Towards Lightweight and Efficient DDoS Attacks Detection for Web Server

Yang Li[1], Tian-Bo Lu[2], Li Guo[3], Zhi-Hong Tian[3], Qin-Wu Nie[4]
[1]China Mobile Research Institute, Beijing China 100053
[2]National Computer network Emergency Response technical Team/Coordination Center of China 100029
[3]Institute of Computing Technology, Chinese Academy of Sciences, Beijing China 100190
[4]Hunan University of Science and Technology, Hunan China 411201
liyangyj@chinamobile.com

ABSTRACT
In this poster, based on our previous work in building a lightweight DDoS (Distributed Denial-of-Services) attacks detection mechanism for web server using TCM-KNN (Transductive Confidence Machines for *K*-Nearest Neighbors) and genetic algorithm based instance selection methods, we further propose a more efficient and effective instance selection method, named E-FCM (Extend Fuzzy C-Means). By using this method, we can obtain much cheaper training time for TCM-KNN while ensuring high detection performance. Therefore, the optimized mechanism is more suitable for lightweight DDoS attacks detection in real network environment.

Categories and Subject Descriptors
C.2.0 [**Computer-Communication Network**]: Security and Protection

General Terms
Security, Algorithms

Keywords
Web server anomaly detection, E-FCM algorithm, TCM-KNN algorithm, Instance selection

1. INTRODUCTION
Web server is a critical and necessary component for Internet applications and web applications dominate the most part of network traffic nowadays. However, they are suffering from a great deal of attacks especially Distributed Denial-of-Service (DDoS) attacks. DDoS significantly degrade service quality experienced by legitimate users. The key point for DDoS defenses is to detect it as soon as possible and neutralize this effect, thereby quickly and fully restore quality of various services to levels acceptable by the users. Currently researchers have designed and implemented numerous DDoS detection methods [1, 2].

However, all these methodologies measure DDoS damage superficially and partially by measuring a single traffic parameter, such as duration, loss or throughput, and showing divergence during the attack from the baseline case. They do not consider Quality-Of-Service (QoS) requirements of different applications and how they map into specific thresholds for various traffic parameters. They fail to measure the service quality experienced by the end users and thus not well suitable for DDoS detection for Web server. In recent years, Jelena, etc. proposed a novel measurement for DDoS towards the web applications from the perspective of end users in [3].

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

In our previous work, we proposed an effective anomaly detection method based on TCM-KNN (Transductive Confidence Machines for *K*-Nearest Neighbors) algorithm to fulfill DDoS attacks detection task towards ensuring the QoS of web server. The method is good at detecting network anomalies with high detection rate, high confidence and low false positives than traditional methods, because it combines "strangeness" with "p-values" measures to evaluate the network traffic compared to the conventional ad-hoc thresholds based detection and particular definition based detection. Secondly, we utilize the new objective measurement as the input feature spaces of TCM-KNN, to effectively detect DDoS attack against web server. Finally, we introduce Genetic Algorithm (GA) based instance selection method to boost the real-time detection performance of TCM-KNN and thus make it be an effective and lightweight mechanism for DDoS detection for web servers [4, 5]. However, we found the computational cost for GA is expensive, which results in high training time for TCM-KNN.

2. OUR METHODS
To further alleviate the expensive computational cost of GA and effectively reduce the training time for TCM-KNN, we developed E-FCM (Extend Fuzzy C-Means) algorithm to solve this problem. It is well known that standard FCM algorithm is usually used to fulfill clustering tasks in traditional applications. We think its core concept lies in the "membership grades" given to a data point for each group, and thus the concept could be effectively utilized in our instance selection task from unsupervised data for network anomaly detection. In other words, we will extend current FCM algorithm and use it to select the most valuable and representative data from the normal training data for efficient network anomaly detection. E-FCM is described as follows:

Step 1. Randomly initialize the membership matrix (*U*) that has constraints in Equation (1).

Step 2. Calculate centroids(ci) by using Equation (3).

Step 3. Compute dissimilarity between centroids and data points using Equation (2). Stop if its improvement over previous iteration is below a threshold.

Step 4. Compute a new *U* using Equation (4). Go to Step 2.

Step 5. Choose the data points the difference of whose membership grades for any two different clusters (named *diff*) is less than 0.5, and then add them to subset *STR* and delete them from the original dataset *TR*.

Step 6. Rank the matrix(*U*) for each formed cluster in descending order.

Step 7. Select the batch of data points with the first *K* highest membership grades and the membership grades (named *g*) must

be bigger than 0.8 for each cluster in the training dataset *TR*, and delete them from *TR*. If the satisfactory data points for each cluster are less than *K*, add the satisfactory ones to the new training dataset *STR*.

Step 8. Output *STR* as new training subset.

Among the steps discussed above, the membership matrix(*U*) is randomly initialized according to Equation (1):

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,...,n \quad (1)$$

The dissimilarity function used in FCM is given as Equation (2):

$$J(U, c_1, c_2, ..., c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} {u_{ij}}^m {d_{ij}}^2 \quad (2)$$

where, c is the number of clusters, n is the number of data point for each cluster, u_{ij} is between 0 and 1, x_j is the jth data point in the dataset, c_i is the centroid of cluster i; d_{ij} is the Euclidian distance between ith centroid and jth data point; $m \in [1,\infty]$ is a weighting exponent. To reach a minimum of dissimilarity function there are two conditions. These are given in Equation (3) and Equation (4).

$$c_i = \frac{\sum_{j=1}^{n} {u_{ij}}^m x_j}{\sum_{j=1}^{n} {u_{ij}}^m} \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (4)$$

3. EXPERIMENTAL RESULTS

To verify the effectiveness and availability of our methods, we apply it to a real web server DDoS attack detection scenario. We setup a web server located in the college running apache http service (version 2.2) on Linux platform. In the meantime, we deploy a remote monitor host as an end user to experience the QoS of web server and collect the normal training dataset for our TCM-KNN, as well as fulfill the detection tasks. The host is equipped with Intel (R) Pentium (R) IV CPU 3 GHz, 1 GB RAM, 80GB hard disk (7200r/min). We conducted many experiments over several days during busy hours and with background traffic generated from more than 5,000 hosts of the college. In the experiments, we utilized the attackers to access the victim web server and launch well-known DDoS attacks using a series of DDoS tools such as Stacheldraht and TFN2K.

We first used TCM-KNN to detect these traffic anomalies, then adopting instance selection mechanisms (GA and E-FCM for comparison) discussed above to optimize it, we finally selected 5,600 data points from the original collected data points (98,000) as normal training dataset, Table 1 shows the detailed experimental results. The followings are some useful conclusions made from the results:

a) The TP (true positive rate) for TCM-KNN holds high (99.53%) and the FP (false positive rate) after utilizing E-FCM algorithm for training dataset selection is still manageable (1.93%) in real network environment, which is comparable to those when using GA for optimizations (see the third row in Table 1).

b) We found the most important and inspiring result we acquired is that the training time for an anomaly is fairly short compared to that when we used GA for instance selection (training time reduced 76%), which means we can further save a large amount of training and detection cost. Therefore, we may claim that the system based on our optimized TCM-KNN could on-line deal with a large amounts of anomalies in the real network environment and thus make corresponding countermeasures as quick as possible to mitigate them.

Table 1. Experimental results

	Training Time	Detection Time	TP	FP
TCM-KNN (original)	22218.62s	0.4164s	100%	1.28%
TCM-KNN (GA)	363.86s	0.1397s	99.38%	1.87%
TCM-KNN (E-FCM)	87.38s	0.1397s	99.53%	1.93%

4. CONCLUSIONS

This poster presents our ongoing work focusing on how to utilize effectively select training data and thus detect DDoS attack against web server based on lightweight TCM-KNN algorithm compared to our previous work based on GA instance selection mechanism [4, 5]. Relevant experiments demonstrate that it is an excellent method for DDoS detection for web server in real applications.

For our future work, we will further optimize its real-time DDoS detection performance in terms of the real application scenarios. It is worth noting that our E-FCM instance selection mechanism for TCM-KNN algorithm is not amenable to incremental updating for the moment. When the patterns of the network traffic or users change over time, the newly normal data should be periodically merged with the previous data and the update process should be done to ensure its effectiveness. Therefore, offline updating is still a limitation when it is necessary to complicated and changing network environments. Therefore, we would focus on constructing an online training dataset collection and selection mechanism.

5. REFERENCES

[1] G. Carl, G. Kesidis, and S. Rai, "Denial of service attack detection techniques," *IEEE Internet Computing*, 10(1): 82-89, Jan. 2006.

[2] E. Gelenbe, and L. George, "A self-aware approach to denial of service defence," *Elsevier Computer Networks*, 51(5): 1299–1314, Apr. 2007.

[3] E. Ikovic, P. Iher, O. Ahmy, and R. Homas, "Measuring Denial of Service," in *Proc. ACM QoP 2006,* pp. 53-58, 2006.

[4] Y. Li, L. Guo, B. X. Fang, Z. H. Tian, Y. Z. Zhang. "Detecting DDoS Attacks Against Web Server via Lightweight TCM-KNN Algorithm". *In Proc. Of ACM Sigcomm 2008 (SIGCOMM 2008*), pp: 497-498.

[5] Y. Li, L. Guo, Z. H. Tian, T. B. Lu. "A Lightweight Web Server Anomaly Detection Method Based on Transductive Scheme and Genetic Algorithm". *Computer Communications (Elsevier Journal)* 31(17):4018-4025, 2008.