Detecting DDoS Attacks Against Web Server via Lightweight TCM-KNN Algorithm

Yang Li[1,2], Li Guo[1], Bin-Xing Fang[1], Zhi-Hong Tian[1], Yong-Zheng Zhang[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing China 100190
[2]Research Institution of China Mobile, Beijing China 100053
liyang@software.ict.ac.cn

ABSTRACT
In this poster, we firstly put forward to 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.

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

General Terms
Security, Algorithms

Keywords
Web server anomaly detection, TCM-KNN, genetic algorithm

1. INTRODUCTION
Web server is a crucial 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].

In essence, the new measurement could evaluate the impact of DDoS more accurately since "they measure DDoS impact as a percentage of transactions that have not met their QoS requirements and aggregate this measure into several metrics that expose the level of service denial". But they did not give any effective method or algorithm to substantially make use of their measures for DDoS detection in [3], therefore, to adopt it for real applications, it should be further improved and the first and the most important problem to be seriously addressed, in our opinion, is how to integrate all these measures into a reasonable and effective DDoS detection algorithm.

2. OUR METHODS
TCM-KNN is an excellent anomaly detection algorithm, which combines transductive learning method with classical KNN classification algorithm effectively, and thus does well in distinguishing normal and abnormal traffic with high confidence for network anomaly detection domain [4]. Unlike traditional methods in data mining, it can offer measures of reliability to individual points one at a time without a previously built model, and directly uses the points (normal points) in hand to determine the new arrival is normal or not. The core concept, p-value serves as a measure of how well the data belongs to a certain class [4]. Users of transduction as a test of confidence have approximated a universal test for randomness (which is in its general form, non-computable) by using a p-value function called strangeness measure.

Before using TCM-KNN for detection, the consideration of input feature space for TCM-KNN is necessary. For each web transaction as mentioned in [3], we also measure five parameters: (1) one-way delay, (2) request/response delay, (3) packet loss, (4) overall transaction duration and (5) delay variation (jitter). Jointly, these parameters capture a variety of application QoS requirements. Therefore, the first thing for us to detect DDoS attacks according to these measures is mapping them into a feature vector. Using these feature vectors, we can build a pattern model for normal status of web server, and thus use them to distinguish the abnormal traffics from the normal ones (see Figure 1).

To alleviate the expensive computational cost, we utilize GA to fulfill the instance selection task for TCM-KNN. Training dataset is denoted as *TR* with instances, and the search space associated with

SIGCOMM'08, August 17-22, 2008, Seattle, Washington, USA.
ACM 978-1-60558-175-0/08/08.

the instance selection of *TR* is constituted by all the subsets of *TR*. Then, the chromosomes should represent subsets of *TR*. This is accomplished by using a binary representation. A chromosome consists of genes (one for each instance in *TR*) with two possible states: 0 and 1. If the gene is 1, then its associated instance is included in the subset of *TR* represented by the chromosome. If it is 0, then this instance does not occur. After running GA algorithm, the selected chromosomes would be the reduced training dataset for TCM-KNN. We employ four well-known GAs, GGA (Generational Genetic Algorithm), SGA (Steady-State Genetic Algorithm), CHC (heterogenious recombination and cataclysmic mutation) adaptive search algorithm, PBIL (Population-Based Incremental Learning), to fulfill the instance selection tasks.

parameters: k (the nearest neighbors to be used), m (size of training dataset), τ (preset threshold), r (instance to be determined)

/***Training Phase***/

for i = 1 to m {

calculate D_i^y for each point in training dataset and store;

calculate strangeness α for each one in training dataset and store;

}

/***Detection Phase***/

calculate the strangeness for r;

calculate the p-values for r;

if ($p \le \tau$)

claim r as anomaly with confidence ($1-\tau$) and return;

else

claim r is normal with confidence ($1-\tau$) and return;

Figure 1. TCM-KNN for DDoS Attack Detection

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 mechanism 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. From these results, we obtain that:

a) We could determine the anomalous points with accuracy (TP) of 100% (2,600 abnormal points are all correctly diagnosed) and only 1.28% false positives (FP) (only 194 out 15,120 normal data points are misjudged) in the real network environment.

b) After the GA based instance selection optimization, the TP (true positive rate) for TCM-KNN keeps high (99.38%) and the FP (false positive rate) is still manageable (1.87%) in real network environment. Moreover, the training time and detection time are all reduced to a great extent.

c) The most important and inspiring result we acquired is that the detection time for an anomaly is fairly short, thus 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 (optimized)	363.86s.	0.1397s	99.38%	1.87%

4. CONCLUSIONS

This poster presents our work focusing on how to effectively detect DDoS attack against web server based on lightweight TCM-KNN algorithm and thus ensure the QoS of web server. 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.

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, B. X. Fang, L. Guo, and Y. Chen, "Network Anomaly Detection Based on TCM-KNN Algorithm," *in Proc. ACM ASIACCS 2007*, pp. 13-19, 2007.