low (w)}.

The basic idea for segmentation points using DFS is:

On the figure of G(V, E) using DFS search method of all the vertices and edges of the inspection process, along information of the left low (v), when the search is over, the root node r is the dividing point, if and only if r has more than one son nodes; v(≠r)) is the dividing point, when and only when any son node w of v does not exist in the "seed" node (including w) to the v's "ancestors" nodes back edge.

From the foregoing, vertex v (≠ r) is the dividing point of G, if and only if v has child w, low(w) ≥ num(v). So, the algorithm steps of graph segmentation point by using DFS are as follows:

(1) on the graph G (V, E) by pruning;
(2) STACK←ø, i←1,
v∈V(G) as
【father(v)←0，mark(v)←0】;
(3) choose a point r satisfies mark (r) =0, as
【v←r，mark(v)←1，num(v)←i，low(v)←i】;
(4) If all v related sides have signs, then turn (6); otherwise, choose one of the edge (v, w) of no sign, add sign to the side (v, w), and added the (v, w) to the stack top STACK, turn(5);
(5) If mark (W) =0, then for
【i←i+1，num(w)←i，low(w)←i，mark(w)←1，father(w)←v，v←w，turn (4)】
If mark (w) =1, for
【low(v)←min { low(v)，num(w)}，turn (4)】;
(6) If father (v) =0, then the algorithm ends; otherwise as
Iflow(v) ≥ num(father(v)), as
[from the STACK shift side (father (v), v) and its elements and the stack to output, turn (7)];
low(father(v))←min {low(v)，low(father(v))}，v←father(v), turn (4).

5. VDU Model

VDU is defined as a collection of some virtual design resources. VDU is taken as the basic execute unit of design task in this paper. The model of VDU is described in the form of five-tuple.VDU={BasicInfo,Resource,DesignActivity,DesignAbility,Constrain}. Moreover, each tuple is explicated as following:

BasicInfo={ID,Name,BuidTime,VmuType,Position,Workshop,Status} denotes the basic information of VDU such ID, name, construction time, type, position, affiliated unit and status).

Resource={ResouceStructure,ResouceRelation} denotes the design resource and resource structure in the VDU. ResourceStructure={M,E,T}.M={m_1,m_2,m_3…}. ∨ m ∈ M, m denotes man resource, M≠0; E={C,Eq,S…} denote the equipment resources include computers, experiment apparatus ,simulation platform. ∨ c ∈ C, c denotes computer resource, C∩E≠0;

DesignActivity={Da_1,Da_2,…,Da_m} denotes the design activities that can be executed in the VDU. Da_m={DesignObject,DesignMethod,DesignActivity,DesignInput,DesignOuput} denotes the design activities attributes set.

DesignAbility={DA_1,DA_2,…,DA_n} denotes the design abilities that VDU takes engage in different tasks. DA_n={DesignActivityID,SuccessRatio,SkilledDegree,Robust, DesignQuality,DesignCost,DesignTime} denotes the design ability attributes.

Constrain={Con_1, Con_1,…, Con_x} denotes the constrains restrain the design activity of VDU.

As the basic design task execute unit, each VDU can engage in at least 1type of design task. In the design activity, designer, computer equipment, simulation apparatus, software, model, tool and knowledge are the mainly factors that can affect quality of design

activity. Inside, all kinds of designers are the dominant body. Designers engage in design, management and maintenance by operating all kinds of hardware and software. So designer is the indispensable element in VDU. In addition, main design computer is as important as designer. Design resources in the VDU connect together by main design computer with network. In the level of atom-resources, main design computer is used to register resources virtual information, accept design task, communicate among atom-resources and collect task information. In the level of VDU, it takes charge the communication and collaboration among different VDUs. Structure diagram of VDU is shown as Figure 4.

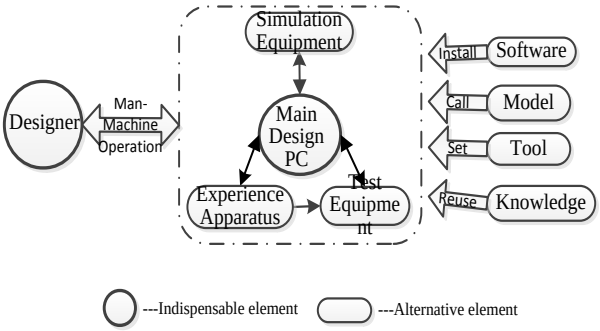

Figure 4. VDU Structure Diagram

6. Conclusion

Due to large space demanding and time-consuming，as the Internet is large and complex, the detection and prevention of the worm has many engineering constraints factors, the classic graph theory algorithms cannot suitable for solving some problems in the large-scale network. Sub-graphs of vertices have a higher density of edges within them while a lower density of edges between sub-graphs. By analyzing the study of graph depth-first search (Depth First Search, referred to as DFS) segmentation algorithm, and thus obtain a description of the network connectivity of undirected connected graph G (V, E).Results show that the segmentation method based on network detection method is a fast and effective method, the size of virus detection sub-graph is moderate and suitable for calculation algorithms applying in the flow path search algorithm, classic graph algorithm.it is also advantageous for the mass virus transmission to network analysis.

References

[1] T Y Lin, “Mining associations by linear inequalities”, Proceedings of International Conference on Data Mining, Washington：IEEE Computer Society (**2004**).
[2] T Y. Lin, “Granular computing: Fuzzy logic andrough sets”, Zadeh I. A., Kacprzyk J. “Computingwith words in information / intelligent systems”, PhysicaVerlag (A Springer-Verlag Company)(**2009**).
[3] D J. Watts and S H. Strogatz, “Collective dynamics of small world networks”, Nature, vol. 4, (**2008**), p. 393.
[4] A. Barabdsi and R. Albert, “Emergence of scaling in random networks”, Science (**2009**).
[5] S. Singh, C. Estan and G. Varghese, “Automated Worm Fingerprinting”, Proceedings of the 6th Symposium on Operating System Design and Implementation(OSDI), USENIX (**2004**).
[6] D. Moore, V. Paxson and S. Savage, “Inside the slammer worm”, IEEE Magazine of Security and Privacy, vol. 4, no. 1, (**2003**).
[7] C. Zoua, D. Towsleyb and W. Gongc, “On the performance of Internet worm scanning strategies”, Performance Evaluation, (**2006**).
[8] H.-A. Kim and B. Karp, “Autograph: Toward Automated, Distributed Worm Signature Detection”, Proceedings of the 13th USENIX Security Symposium (**2004**).

Author

Yuqi Tang, male; birth year: 1983; ethnic group: Han; native place: Luannan, Hebei; educational background: on-job postgraduate; academic degree: master; professional title: lecturer; place of work: Tangshan Vocational & Technical College; main research direction: computer application.