Traffic-Based Credit Network Management System

Yizhe Zhang
Computer Science & Technology
Tsinghua University
Beijing, China
zhangyizhe4503@yahoo.com.cn

Jilong Wang
Tsinghua National Lab. for IS &
Tech Network Research Center
Tsinghua University
Beijing, China
wjl@cernet.edu.cn

Haizhuo Lin
Computer Science & Technology
Tsinghua University
Beijing, China
lin.haizhuo@163.com

***Abstract*—With the rapid development of Internet, some bad behaviors such as network attack and bandwidth abuse become more and more serious, result in wasting the limited resources of network. Cernet2, the experimental IPV6 network, is suffering from the shortage of monitoring resources, so how to maximize the utilization of the limited resources is an urgent problem. In this paper, we propose the traffic-based network credit management system. According to the credit status of each node, we can apply the appropriate monitoring force. We firstly capture the traffic of Cernet2 by NetFlow, and use statistical methods to identify network bad behaviors. Then assign each node a credit score. Finally, display the credit status and corresponding statistical analysis of each node on the credit management website in real time.**

Keywords-component; network credit management; Cernet2; credit score; statistics;

I. INTRODUCTION

Nowadays, Internet plays a more and more important role in our life. According to statistics of the International Telecommunication Union in early 2012, the number of global Internet users has already exceeded 2 billion, with a growth rate more than 10%. However, along with the increasing size of Internet, the job of managing the network becomes one of the most challenging tasks. According to statistics, the average annual pecuniary loss due to Internet attacks such as Dos, DDos, Trojan horses, viruses and spam exceeds 100 billion dollars. A safe and stable network is an urgent need.

At the same time, IPV4 addresses are close to depletion and we are in a transition period to IPV6. With an important role in the Internet research, China is making big effort to develop the next generation Internet. To this end, China established Cernet2, the second generation of China Education and Research computer network, and actively research on new technologies and applications.

As an emerging experimental network, Cernet2 is also facing the challenging security situation. Compared to the developed IPV4 network, human capital, infrastructure and financial resources are more limited, so it is difficult to apply adequate monitoring force on each node. Therefore, how to maximize the utilization of the limited resources is an urgent need. If we can pre-determine which IP or IP segment is more likely to make bad behaviors, we can effectively strengthen the monitoring force of this area. To the contrary, for the segment of network with better performance, we can reduce monitoring force to avoid of waste, so as to maximize the utilization of resources.

For this purpose, we can draw lessons from the credit system commonly used in financial industry. Based on the historical traffic records, we give each node a real-time credit evaluation by a credit score. The score will be deducted when the node makes bad behaviors. Conversely, the score will be increased after a period of time without bad behaviors. According to the credit status, we use different standards of supervision or service provision. For nodes with low credit score, we can correspondingly strengthen monitoring force or provide services with restrictions.

The paper is organized as below. Section 2 briefs the whole procedure of the credit management system followed by the introduction of data acquisition and processing in section 3. Section 4 discusses the identification of abnormal behaviors. Section 5 describes the rule of the credit management system in detail. Section 6 shows the display platform and the analysis of the results, and then concludes the paper and look forward to the future research in section 7.

II. AN OVERVIEW OF THE SYSTEM

The proposed credit management system is shown in Fig.1 including 4 steps.

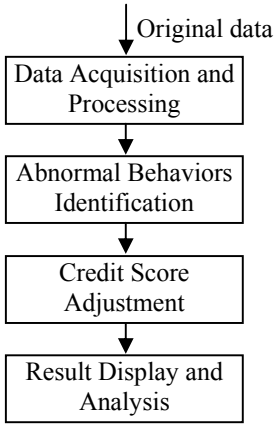

Figure 1. Overview of the system.

Published by Atlantis Press, Paris, France.

III. DATA ACQUISITION AND PROCESSING

We capture the sample traffic from the Beijing export of Cernet2 every fifteen minutes using NetFlow, and then output the traffic in a report. The sampling fraction is 1/10. The traffic records in the report contain source addresses, source ports, destination addresses, destination ports, protocol types, packet sizes and other information.

To reduce the burden of the real-time statistical system and improve the operation efficiency, we only output the top 20 records ranked by the usage of network resources or the number of connection attempts, making the system three times faster as a result. In fact, most of the bad behaviors require large amount of network resources so out treatment will seldom miss them.

One of the reports is shown in Fig.2, ordered by the usage of bandwidth.

Sort Order : Destination Bandwidth

Rank	IP	Bandwidth(Mbps)		Packets/sec		Avgpkt(Bytes)
1	210.45.224.186	112.07	(0.93%)	20357	(0.90%)	688
2	118.228.24.237	106.07	(0.88%)	12322	(0.54%)	1076
3	58.205.218.18	96.34	(0.80%)	8826	(0.39%)	1364
4	125.222.31.159	95.53	(0.79%)	8000	(0.35%)	1492
5	125.222.31.156	92.36	(0.77%)	7714	(0.34%)	1496
6	125.222.31.141	75.24	(0.63%)	6321	(0.28%)	1487
7	58.205.218.9	59.08	(0.49%)	5167	(0.23%)	1429
8	118.228.24.218	56.90	(0.47%)	8246	(0.36%)	862
9	210.45.224.114	55.97	(0.47%)	8492	(0.37%)	823
10	210.45.224.116	54.11	(0.45%)	8499	(0.37%)	795
11	210.45.224.125	53.94	(0.45%)	8246	(0.36%)	817
12	210.45.224.126	53.04	(0.44%)	8130	(0.36%)	815
13	210.45.224.124	51.80	(0.43%)	8110	(0.36%)	798
14	210.45.224.121	51.20	(0.43%)	7325	(0.32%)	873
15	210.45.224.122	50.36	(0.42%)	8471	(0.37%)	743
16	210.45.224.113	50.28	(0.42%)	8185	(0.36%)	767
17	210.45.224.115	48.88	(0.41%)	7686	(0.34%)	794
18	210.45.224.120	48.34	(0.40%)	7550	(0.33%)	800
19	210.45.224.118	47.09	(0.39%)	7427	(0.33%)	792
20	210.45.224.117	46.71	(0.39%)	7270	(0.32%)	803
TOTAL= 43965	*.*.*.*	12019.269	(100%)	2274044	(100%)	660

Figure 2. NetFlow report

IV. BAD BEHAVIORS IDENTIFICATION

We need to establish a standard to identify abnormal behaviors by the traffic report. There are two kinds of standard, absolute standard and relative standard. Considering of the unique volatility of the network, we choose the latter which is more flexible.

According to the existing theories, if the data shows a normal distribution, there will be more than 99% of the data falls into the interval [μ- 3σ, μ+ 3σ], where μ is the mean andσ is the standard deviation of the dataset. According to Chebyshev's conclusion, regardless of the distribution of the data, at least 1-1/k^2 of the data falls into the interval [μ- kσ,μ+ kσ]. For k=3, there will be at least 89% of the data falls into [μ- 3σ,μ+ 3σ].

As the distribution of the top 20 records ranges between the normal distribution and the extreme distribution, we therefore examine a record by whether it exceeds the mean of other data plus three times of the standard deviation. The significant level is around 0.05, which is quite accurate.

So we establish the criteria as follows. For the first record, we use the other 19 records as the sample dataset and calculate the mean and the standard deviation. If the first record exceeds the mean plus three times of the standard deviation, we will identify it as abnormal. If the first record is abnormal, we will perform the same procedure to identify the second record using the other 18 records as the sample dataset. If the second record is identified as abnormal as well, we will go for the third. Whatever the result of the third is, we will not check others.

V. CREDIT SCORE ADJUSTMENT

A. 15Minutes Abnormal Point

Reports are generated every 15 minutes by NetFlow. As previously described, we output the top 20 records ranked by the usage of network resources or the number of connection attempts. The abnormal test is conducted automatically by a java program in background. It uses a regular expression to match the data from the page with the records, then do the test and finally give the abnormal record a point. The abnormal point is used for the adjustment of credit score at the end of the day.

When a record is identified as abnormal, the source address should be assigned a greater value as the initiator of bad behaviors. For the passive side, the destination address will be given a smaller value. Meanwhile, the points should be different based on the rank of network resources usage or connection times. The higher the rank is, the greater value the point assigned. Considering the actual network environment of Cernet2, we set the rules as follows.

TABLE I. ABNORMAL POINTS

Address\Rank	1st	2nd	3rd
Source	6	4	2
Destination	3	2	1

In particular, if an IP is both the source and the destination in the abnormal records at the same time, its abnormal point is the sum of them. For example, if an IP is the 2nd as the source and the 1st as the destination, its mark value is 4 + 3 = 7.

Universities are the nodes of our credit management system. Every university has its fixed IP segment. The abnormal point of a university is the sum of the IPs'.

B. Daily Credit Score Adjustment

Because of the great volatility of network traffic, it is inappropriate to identify a university merely by 15 minutes traffic. We evaluate the universities and adjust their credit scores daily. NetFlow outputs a report every 15 minutes, so there are 96 reports every day. We add up the abnormal points and set a limit. A sum above the limit means a university with bad behaviors. Therefore, only the universities having persistent negative impact on the network will be punished. This method greatly reduces the probability of identification error due to the fluctuation of network activities.

The limit is set based on the actual network environment, the higher the limit, the larger probability of missing bad behaviors but the smaller probability of erroneous judgment, and vice versa. The goal of the adjustment is to make it distinguished between universities. We attempt to avoid the situations such as all the universities have too many or zero

bad behaviors. For the real environment of Cernet2, we set the standard as follows: In a day, if the accumulated abnormal point of a university exceeds 50, 0.2 points will be subtracted from its credit score. If it exceeds 100, 0.5 points will be subtracted. 1 point will be subtracted if it exceeds 200.

C. *Weekly Credit Score Adjustment*

For the universities without bad behaviors for a period, we increase its credit score. The length of the period and the extent of increase can also be adjusted in accordance with the actual network environment. This paper uses the following criteria: If a university doesn't have any bad behaviors for a week, its credit score will be increased by 1 point. In addition, the initial credit score is 80. The highest is 100 and the lowest is 0.

VI. RESULT DISPLAY AND ANALYSIS

A. *Result Display*

To display the real-time credit status of universities and the corresponding detail information intuitively, we make a website for the credit management system. There is a java program for collecting, processing, statistics and analysis. Then it outputs the result into an xml file in background. The xml file is used by a flash to show the result as well as the details on the website in real time. The website is shown below in Fig.3.

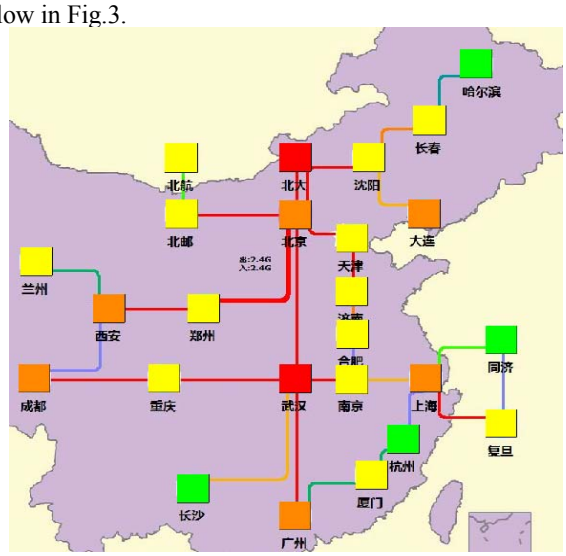

Figure 3. Website of the credit management system

TABLE II. COLOR, CREDIT SCORE AND CREDIT STATUS

Color	Credit Score	Credit Status
Green	85~100	Excellent
Yellow	70~80	Good
Orange	60~70	Mediocre
Red	0~60	Poor

The relationship among color, credit score and credit status is shown in Table Ⅱ.

Users can get detailed information by clicking the nodes. There are many statistical results on the subpages, such as the history of credit score, the history of daily abnormal points, and the distribution of abnormal points. Fig.4-Fig.6 are shown below:

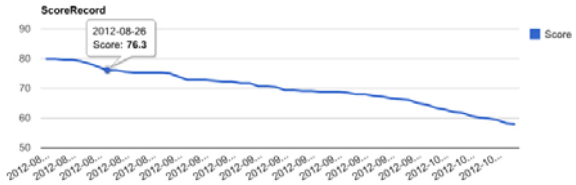

Figure 4. History of credit score

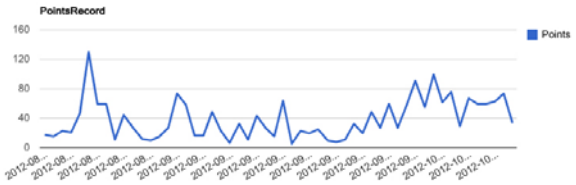

Figure 5. History of abnormal points

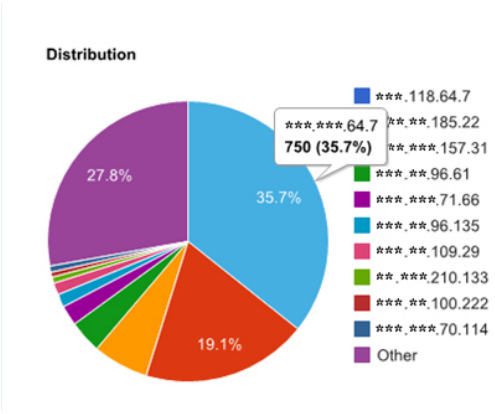

Figure 6. Distribution of abnormal points

B. *Result Analysis*

From the point of view of universities, the credit status differentiated significantly. As Fig.3 shows, there were four green nodes which possess excellent credit status but also two nodes in poor credit status marked by red color. From the initial 80, the highest score was above 90 by Tongji University while the lowest was below 40. Besides, the bad behaviors displayed notable continuity. As shown in Fig.4, the university demonstrated bad behaviors continually so its credit score declined persistently. The Hurst exponent is used as a measure of long term memory of time series. It quantifies the relative tendency of a time series either to regress strongly to the mean or to cluster in a direction. A value H in the range $0.5 < H < 1$ indicates a time series with long-term positive autocorrelation, meaning both that a high value in the series will probably be followed by another

high value and that the values a long time into the future will also tend to be high. The Hurst exponent is 0.68 in this paper indicating that the universities with bad behaviors are much more likely to act again soon.

Bad behaviors are judged by the usage of Internet. We investigate whether there is a relationship between the credit score and the traffic load. But in fact, Beijing University of Posts and Telecommunication, which has the most developed network of IPV6 in China, has the largest traffic but with a credit score more than the initial score. By t-test, we conclude that there is no exact connection between credit score and traffic load.

From the perspective of IPs, the distribution of abnormal points illustrates a high degree of concentration. As shown in Fig.6, more than 80% of the abnormal points occur due to 3 IPs. We need more investigations to determine whether the 3 IPs are doing network tests or something requiring large amount of network resources. If not, universities can put some restriction on them. In most universities, abnormal points are concentrated, which validates the credit management system. We can apply more monitoring force on IPs with poor credit status so that resources can be allocated in an optimal manner.

VII. FUTURE RESEARCH

In this paper, we only used NetFlow to get some unsophisticated information of the traffic. In the future research, we can capture more detailed information of the traffic so that the judgment of bad behaviors will be more accurate.

It might be problematic using the bad behaviors judgment with traffic information only. We need more information to conclude bad behaviors, or to establish a complaint mechanism. If the explanations of the bad behaviors are reasonable, we can resume the reduced credit scores.

Nowadays, the credit system in network management is far from mature. We can learn more from financial industry, for example, the credit evaluation is not only determined by the history but also by the prospect, and how to use the result to guide our future management. This paper hopefully can bring more people's attention to this field, continuously improving the traffic-based credit network management system.

ACKNOWLEDGMENT

Thanks to Pengfei Li for the assistance in data collecting.

REFERENCES

[1] Fei Wang, Qin Yan and Jilong Wang, "Credit Based network Management by Discriminate Analysis", International Conference on Funture Internet Technologies, 2010.

[2] Jilong Wang, Dah Ming Chiu and John C.S. Lui, "Credit-based Network Management", Communication System and Network and Workshops, 2009. pp: 5-10

[3] Peng Chen, "The Analysis of Bad Online Communication Behavior and Credit Management", unpublished.

[4] D. West, "Neural network credit scoring models. Computers and Operations Research", 27: 1131-1152, 2000.

[5] L. C. Thomas, "A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. International Journal of Forecasting", (16): 149-172, 2000.

[6] Xiangyang Xu and Jike Ge, "Research on Personal Credit Scoring Model based on Clustering", Microcomputer Information, 2006, (27): 229-231.

[7] David Durand and Muhittin, "A credit scoring approach for the commercial banking sector", Socio-Economic Planning Sciences, 2001, 37(2):103-123

[8] P. Makowski, "Credit scoring branches out", *Credit World*, 74, 1985.