IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Simulation of Semi Markov Process to Detect Mimicking Attacks Based On User Behavior

M.Anitha[1], A.Kanchana[2], R.Padmapriya[3], N.Malathi[4]

Computer Science and Engineering, Panimalar Engineering College, Chennai, India[1,2,3,4]

Abstract: Botnets have turn out to be a most important engines for malicious activities in cyberspace these days. Botnets are the major drivers of cyber attacks, such as distributed denial of service (DDoS), flash crowds, email spamming and information phishing. Both flash crowds and DDoS attacks have extremely related properties in terms of internet traffic. Flash crowds are legal flows whereas DDoS attacks are illegal flows. To maintain their botnets, botnet owners are mimicking valid cyber behavior. This poses a critical confront in anomaly detection. In this work, study of mimicking attacks and detections from both sides, as attackers and defenders is made. First of all, a semi-Markov model for browsing behavior is recognized. Based on this model, a botmasters can create flash crowd effectively in terms of statistics, with a adequate number of active bots(not less than the number of active valid users). But it is hard for botnet owners to gratify the situation to carry out a mimicking attack most of the time. With this new finding, we conclude that mimicking attacks can be discriminated from real flash crowds using second order statistical metrics. When the adequate number condition does not hold for botmasters we detect the mimicking attacks. Detection is proclaimed to the user. Furthermore, the findings can be widely functional to related situations in further research fields.

Keywords: detection; flash crowd attack; mimicking; second order metrics;

I. INTRODUCTION

The development of World Wide Web (WWW) is making it the standard information system for an increasing sector of the world's population. The Internet was initially designed for openness and scalability. For example, the Internet Protocol (IP) was designed to sustain ease of attachment of hosts to networks, and provides little support for verifying the contents of IP packet header fields. This makes it possible to fake the source address of packets, and hence difficult to identify the source of traffic. In addition, there is no natural support in the IP layer to ensure whether a source is allowed to access a service. Packets are delivered to their destination, and the server at the destination must choose whether to accept and service these packets. A denial of service (DoS) attack aims to deny access by valid users to shared services or resources. This can happen in a large range of contexts, from operating systems to network-based services. When the traffic of a DoS attack comes from multiple sources, it is called a distributed denial of service (DDoS) attack [1].
On the other hand, these days Botnets are the main drivers of cyber attacks, such as distributed denial of service (DDoS), anti attacks and information phishing. Additional examples of mimicking attacks, such as membership recruitments of botnet, email spamming , ,etc., The term *bot* (derived from the word *robot*) is used in industry terminology to depict an machine or automated process in both the real world and the computer world. A bot normally supports a communication channel with the attacker, as well as the ability to execute particular tasks, for example, launching mimicking attacks, according to the attacker's instructions [1].Well experienced attackers usually simulate the phenomenon of flash crowds to halt intrusion detection systems (referred to as a flash crowd attack) [6], [7].

Discriminating flash crowd attacks from legitimate flash crowds has been explored for approximately a decade. Previous work [8]–[10] has paying attention on extracting DDoS attack features, followed by detecting and filtering DDoS attack packets using the known elements. However, these strategies cannot actively detect DDoS attacks. The current popular defence against flash crowd attacks is the use of graphical puzzles to distinguish between humans and bots [11]. This method involves human responses and can be annoying to users. From the botnet programmer's perspective, in order to simulate the legitimate behavior of a web browser, we require three key pieces of information: web page popularity of the target website, website page asking for time interval for a user, and number of pages a user usually browses for one browsing session (referred to as browsing length).

This paper endeavor to reveal that legitimate cyber behavior can be be effectively simulated, accordingly, it is impractical, to distinguish mimicking attacks from legitimate cyber events using statistical methods. However, in order to attain this, attackers have to acquire a adequately large number of active bots, than the number of active legitimate users. In view of the study of mimicking attacks, we found four parameter semi-Markov model to characterize browsing behavior. Using this model, we effectively simulate browsing behavior of victim client. We discover that the first order statistical metric does not provide our discrimination task, and the traditional second order metric (e.g. the standard deviation) is not good enough as far as detection granularity. Hence we invent a new second order statistical metric based on the traditional correntropy to provide the detection tasks with fine detection accuracy.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

II. ANALYSIS

In 2007,TAO PENG investigated and build up that, interestingly to direct attacks, indirect attacks can exploit insecure actions that may be performed by real clients. These attacks generally require human interaction.

A. Traffic Model Analysis

In 2009, Ke Li, Wanlei Zhou gave an analysis of traffic model. Firstly, Flash crowds and DDoS attacks are very related in traffic behavior from macroscopic observation; though there are also several essential differences in the aspects of access intents, distributions of source IP address and speed of the increased and decreased traffic. Flash crowds are the results of the legitimate users respond to special events such as breaking news or popular products (movies, music and software) release. All the users just want to acquire the information or material quickly from the server. If the server is slowed down they will even shut down. However, DDoS attacks are not social events and all the requests are launched by attackers and are illegitimate. Secondly, the distributions of the source IP address are also quite different between Flash crowds and DDoS attacks [2].If we combined the IP addresses of flash crowd attack, the distribution of source IP addresses will be subject to the fractional Gaussian noise distribution [5] However, If we aggregate these source the distribution of source IP addresses of DDoS attack, it will subject to the Poisson distribution [4]. Thirdly, there is a big difference in the increased and decreased speed of traffics between them.

B. Web Browsing Dynamics

Breslau et al. analyzed web accessing behavior and found that page popularity follows the Zipf-like distribution [26]. A general form of the popularity distribution is called the Zipf-Mandelbrot distribution [13]. These findings are widely used in research papers, such as [17] and [18].For a given website, if all the bots of a botnet requests page based on the Zipf-Mandelbrot law,we will not be able to recognize which ones are attack requests. Therefore, attackers can easily disable statistics based detection algorithms using this strategy in their bot programs.

Crovella and Bestavros found that viewing time distribution on web pages follows the Pareto distribution [14] (confirmed also by [19] and [20]). This information is very useful for botnet writers. Once a browsing page has been decided, a bot submits the page request to the victim and downloads the page to the host computer without displaying it (e.g. discarding it or depositing it to the cache). When the requested page has been downloaded, the bot decides a “reading" time interval following the Pareto distribution before requesting another web page.

The last component for browsing dynamics is browsing length namely the number of pages a user generally views during a browsing session. Huberman et al. demonstrated that the probability follows the two – parameter inverse of guassian didtrubution. This information can be engaged by botnet writers to choose how many pages to request for a bot, or else, the defender may become aware of that many “clients” have a long browsing length, and thus detect the attack. This fact services bot-masters to possess a sufficient number of active bots to carry out flash crowd attacks.

C. Mimcking attack detection

Shui Yu in 2012 gave a mechanism for attack detection. If the sufficient number condition holds for a botnet owner, then botmaster can perfectly simulate a cyber event such as mimicking attacks. He took World CUP 98 data set [12] and the Auckland data set [13],and based on research [3], concluded that the number of active bots of a botnet is usually only at the hundreds or a few thousands level. Therefore, in order to carry out a flash crowd attack, the sufficient number condition is hard to meet. He demonstrated the effectiveness with two examples such as the Gaussian distribution and the Poisson distribution as both of them are typical and widely used for network traffic modeling.[16]

III. PROPOSED SYSTEM

Mimicking attacks and detections are scrutinized both as botmasters and victim clients. From the botnet developers(botmaster) point of view, regarding simulate the legitimate behavior of a web browser, we require three information such as popularity of target website, time interval between two successive request for a web page, number of pages the browser usually browses for one browsing session.

If attackers have a adequate number of active client nodes, then each bot can simulate one genuine user avail these three statistical distributions. Although, it is hard for botnet owners to set up the adequate number condition for definite mimicking attacks.

Following are the contributions of our paper

1. We demonstrate that botmasters can simulate a flash crowd successfully in terms of statistics. With a adequate number of active bots, a botmaster can make use of one bot to simulate one legitimate user with the information of web browsing dynamics.
2. We establish four parameter semi-Markov model to differentiate browsing behavior. Using this model, we can efficiently simulate browsing behavior, and so can productively begin a cyber event.
3. Also we intend a new second order statistical metric for the use of detection. We as a result invent a new second order statistical metric based on the traditional correntropy to give the detection tasks with fine detection accuracy.
4. Therefore the attack is detected and the event is proclaimed to the user.

A. Browsing behavior tracking

We count the amount of HTTP requests of every flow for the given time intervals and to illustrate the browsing behavior of a legitimate web viewer or user. This preparation should be taken occasionally to update the parameters to replicate the ever altering web browsing behavior. To depict the browsing behavior of a legal web

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

viewer, we develop the classical Markov model to a four parameter semi-Markov model as follows,

$$\Delta = (S, M, l, \pi)$$

where (S, M, l, π) the state transition matrix, duration at the current state, browsing length, and the initial probability distribution of the states, respectively.

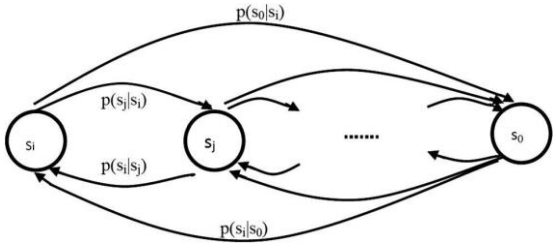

Fig.1. State transition

The state transition is shown in Fig. 1. For a given point of time, we expect to know the number of total page requests to a web site, and number of requests for a particular web page of the web site.

We need one more parameter: the number of active web viewers for a specified time point t, which we represent as m(t). m(t) varies adjacent to the time point of a day. Naturally, there are more web viewers in the course of working time than early morning.

B. Implementation of Mimicking Attack

In this part, using the composed details about victim, the botmaster will effectively produce mimicking attack. If we carry out any modification in the botmaster side, it will automatically get reproduce in the victim client page. Examine the target website for extracting all the markov model parameters and initialize these parameters. Decide illustration set of bots s, from the set of active bots S and initiate these bots to run separately. Generate a random number rm. As per the client browsing behavior tracking reviewed earlier in this paper, first choose the initial page .Next choose the browsing length L of the current bot based on clients browsing history. When the browsing length is within L, request a page and download the content.

C. Attack Detection

After analyzing the client response from the server, we can able to detect the mimicking attack. A bot has to make many more requests compared to a legitimate browser for a given time interval in order to generate the same number of requests to the web site, and therefore, the standard deviation of the attack flow is much smaller than that of legitimate browsers'. Therefore, we can differentiate them. In general, we can use any second order statistical metric to carry out the detection task. The only differentiation is the accuracy of the result, which depends on the granularity of the metric. Mimicking attacks can be detected using the standard deviation under the condition that the sufficient number condition is not detained for attackers.

On the other hand, there exists a problem of how precisely we detect mimicking attacks. Accuracy depends on the metric that we prefer. In this paper, we have to utilize second order statistical metrics. There are many candidates, such as the standard deviation, or the traditional correntropy. Though, in our experiments, we establish that both of them are not as good as we expected, therefore, we proposed a new second order metric based on the correntropy. Correntropy is a recently made-up local tool for second-order similarity measurement in statistics. It works separately on measuring pair-wise arbitary samples.

Correntropy metrics are symmetric, positive, and bounded. For any two finite data sequences A and B, assume we have sample $\{(A_j, B_j)\}_{j-1}^{m}$, $m \in$ N, then the similarity of the sequences are estimated as,

$$C_{m,\sigma} = \frac{1}{m} \sum_{i=1}^{m} k_\sigma (A_i - B_i)$$

where $k(\cdot)$ is the Gaussian kernel, which is usually defined as follows.

$$k(\cdot) = \exp\left(-\frac{x}{2\sigma^2}\right).$$

Algorithm for discriminating Mimicking Attack

1.Monitor the clients number of page request for a 24 hour period. Denote it C(t)

2. Establish a mapping of the difference of flow fine correntropy of page request flows against C(t) and denote as Tf(n(t))

3. while {true}do

Monitor the number of page request of current website. Denote it as C'(t).

While C(t)<C'(t)

a. Collect the sample points of request flows based on our statistics methodology

b.Claculate the flow fine correntropy C'(t)

c.$\Delta C(t) = C(t) - C'(t)$

if $\Delta C(t)$ is minimum then it is detected as mimicking attack

else

do nothing

end

end

end

IV.EXPERIMENT RESULTS

We have done an experiment by designing both server side and the attacker side. At server side we designed a web page to supply service for uploading image files.

The client can login into the server after registering and can upload the files as shown in Fig.1

We also designed an attacker page with different attacks as shown in Fig.2.

The attacker will attack the server, intruding as legitimate user based on the browsing behavior of genuine user. We performed attacks such as mimicking attack, flash crowd attack, DDos attack and phishing attack.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Fig.1 Client Login Page

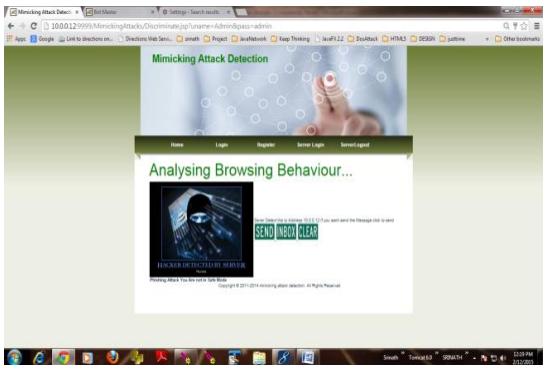

Fig.2 Attack Detection

At server side we analysed the browsing behavior of genuine client as per statistical method and Current behavior. The similarity measure was calculated by applying fine correntropy. We also calculated based on standard deviation and other metrics. The result is shown in fig 3.

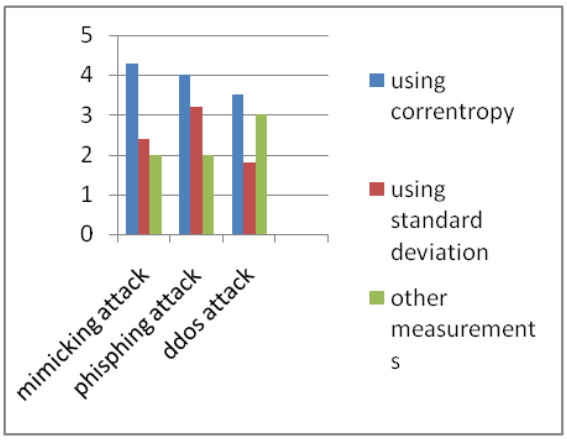

Fig.3 Similarity measure

V. CONCLUSION

Detection of legitimate mimicking attacks is taken as a most important analysis in this paper. We have established markov process model to simulate the browsing dynamics of genuine web browsers. The theoretical analysis and real world data experiments demonstrated that we cannot identify this kind of simulation in statistics. However, there is a significant condition for a successful mimicking attack. That is, the number of active bots of the botnet must not be lower than the number of active genuine users. We find that it is impossible for botnet owners to satisfy this sufficient number condition in the case of performing large scale attacks. Based on this new finding, we proposed a second order statistics based differentiation algorithm to detect this kind of attack. We done theoretical analysis and confirmed the effectiveness of the proposed detection method.

VI. FUTURE WORK

Mimicking attacks such as membership recruitment, performance degradation attacks etc can be performed in networks which have less number of users. We can address this kind of problem by finding new methodologies. Botnet owners can also interact with other botner owners to establish a super botnet for satisfying the sufficient number condition to execute mimicking attacks. We can analyse this kind of attacks in future.

REFERENCES

[1] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of network-based defense mechanisms countering the DOS and DDoS pro-blems," ACM Comput. Surv., vol. 39, no. 1, 2007.

[2] Ke Li, Wanlei Zhou, Ping Li, Jing Hai and Jianwen Liu," Distinguishing DDoS attacks from flash crowds using probability metrics" 2009 Third International Conference on Network and System Security

[3] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "My botnet is bigger than yours (maybe, better than yours): Why size estimates remain challenging," in Proc. 1st Conf. Workshop Hot Topics Under-standing Botnets (HotBots'07), 2007.

[4] W. Willinger. Traffic modeling for high-speed networks: Theory versus practice. In Stochastic Networks. 1995: Springer-Verlag

[5] S. Ledesma and D. Liu, Synthesis of fractional gaussian noise using linear approximation for generating self-similar network traffic. Computer Communication Review, 2000. vol.30.

[6] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-Gaussian and long memory statistical characterizations for internet traffic with anomalies," IEEE Trans. Dependable Secure Comput., vol. 4, no. 1, pp. 56–70, Jan./Mar. 2007.

[7] A. El-Atawy, E. Al-Shaer, T. Tran, and R. Boutaba, "Adaptive early packet filtering for protecting firewalls against DOS attacks," in Proc. IEEE Conf. Comput. Commun. (INFOCOM), 2009.

[8] J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash crowds and denial of service attacks: Characterization and implications for CDNS and web sites," in Proc. World Wide Web (WWW), 2002, pp. 252–262.

[9] G. Carl, G. Kesidis, R. Brooks, and S. Rai, "Denial-of-service attack-detection techniques," IEEE Internet Comput., vol. 10, no. 1, pp. 82–89, Jan./Feb. 2006.

[10] Y. Chen and K. Hwang, "Collaborative detection and filtering of shrew DDoS attacks using spectral analysis," J. Parallel Distrib. Comput., vol. 66, no. 9, pp. 1137–1151, 2006.

[11] S. Kandula, D. Katabi, M. Jacob, and A. Berger, "Botz-4-sale: Sur-viving organized DDoS attacks that mimic flash crowds (awarded best student paper)," in Proc. Symp. Netw. Syst. Des. Implement. (NSDI), 2005.

[12] S. Yu, S. Guo, and I. Stojmenovic, "Can we beat legitimate cyber behavior mimicking attacks from botnets," in Proc. IEEE Conf. Comput. Commun. (INFOCOM), 2012, pp. 3133–3137.

[13] Z. K. Silagadze, "Citations and the Zipf-Mandelbrot's law," Complex Syst., vol. 11, p. 487, 1997.

[14] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," IEEE/ACM Trans. Netw., vol. 5,no. 6, pp. 835–846, 1997.

[15] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in Proc. IEEE Conf. Comput. Commun. (INFOCOM), 1999,126–134.

[16] S. Yu, G. Zhao, S. Guo, Y. Xiang, and A. Vasilakos, "Browsing behavior mimicking attacks on popular websites," in Proc. IEEE Conf. Comput. Commun. (INFOCOM) Workshops, 2011.

[17] A. Klemm, C. Lindemann, M. K. Vernon, and O. P. Waldhorst, "Characterizing the query behavior in peer-to-peer file sharing systems," in Proc. 4th ACM SIGCOMM Conf. Internet Meas., 2004, 55–67

[18] M. Hefeeda and O. Saleh, "Traffic modeling and proportional partial caching for peer-to-peer systems," IEEE/ACM Trans. Netw., vol. 16, no. 6, pp. 1447–1460, Dec. 2008.

[19] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," Internet Math., vol. 1, 2004.

[20] W. J. Reed and M. Jorgensen, "The double pareto-lognormal distribution—A new parametric model for size distributions,"Commun. Stat. Theory Methods, vol. 33, no. 8, pp.1733-1753,2003.