Network Traffic Anomaly Detection based on Ratio and Volume Analysis

Hyun Joo Kim, Jung C. Na, Jong S. Jang

Active Security Technology Research Team
Network Security Department Information Security Division,
Electronics and Telecommunications Research Institute (ETRI)
161 Gajeong-Dong, Yuseong-Gu, Daejeon, 305-350, Korea

Summary
Recent attack targets on a public network as well as an enterprise/edge network or system because the damage of the public network-attack is far stronger than an enterprise network or systems-attack and the speed of its propagation is far faster. Theses attacks typically cause not only traffic congestion but also network failure exhausting network bandwidth, router processing capacity using the abnormal traffic or excessive network traffic, so that they can have an extremely large impact on the public network. Therefore in this paper, we propose the detection mechanism of network traffic anomalies. This mechanism analyzes flow data based on the statistical anomaly detection. Besides, it supports the two analysis method- ratio based analysis and volume based analysis, and it correlates the results from these two models or the results of analysis according to each traffic characteristic parameter to solve the problem of each model and reduce the false-positive.

Key words:
anomaly detection, ratio analysis, volume analysis, population proportion

1. Introduction

Today's attack usually targets on a public network as well as an enterprise/edge network or system because the damage of the public network attack is far stronger than an enterprise network or systems attack and the speed of its propagation is far faster. It can cause not only traffic congestion but also network failure using the abnormal traffic or excessive network traffic.
Therefore this paper is focused on the traffic anomalies such as failures and attacks especially on the excessive abnormal network traffic. Identifying, diagnosing and treating anomalies in a timely fashion are the fundamental part of day to day network operations. Without this kind of capability, networks are not able to operate efficiently or reliably [1].
To achieve it, we developed the Security Management System (SMS) which can provide detecting, diagnosing, and responding to anomalies in real-time. SMS diagnoses the network state analyzing security alerts from network security equipments such as IDS and firewall and flow data from the measuring equip. such as Netflow[2] of Cisco, and it takes the response actions to mitigate and treat the security threat.
However in this paper, we introduce the only detection mechanism of network traffic anomalies.

2. Related Works

General anomaly detection techniques in networks have been widely treated due to their importance in network management [10]. Katzela and Schwartz which focuses on methods for isolating failure in networks [11], Feather et.al which shows the faults can be detected by statistical deviations from regularly observed behavior [12], Brutlag which applies thresholds to time series models to detect aberrant network behavior [13]. But anomaly detection models must be trained on the specific network to be monitored. It is naive to assume that a network with a connection to the Internet is clean when the anomaly detector is being trained. So our approach provides the init-threshold configured by manager, which is used to prevent the anomalous traffic from modeling the normal traffic when the anomaly detector is being trained.
Recently statistical analysis of aggregate traffic data has been studied [1, 8, 9]. The works in [1, 8] have studied traffic volume as a signal for wavelet analysis. The work in [9] proposed a technique for traffic anomaly detection based on analyzing correlation of destination IP addresses in outgoing traffic at an egress router. This correlation data are transformed through discrete wavelet transform for the detection of anomalies through statistical analysis. And most traffic monitoring and analysis products have used the traffic volume for the anomaly detection. But using the only traffic volume for the anomaly detection is limited to explain traffic increase because of network extension. Our approach proposes the statistical anomaly detection mechanism based on the absolute traffic volume and relative traffic ratio.

Manuscript received May 5, 2006.
Manuscript revised May 25, 2006.

3. Detection of Excessive Network Traffic

Our goal is to prevent the network traffic congestion and network failure to let the network support many kinds of services without interrupt and limit. So our analysis mechanism focuses on the public network (ISP) not an enterprise network (private network). Surely, security of the private network is very important, but if the public network is not operated well, the private network can not be done as well.
Therefore we use flow data from some routers of the public network and the border routers connected to the private network. We collect flow data from some NetFlows of Cisco routers.
And we analyze flow data based on not signature based detection but anomaly detection. Besides, we support the two analysis method- ratio based analysis and volume based analysis and correlate the results from these two models to solve the problems of each model and reduce the false-positive error. Volume based analysis has false-positive when total traffic volume including the specific traffic which will be analyzed increased because the number of legitimate user increased accidentally in that time. On the other hand, ratio based analysis has false-positive when total traffic volume decreased because other kind of traffic (background traffic) decreased. Hence it is necessary to correlate them.

3.1 Detection Environment (Network Architecture)

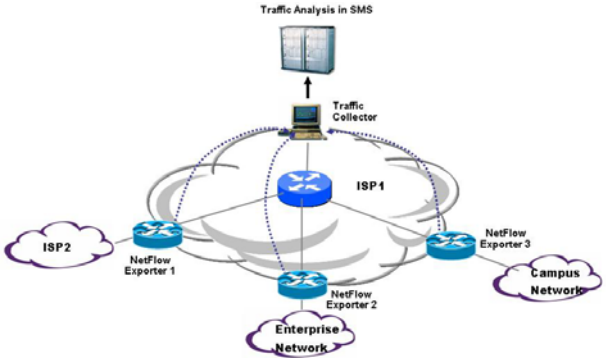

Fig. 1 Detection Environment

Netflow-exporter, which is embedded in a Cisco router, is responsible for measuring the traffic passed through the Cisco router and exporting it to the traffic-collector configured in Netflow[6].
Traffic-collector periodically collects traffic data exported from the NetFlow-exporter and it combines the collected traffic data and transfers them to the traffic analysis system. At this time, traffic-collector converts the collected traffic data into a designated format for transmission. The traffic information transmitted is following- the source and destination IP addresses, destination port number, destination protocol number, packet count, byte count, and start/end time.
Traffic analysis system periodically pre-processes and analyzes traffic data collected from each traffic collector according to characteristic parameter for analysis.

3.2 Traffic Analysis Model

Traffic analysis model must have the low computing complexity to analyze traffic in near real time considering the speed of attack propagation. So our detection system uses exponential smoothing model [3] for volume based analysis and population proportion testing [4] model for ratio based analysis.
Exponential smoothing model (formula (1)~(3)) is a popular scheme to produce a smoothed Time Series [5, 7], and calculates the expected value of next term adjusting weight on an average of past and current value of this term. To want to know the traffic analysis mechanism of our SMS applied this model, refer to the paper [3]. Y_{t+1} is traffic forecast value at time t+1 and X *is* current measured traffic value at time t. α *and* γ are smoothing constants and *MAD* is mean absolute deviation.

$$Y_{t+1} = \alpha X_t + (1-\alpha) Y_t \quad \textbf{(1)}$$

$$0 < \alpha \leq 1 \text{ , where } Y_1 = X_1$$

$$MAD_t = \sum^{-1} (|Et| / n) = \gamma |Et| + (1-\gamma) MAD_t \quad \textbf{(2)}$$

$$\text{where } MAD_0 = \text{initail set value, } E_t = X_t - Y_t$$

$$(Y_{t+1} - n\,MAD_t) \leq X_{t+1} \leq (Y_{t+1} - n\,MAD_t) \quad \textbf{(3)}$$

Hence in this paper, we introduce simply the population proportion testing used as the ratio based analysis model.
Formula (4)~(7) shows some formulas of proportion testing. In formula (4), $T_i(t+1)$ means the rate of specific traffic at time t+1. $T_v(t, t+1)$ is the volume of specific traffic and $T_{total}(t,t+1)$ is the total traffic volume at time t+1. Formula (5) and (6) are made heuristically through the expectation method used in exponential smoothing for it is very difficult to save a number of data which have been measured since the analysis was started. μ and σ is mean and deviation of standard distribution. β, exponential constant, generally weights the measured values with the range of 0.1~0.3 in exponential smoothing, but in this model, it is appropriate that β is 0.02 because the larger β is, the more fluctuant the threshold(the upper confidence interval) is. *N* is sample size (number of analysis period)

and Z_α is the level of significance. The range of ratio merely is limited with 0~1.
With the values calculated from formula (5) and (6), we finally calculate the ratio based threshold in the formula (7). This threshold is used as a means of detecting the traffic anomalies and updated continually by μ and σ updated with the normal state values.

$$T_i(t+1) = T_v(t, t+1) / T_{total}(t, t+1) \quad (4)$$

$$\mu = N\mu + T_i(t+1) / N+1 \quad (5)$$

$$\sigma^2 = \beta\sigma^2 + (1-\beta)\{T_i(t) - \mu\}^2, \quad \sigma = \sqrt{\sigma 2} \quad (6)$$

$$Threshold = \mu + z_\alpha \sigma \quad (7)$$

3.3 Architecture of Traffic Analysis System

In this section, we explain the architecture of traffic analysis system. Figure 2 illustrates a block diagram of the traffic analysis system detecting the network traffic abnormality.
This analysis system is consisted of some components – TrafficPre-Processor, Analyzer, Profiler, Correlator, Analysis Model, and Database.

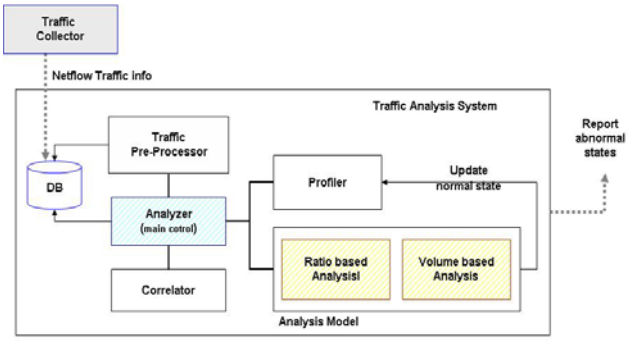

Fig. 2 Block diagram of the traffic analysis system.

- TrafficPre-Processor: It pre-processes the traffics received by traffic-collectors as information required by Analyzer.
- Analyzer: It receives the pre-processed data from the TrafficPre-Processor and calculates the threshold with the mean and deviation of Profiler. Then it compares the present traffic with the threshold so that it decides the security state of network.
- Profiler: It performs a normal traffic modeling during the traffic learning(training) period and updates the normal state model with the normal traffic data during the traffic analysis period.
- Correlator: It correlates the results from the two analysis models or the results of each analysis(characteristic) parameter.
- Analysis Model: It is composed of population proportion testing model and exponential smoothing model for ratio-based and volume-based analysis method.
- Database: It has the network traffic data(Netflow data) and analysis options which affects the analysis model and profiler. Also it stores analysis results about normality and abnormality generated in each analysis period.

3.4 Detection Algorithm for Network Traffic

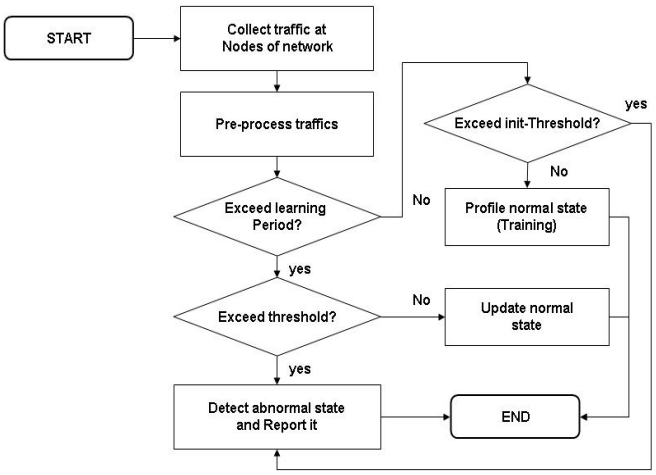

Fig.3 Flow chart of detection algorithm

This flow chart describes a method of detecting the network traffic anomalies and includes the operations of the traffic-collectors as well as the traffic analysis system.
1) Firstly traffic collectors can collect the network traffic data passed through the network equipments using Netflow Exporter and stores them in database for traffic analysis system.
2) And TrafficPre-processor of traffic analysis system processes the traffics data to make format required by Analyzer.
3) Then if the analysis count does not exceed the traffic learning count, it is a traffic learning period. Otherwise it is a traffic analysis period.
4) During the traffic learning period(initial training period), Profiler performs a normal traffic modeling (normal state model) using some values used in each analysis model. Also Profiler updates the normal state values with the normal traffic data which was determined as the normal state by Analyzer during a traffic analysis period. In the learning period, the traffic volume or ratio which exceeds

the init-threshold is removed and reported for the accuracy of normal traffic model.
5) According to the analysis model, Analyzer calculates the threshold - ratio and volume based - with mean and deviation generated from the Profiler. And it compares the threshold with the observed traffic in this analysis period. At this time, if the present observed traffic value exceeds the threshold, it decides that the traffic is abnormal. After this, for the accuracy of detection, Analyzer requests Correlator to correlate the results of each analysis model. At this time, to correlate results analyzed by each model, we propose the composite severity level and reliability level. Composite severity level is created by severity levels resulted from each model. Reliability level means how trustworthy the composite severity level is. The closer to zero difference of each severity level is, the larger reliability level is.
6) Last if analyzed state is determined to be abnormal, Analyzer must notify to a manager the information of abnormal state and correlation result.

4. Conclusion and Future Works

In this paper, we present the anomaly detection mechanism for detecting excessive network traffic based on correlation model using both ratio-based analysis model and volume-based analysis model. The result of our anomaly detection can be used to manage a network in combination with the security response policy, thereby we can provide an automatic detection and response. Also because our mechanism integrates and analyzes the traffics of not the private network but all managed networks, it can detect more quickly abnormal situations such as network performance degradation, traffic congestion, etc., in the initial step of network attack.

In the future, we intend to investigate how well this mechanism can be implemented to detect the network traffic anomalies using real network traffic not experiments traffic data in real world and real-time. And we plan to improve the performance to search and insert flow data in Database quickly through the database aging-out and multiple table space and to verify formulas of the severity level and reliability level.

References

[1] P. Barford, J. Kline, D. Plonka and A. Ron, “A Signal Analysis of Network Traffic Anomalies”, in Proceedings of ACM SIGCOMM Internet Measurement Workshop, Marseille, France, November 2002.
[2] “Cisco IOS NetFlow Overview”, http://www.cisco.com/univercd/cc/td/doc/product/software/ios124/124cg/hnf_c/nfb_ov.pdf
[3] S.H. Lee, H.J. Kim, J.C.Na et al., “Abnormal Traffic Detection and Its Implementation”, ICACT2005 , Feb. 21~23, 2005
[4] “Estimating the Population Proportion“, http://www.richland.edu/james/lecture/m113/
[5] G. Box, G. Jenkins, G. Reinsel, “Time Series Analysis”, 3rd edition, Prentice Hall, 1994
[6] “Configuring the Exporting of Statistics from the NetFlow Main Cache”, http://www.cisco.com/univercd/cc/td/doc/product/software/ios124/124cg/hnf_c/ch05/nfb_bexp.pdf
[7] P. Brockwell and R. Davis, “Introduction to Time Series and Forecasting”, Springer, 1996
[8] Anu Ramanathan, “WADeS: A Tool for Distributed Denial of Service Attack Detection”, TAMU-ECE-2002-02, Master of Science Thesis, August 2002. http://dropzone.tamu.edu/techpubs/2002/thesis_ramanathan.pdf
[9] S.S Kim, A.L Narasimha Reddy, M. Vannucci, “Detecting Traffic Anomalies at the Source through aggregate analysis of packet header data”, TAMU-ECE-2003-03, 2003
[10] P.Barford, D. Plonka, “Characteristics of network traffic fbw anomalies;' in Proceedings of ACM SIGCOMM Internet Measurement Workshop, Francisco, CA, November 2001.
[11] I. Katzela and M. Schwartz, “Schemes for fault identification in communications networks,” IEEE/ACM Transactions on Networking, vol. 3(6), pp. 753-764, December 1995.
[12] F. Feather, D. Siewiorek, and R. Maxion, “Fault detection in an ethernet network using anomaly signature matching,” in Proceedings of ACM SIGCOMM ‘93, San Francisco, CA, September 2000.
[13] J. Brutlag, “Aberrant behavior detection in time series for network monitoring,” in Proceedings of the USENIX Fourteenth System Administration Conference LISA XIV, New Orleans, LA, December 2000.

Hyun Joo Kim received the B.S. and M.S. degrees in Information Engineering from Sungkyunkwan Univ. in 2000 and 2002, respectively. Since 2002, she has worked in Information Security Research Division of Electronics and Telecommunications Research Institute (ETRI) of Korea to study network security, active network, and network management. She is interested in and researches anomaly detection of network traffic and security incident sharing.

Jung C. Na received his M.S. and and Ph. D degree in computer science in 1989 and 2005 from Soongsil and Chungnam University, Korea. He has been a Principal Engineering Staff in ETRI since 1989. His research interests include network security, active networks, and real time system.