A Note on Statistically Detecting Tampered Type Attacks

Ming Li [1] and Wei Zhao [2]

***Abstract*—** Information integrity of communication data is crucial to e-business systems. An adversary may delete or insert packets into a normal communication data series in e-business systems to destroy information integrity to produce deserters (likely informal, this paper uses the term tamped type attacks to specifically describe those that attackers detect or insert packets into a normal data series). Consequently, a security issue in an e-business system is how to detect tamped type attacks.

From a view of intrusion detection systems, statistic methods of intrusion detection are increasingly paid attention to [1-3]. This short paper gives a method for statistically real-time detection of tampered type attacks. The method is based on power spectra of monitored traffic time series with finite length. The detection probability is derived. A case study is demonstrated with a real-traffic series to suggest that the present method can be used to give an alarm at the early stage of intrusions.

***Keywords*—** Network-based IDS, anomaly intrusion detection, information integrity, pattern matching.

I. INTRODUCTION

DETECTING intrusions with unknown patterns is greatly desired in practice. In this regard, the challenge is to develop a system that detects close to 100 percent of attacks (i.e., reliable detection) as can be seen from [1]. This requires that detection should be reported with detection probability [3]. Since a basic characteristic of intrusion detection system (IDS) is that an IDS is objective-dependent [4], this paper presents a statistical method of detecting tampered type attacks.

Let $x(t)$ be a normal arrival traffic time series at a protected site in an e-business system, indicating the number of bytes in a packet at t. If an adversary tampers with $x(t)$ by purposely deleting or inserting some packets into $x(t)$, we say that the site suffers from tampered type intrusions or attacks. An obvious consequence of tampered type attacks is that the information integrity of normal communication data series is destroyed. It is noted that the term tampered type attacks may be informal but we use it in this presentation to mean as mentioned above. If the site suffers from tampered type attacks, the arrival traffic series becomes abnormal and we denote it as $\hat{x}(t)$. The importance of detecting tampered type attacks in an e-business system is obvious.

From a view of reliable detection, there are two key specifications for an intrusion detection system (IDS). One is that an IDS should tell us when the monitored site is attacked (alarm time) and the other is what the detection probability is. A real-time reporting of intrusions is required to be on the order of minutes or hours as implied in [5].

Traffic time series is random [6,8]. Usually, an intruder may not delete or insert only one packet into a dada series but some periodically or randomly from time to time. Therefore, both normal and abnormal traffic series are random. Though traffic has the property of long-range dependence (LRD), its power spectrum exists if it is of finite length. Due to the efficiency of fast Fourier transform (FFT) in practice and a series practically encountered in engineering being finite length, this paper presents a method for reliable detection of tampered type attacks based on power spectrum.

In what follows, section 2 describes the detection method. Section 3 demonstrates a case study. Discussions are given in Section 4 and conclusions in Section 5.

II. DESCRIPTION OF DETECTION METHOD

Taking the power spectrum $S_x(f)$ of finite length $x(t)$ as a template, $\hat{x}(t)$ can be detected with the correlator

$$corr[S_x(f), S_{\hat{x}}(f)],$$

where $S_{\hat{x}}(f)$ is the power spectrum of $\hat{x}(t)$. From a view of statistical pattern matching, power spectrum of a normal series can be taken as a feature of that series [7]. During detection, the monitored traffic series is sectioned. Let j be the sections index. Then,

$$corr[S_x(f), S_{\hat{x}}(f)]$$

is the function of j, which is denoted as

$$c(j) = corr[S_x(f), S_{\hat{x}}(f)].$$

Denote h as the threshold of $c(j)$. Then, the detection hypothesis is

$$|c(j)| \le h,$$

where

$$0 \le |c(j)| \le 1.$$

Manuscript received March 2, 2007; Revised version received June 19, 2007. This work was supported in part by the National Natural Science Foundation of China under the project grant number 60573125. Wei Zhao's work was also partially supported by the NSF (USA) under Contracts 0808419, 0324988, 0721571, and 0329181. Any opinions, findings, conclusions, and/or recommendations in this paper, either expressed or implied, are those of the authors and do not necessarily reflect the views of the agencies listed above.

[1] Ming Li (corresponding author) is with the School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, PR. China. (Tel.: +86-21-5434 5193; fax: +86-21-5434 5119; e-mails: ming_lihk@yahoo.com, mli@ee.ecnu.edu.cn).

[2] Wei Zhao is with Rensselaer Polytechnic Institute, 110 Eighth Street, 1C05 Science, Troy, NY 12180-3590, USA. (e-mail: zhaow3@rpi.edu).

When

$$|c(j_0)| \le h,$$

we say $S_{\hat{x}}(f)$ is uncorrelated with $S_x(f)$ in the sense of the threshold being h and the site of $x(t)$ is attacked at j_0. Define detection probability as

$$\text{Prob}(h) = m/M, \quad (1)$$

where M stands for the total number of outcomes that the site of $x(t)$ is attacked and m for the number of outcomes that the attack behavior is detected. Extremely,

$$\text{Prob}(0) = 1 \quad (2)$$

and

$$\text{Prob}(1) = 0. \quad (3)$$

Let $\mathcal{A}$ be alarm time. Then, $\mathcal{A}$ consists of two parts:

$$\mathcal{A} = \mathcal{A}_l + \mathcal{A}_p, \quad (4)$$

where $\mathcal{A}_l$ is the time when an intrusion is detected and $\mathcal{A}_p$ indicates the time for data processing.

III. A Case Study

We use a real-traffic trace named Apr0502.sytu, which was measured on the Ethernet at Southern Yangtze University, China, in 5 April 2002. The settings for signal processing are:

- The section index is j (= 1, 2, …);
- each section contains 12 non-overlapping segments for averaging, denoting L = 12;
- each segment contains 128 points of data, N = 128;
- the average time resolution Δt is 3.14 ms;
- the section size is $12\times128\times\Delta t$.

The normal trace of Apr0502.sytu is indicated in Fig. 1. The abnormal trace is arranged in this way. For every 20 points of original series, there is one point whose packet size is randomly set, as shown in Fig. 2. Fig. 3 shows the correlation curve. Fig. 4 indicates the detection probability. One is able to determine what detection probability is with the function of Prob(h). For instance, if

$$h = 0.9,$$

the "attack" can be detected at

$$j \approx 11,$$

see Fig. 3. That is, the "attack" can be detected at 16,896-th (=11×12×128) point of that series. Hence,

$$\mathcal{A}_l = 11\times12\times128\times\Delta t = 53 \text{ seconds}.$$

With an FFT-based correlator, 4356 times of multiplication and 12960 times of addition are needed for processing when the "attack" is detected, referring [9] for the complexity computations regarding FFT. As $\mathcal{A}_p$ < 1 second when the computer (Pentium 450) is used, one has

$$\mathcal{A} \approx \mathcal{A}_l = 53 \text{ seconds}.$$

The detection probability is

$$\text{Prob}(0.9) \approx 0.94,$$

see Fig. 4.

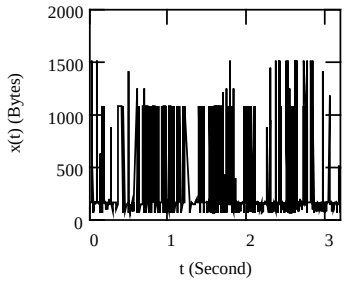

Fig. 1. Normal series of Apr0502.sytu.

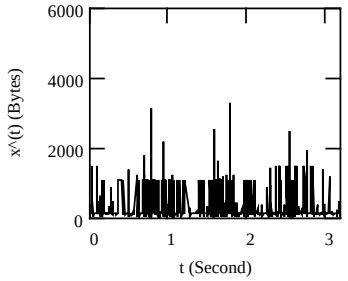

Fig. 2. Abnormal series of Apr0502.sytu.

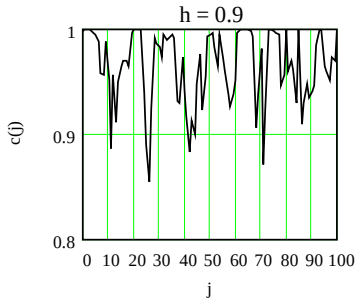

Fig. 3. Correlation curve between $S_x(f)$ and $S_{\hat{x}}(f)$.

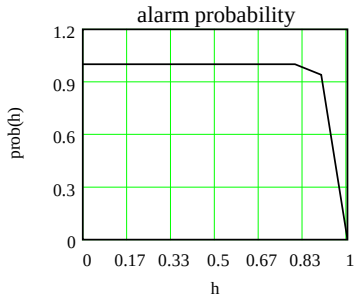

Fig. 4. Detection probability curve.

IV. Discussions

The previous explanations stand for a profile regarding the statistical detection of tampered type attacks. A detection does not give an alert which packet is deleted or inserted but statistically implies an alert that the monitored site is suffering tampered type attacks. The classical probability used above mainly shows the detection method in principle for the simplicity. Using other probabilistic approaches may yield other expressions of detection probability. Our further task is to seek a method how to maximize the detection probability for a given probability of false alarm.

V. CONCLUSION

A statistical method to detect tamped type attacks in e-business systems has been proposed and discussed. A case study has been given to show that the method meets the requirements of real-time reporting for detection.

REFERENCES

[1] R. A. Kemmerer and G. Vigna, "Intrusion detection: a brief history and overview," *Supplement to IEEE Computer (IEEE Security & Privacy)*, 35 (4), April 2002, 27-30.

[2] J. Leach, "TBSE— an engineering approach to the design of accurate and reliable security systems," *Computers & Security*, 23 (1), Feb. 2004, 22-28.

[3] M. Li, "An approach to reliably identifying signs of DDOS flood attacks based on lrd traffic pattern recognition," *Computers & Security*, 23(7), 2004, 549-558.

[4] E. Schultz, "Intrusion prevention," *Computers & Security*, 23 (4), Oct. 2004, 265-266.

[5] E. Amoroso and R. Kwapnieski, "A selection criteria for intrusion detection systems," *Proc. of 14th Annual Conf. on Computer Security Applications*, Dec., 1998, 280-287.

[6] M. Li and C.-H. Chi, "A correlation based computational model for synthesizing long-range dependent data," *Journal of the Franklin Institute*, 340 (6/7), Sep.-Nov. 2003, 503-514.

[7] K. S. Fu, editor, *Digital Pattern Recognition*, Springer-Verlag, 2nd edition, 1980.

[8] M. Li, "Change trend of averaged Hurst parameter of traffic under DDOS flood attacks," *Computers & Security*, 25 (3), May 2006, 213-220.

[9] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing*, John Wiley & Sons, 1993.

Ming Li was born in 1955 in Wuxi, China. He completed his undergraduate program in electronics engineering at Tsinghua University. He received the M.S. degree in ship structural mechanics from China Ship Scientific Research Center (CSSRC) and Ph.D. degree in computer science from City University of Hong Kong, respectively. From 1990 to 1995, he was a researcher in CSSRC. From 1995-1999, he was with the Automation Department, Wuxi University of Light Industry. From 2002 to 2004, he was with the School of Computing, National University of Singapore.

In 2004, he joined East China Normal University (ECNU) as a professor in both electronics engineering and computer science. He is currently a Division Head for Communications & Information Systems at ECNU. His research areas relate to applied statistics, computer science, measurement & control. He has published over 70 papers in international journals and international conferences in those areas. Li is a member of IEEE.

Wei Zhao is a professor of computer science and the dean for the School of Science at Rensselaer Polytechnic Institute. His research interests include distributed computing, real-time systems, computer networks, and cyberspace security. Zhao received a PhD in computer and information sciences from the University of Massachusetts, Amherst. He is a Fellow of the IEEE. He has published over 280 papers in international journals, international conferences, and book chapters.