PROCEEDINGS OF SCIENCE

SMCSN: A New Secure Model of Content Sharing Network by Using Multi-roles Sybil Nodes

Qiang Lu[1 2]
School of Computer Science, National University of Defense Technology, Changsha, 410073, China
E-mail: luqiangjfj@163.com

Bo Liu
School of Computer Science, National University of Defense Technology, Changsha, 410073, China
*E-mail:*kyle.liu@aliyun.com

Huaping Hu
School of Computer Science, National University of Defense Technology, Changsha, 410073, China
*E-mail:*hphu@nudt.edu.cn

Tianzuo Wang
School of Computer Science, National University of Defense Technology, Changsha, 410073, China
*E-mail:*phoenixwtz@163.com

Considering the serious security trend of Content Sharing Networks (CSNs) and the advantages of Sybil nodes, a new Secure Model of Content Sharing Network by using multi-roles Sybil nodes (SMCSN) is hereby proposed. In SMCSN, we introduce three different kinds of Sybil nodes and discuss some issues of exploiting Sybil nodes like infiltrating collision, hotspots sensing and importance improvement, etc.. Further simulation experiments and analysis indicate the effectiveness and feasibility of SMCSN to enhance the security of CSNs.

CENet2015
12-13 September 2015
Shanghai, China

[1]Speaker
[2]Correspongding Author

PoS(CENet2015)058

1. Introduction

Nowadays, it is prevailing to share various contents (e.g. files, information and applications) via network; however, numerous malicious files and spy applications have swarmed into CSNs, which has posed great threat to the network security.

Speaking of Sybil nodes, attacks will emerge in mind firstly. Researchers mainly concentrate on detection, mitigation and elimination of Sybil nodes. Cai Z et al. used statistical methods and learning algorithms to detect Sybil nodes [1]. Mohaisen A et al. surveyed three main methods to defend Sybil attacks, namely the trusted certification, the resources testing and the social networks mitigation [2]. Most elimination solutions were combined with detection and mitigation to defend Sybil nodes [3]. As far as I know, few articles have paid attention to positive effects of Sybil nodes; nevertheless, absolute elimination has been proved impossible by Douceur J R for many years [4]. Sybil nodes can also do well in improving the security of CSNs just like a magic poison.

Recently, Sybil nodes gradually show some positive impacts such as mitigating P2P (Peer-to-Peer) botnets [5] and controlling message propagation in OSNs (Online Social Networks) [6], etc.

2. Related Works

In this paper, we specialize CSNs to P2P-based sharing networks for their wide effects on content sharing. Despite a few initial adoptions, Sybil nodes are just applied to cope with partial security problems. To the best of our knowledge, a secure model of content sharing network has never been considered before.

Wang T et al. made a trial to use some Sybil nodes with special settings to infiltrate P2P botnets, mainly to break their C&C (Command and Control) mechanisms [7]. It was a good solution to defend some activities of botnets like commands distribution, but it neglected the increasing malicious sharing contents.

In order to mitigate attacks from P2P botnets, a proper number of Sybil nodes can get a good control of the processes of malicious information searching and publishing among nodes [8]. From [8] we can get a clear understanding about the discriminable influences on network security as to different percentages of Sybil nodes. Whereas different roles of Sybil nodes are important to network security as well, less attention has been yet paid.

In order to enhance the security of CSNs, we will introduce some Sybil nodes with different roles for actual deployments in Kad (a typical CSN, mainly in eMule) [9].

3. Secure Model of Content Sharing Network (SMCSN)

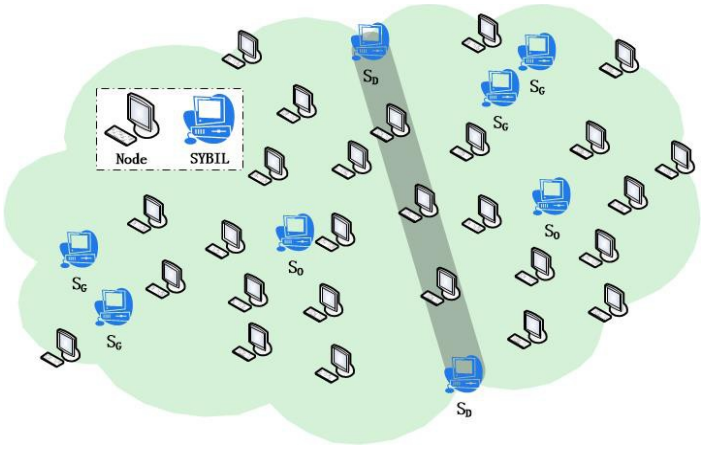

Figure 1: SMCSN with Multi-roles Sybil Nodes

Kad, based on the typical DHT (Distributed Hash Table) protocol Kademlia [10], is playing an important role in contents sharing. In Kad network, each node has a fixed 128-bit DHT ID (identifier). The routing table of every node consists of L (L equals the number of node ID's bits and is usually given a value 128) lists, and each list is called a "K-bucket". A "K-bucket" is used to store the information of neighboring nodes with the tri-tuple <IP address, UDP port, node ID>. As eMule is a main implement of Kad, it is proper to take it for example to present SMCSN, as shown in Fig.1.

3.1 Roles of Sybil Nodes

There are three kinds of Sybil nodes: S_G (Guard nodes), S_O (Observing nodes), S_D (Dam nodes).

1) Guard nodes

Obviously, the guard nodes are similar to those who assure the safety of a special system like a campus building; however, such nodes have more functions than the so-called gatekeepers. Generally speaking, these nodes are in charge of the normal nodes in a LAN (Local Area Network) and maintain information of them. Here the special LAN is called the guarding subnet shown in Fig.2. A pair of guard nodes with the same ID is deployed to fulfill the guarding function. Guard nodes can be added to form a triple or quadruple according to the subnet size. One of them is the captain with others as the spares and assistants. They work in collaboration to control, register and audit the importing sharing contents with a database; besides, a guarding node can serve as a CA (Certificate Authority) of the subnet for nodes' identity authentication.

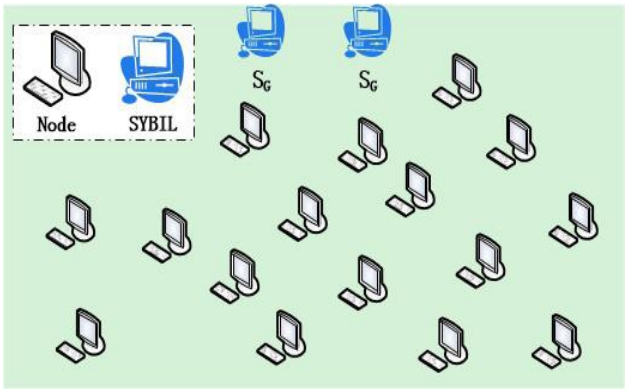

Figure 2: the Guarding Subnet

2) Observing nodes

Sybil nodes with the role of observation are outstanding in capability of infiltration, generally located in the key path of Kad network. To achieve the best observing effects, the location of observing nodes is chosen according to the clustering degree of selected domain. In the domain, the observing nodes attain a high in-degree just like the center of a circle and infiltrate information into routing tables of other nodes as soon as possible. The domain is named the observing subnet as shown in Fig. 3. As the center, the observing nodes are in possession of not only high capability of infiltration, but also absolute power to control and conduct the routing and sharing information propagation. They can apperceive the sharing hotspots easily and accurately at a quick speed.

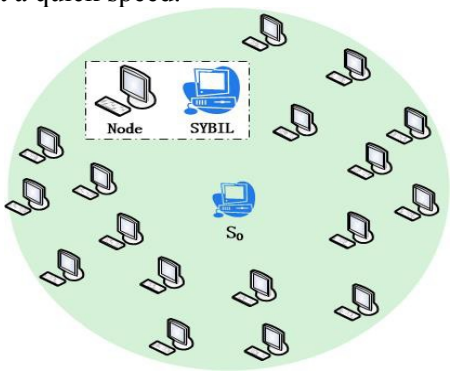

Figure 3: the Observing Subnet

3) Dam nodes

Similar to the significant Three Gorges Dam, the deployment of dam nodes is crucial to the CSNs. From Fig. 1 above, we can draw that the dam nodes are related with the information transmission between subnets. In my opinions, we have attached too much importance to the sharing function and the dam nodes just add up some complementary security restricts. Since the inherent equality, anonymity and openness of P2P, a massive amount of sharing contents are pouring into the network. It is not sensible to share all kinds of contents without any limitation. Given malicious attacks, the dam nodes are very important and necessary as well as guard ones and observing ones.

3.2 Key issues

In SMCSN, in addition to a proper number of multi-roles Sybil nodes, there are some key issues that cannot be ignored among those nodes, shown as follows:

1) Infiltrating collision

Require:

LCrawler

Basic Steps:

1. Divide infiltration domain;
2. Organize Sybil nodes into groups;
3. Distribute IDs to Sybil nodes;
4. Crawl neighboring nodes with LCrawler;
5. Generate and distribute infiltration lists;
6. Infiltrate in a parallel way with a cycled condition **goto** 4.

Figure 4: the Method of Solving the Infiltrating Collision Problem

Naturally, there will be some collisions in the process of infiltration among Sybil nodes. For example, the simultaneous online nodes of eMule (only Kad network) can be always up to one third of a million according to the tests with our own eMule crawler (LCrawler). The collisions are not unavoidable. Here we propose a simple and feasible method. The basic process is shown in Fig. 4. In order to organize the Sybil nodes into groups, the number of Sybil nodes is noted as N shown in (3.1); thus α must be a certain power of 2 like 64. Then the number of both Sybil groups and infiltration domains will be set as α. The distribution of IDs follows the XOR operation with details described in Kademlia protocol [10]. The cycled conditions need a period T and several timers.

$$N = \alpha * \log_2 \alpha \tag{3.1}$$

2) Hotspots sensing

The malicious sharing contents will probably be hotspots during the fast propagation; thus hotspots sensing is vital to Sybil nodes in SMCSN. It is not necessary to make all Sybil nodes to sense the hotspots of CSN at a high cost. A certain amount of Sybil nodes will be just fine. Here we use M, S to note the number of required Sybil nodes and the total amount of eMule nodes respectively. Then M can be expressed with (3.2). For example, when S equals one million, the value of M is 20. The selected Sybil nodes form a set named SNS (Sensing Nodes Set). For every member of the SNS, a few additional function modules are needed, such as setting parts for sensing period named ST and hotspot threshold value named TH, distribution part for searching records and reporting part; and they will share a hotspot list created with a default initialization. Given the perfect clustering trait, the observing nodes are usually selected to form the SNS. The sensing method is shown in Fig. 5.

$$M = \lceil \log_2 S \rceil \tag{3.2}$$

Require:

Hotspot list

Basic Steps:

1. Create SNS, hotspot list;
2. **for** i=0; i<M; i++ **do**
3. set the sensing period and hotspot threshold value;
4. **End for**;
5. Sense and record the searching information;
6. Judge and report the hotspots;
7. Update the hotspot list with a cycled condition **goto** 5.

Figure 5: the Method of Sensing Hotspots

3) Importance improvement

When we say the importance of Sybil nodes, the in-degree is a main evaluation parameter. The greater the in-degree of a Sybil node is, the more neighboring nodes and higher probability of infiltration will have. Then the Sybil node will be emerged in routing tables of those neighboring nodes and attain more content sharing information including malicious activities. In order to make the in-degree value bigger, the fast variability of the nodes' status must be taken into account. We have understood the dynamic character of network well from crawling eMule Kad network with LCrawler. Upon contemplation, we've found a cycled filtration method to overcome the infections brought by the dynamic Kad network; besides, a principle is proposed to make Sybil nodes prior to the neighbors of nodes with a far distance or great activity.

3.3 Collaboration of Different Sybil Nodes

Apart from the roles of Sybil nodes and some key issues as discussed above, the collaboration of different Sybil nodes is another emphasis. As an indispensable part of the SMCSN, it has a compact relation with other two parts of Section 3. The collaboration is a complex mechanism problem and here we just offer some of our own ideas: the guarding nodes and observing ones can work together to make the overlap part of subnets have a higher probability to avoid malicious pollution and propagation; the cooperation between observing nodes and dam will play a significant role in the mitigation of two-way transmission of malicious sharing contents; the power and impacts, brought by the collaboration of three kinds of Sybil nodes, will be much more than treble of those different Sybil nodes work independently.

4. Experiments and Analysis

In order to validate SMCSN, we carry out some simulation experiments and relevant analysis. The experimental CSN is set in reference to a subnet of eMule Kad network with one tenth million nodes. Namely, the value of S is 100,000, and then M equals to 17 according to (2). Values of other parameters are listed in Table 1.

Parameter name	Meaning of each parameter	Parameter value
α	the number of Sybil nodes' groups/infiltration domains	32
N	the total amount of the Sybil nodes	160
T	infiltrating period	1 hour
ST	sensing period	1 hour
TH	hotspot threshold value (according to real sharing activities)	25,000

Table 1: Values of Some Parameters

4.1 Simulation tool

In view of the large scale of CSN like eMule (Kad), our experiments are carried out with a modified simulation tool based on PeerSim [11] and the protocol of Kademlia implemented by Furlan and Bonani [12]. The modifications are concentrated on the Kademlia protocol and necessary content-sharing functions. Some associated settings about simulation are listed in Table 2, in which, the event-driven mode is selected from the two simulation modes on PeerSim: event-driven and cycle-based. Experiments are carried out with three different initiations respectively (2%, 5%, and 10%).

Setting Items and the Meanings	Setting Values
the mode of simulation	event-driven
the repeated times of simulation	50
the least content sharing activities of every node	2
the initial proportion of malicious sharing contents	{2%, 5%, 10%}

Table 2: Associated Settings of Simulation

4.2 Experimental Results and Analysis

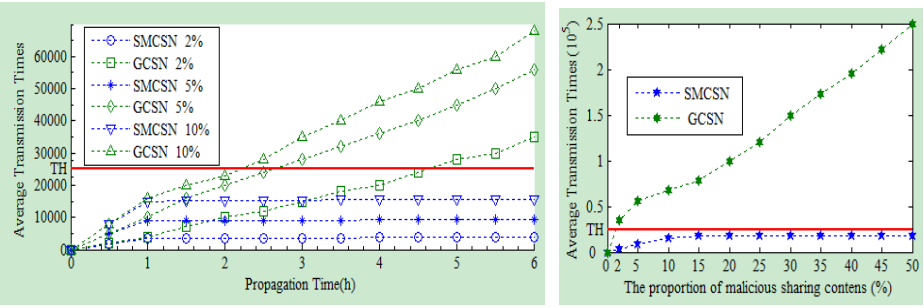

Figure 6:a) Results of Six Group Experiments **b)** Results of Supplementary Experiments

We will compare the security of SMCSN with that of general CSN (GCSN). The current focus of CSN security is mainly on malicious nodes and malicious sharing contents. The former has been proved in some papers [6] [7] [8], but the latter is still not be tested. As there are three different initiations for malicious sharing contents, the experiments will be 6 groups in all. For every group, the simulation tool runs 50 times to get the average transmission times of malicious sharing contents. The experimental results are shown in Fig.6 a). Evidently, the malicious propagation in SMCSN is slower especially after the first infiltrating period (1 hour). It has almost nothing to do with the initial proportion of malicious sharing contents. Then we do some supplementary experiments to continue a further comparison by gradually increasing the initial proportion of malicious sharing contents up to 50% (a maximum according to our monitoring on some CSNs). Getting the average transmission times of malicious sharing contents at the sixth period of propagation (6 hours), we show our results in Fig. 6 b). In our new secure model, malicious sharing contents can be suppressed effectively and they are nearly impossible to become hotspots.

5. Conclusion

This paper proposed a new secure model of content sharing network with deploying different kinds of Sybil nodes like guard nodes, observing ones and dam ones over eMule Kad network. From simulation results, we can conclude that our new model is effective to enhance the security of CSNs and mitigate the propagation of malicious sharing contents. The future work will consider the actual deployments of the SMCSN, the collaboration mechanism of different Sybil nodes, more Sybil roles and other issues in terms of mitigation of malicious sharing contents.

References

[1] Z. H. Cai, C. Jermaine. *The latent community model for detecting sybil attacks in social networks*[C]//Proc. Very Large DataBase Endowment 2011, Seattle, ACM, 2011:1-9

[2] A. Mohaisen, J. Kim, *The Sybil Attacks and Defenses: A Survey* [J]. Cryptography and Security, 2013, 3(6): 480-489

[3] S. Krishnaveni, A. V. S. Kumar, *A Survey on Defense Mechanism for Sybil Attacks in Large Social Networks* [J]. International Journal of Advanced Research in Computer Science, 2014, 5(1): 105-110

[4] J. R. Douceur. *The Sybil attack* [M]//Peer-to-peer Systems. Springer Berlin Heidelberg, 2002: 251-260

[5] T. Holz, M. Steiner, F. Dahl, E. Biersack, F. Freiling. *Measurements and Mitigation of Peer-to-Peer-based Botnets: A Case Study on Storm Worm* [J]. Larger-scale Exploits and Emergent Threats, 2008, 8(1): 1-9

[6] T. Z .Wang, H. M. Wang, B. Liu, B. Ding, J. Zhang, P. C. Shi. *Further Analyzing the Sybil Attack in Mitigating Peer-to-Peer Botnets* [J]. KSII Transactions on Internet & Information Systems, 2012, 6(10):2731-2749

[7] T. Z. Wang, H. M. Wang, B. Liu, H. Ren, X. L. Ma. *A study of strategies to restrain the C&C activities of structured P2P botnets*[C]//Computing and Convergence Technology (ICCCT), 2012 7th International Conference on. IEEE, Seoul, 2012: 537-542

[8] T. Z. Wang, H. M. Wang, B. Liu, P. C. Shi. *Model the Influence of Sybil Nodes in P2P Botnets [M]//Network and System Security.* Springer Berlin Heidelberg, 2013: 54-67

[9] M. Steiner, T. En-Najjary, E. W. Biersack, *A global view of Kad*[C]//Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, New York, 2007: 117-122

[10] P. Maymounkov, D. Mazieres, *Kademlia: A peer-to-peer information system based on the xor metric* [M]//Peer-to-Peer Systems. Springer Berlin Heidelberg, 2002: 53-65

[11] A. Montresor, M. Jelasity, *PeerSim: A scalable P2P simulator*[C]//Peer-to-Peer Computing. P2P'09 IEEE Ninth International Conference, Seattle , IEEE, 2009: 99-100

[12] http://peersim.sourceforge.net/code/kademlia.zip [Z], 2013(4)