© 2010 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

An Improved Wavelet Analysis Method for Detecting DDoS Attacks

Liang Fu Lu [1], Mao Lin Huang [2,*], Mehmet A. Orgun [3], Jia Wan Zhang [4]

1. Mathematics Department, Tianjin University, Tianjin, P. R. China
2. Faculty of Engineering and IT, University of Technology, Sydney, Australia
3. Department of Computing, Macquarie University, Sydney, Australia
4. School of Computer Science and Technology, Tianjin University, Tianjin, P. R. China
*Corresponding Author E-mail: maolin@it.uts.edu.au

Abstract—**Wavelet Analysis method is considered as one of the most efficient methods for detecting DDoS attacks. However, during the peak data communication hours with a large amount of data transactions, this method is required to collect too many samples that will greatly increase the computational complexity. Therefore, the real-time response time as well as the accuracy of attack detection becomes very low. To address the above problem, we propose a new DDoS detection method called Modified Wavelet Analysis method which is based on the existing Isomap algorithm and wavelet analysis. In the paper, we present our new model and algorithm for detecting DDoS attacks and demonstrate the reasons of why we enlarge the Hurst's value of the self-similarity in our new approach. Finally we present an experimental evaluation to demonstrate that the proposed method is more efficient than the other traditional methods based on wavelet analysis.**

Keywords-network intrusion detection; distributed denial of service; self-similarity; wavelet analysis; Isomap algorithm

I. INTRODUCTION

With the rapid growth of networked data communications in size and complexity, network administrators today are facing more and more challenges to protect their networked devices from a variety of network attacks. Distributed Denial of Service (DDoS) attack, as one of the most popular attacks, is a very powerful and simple type of the network attack from Internet resources. It uses a large number of zombies to generate a huge volume of useless network transactions sending towards the victim. This "many to one" attacks typically exhaust bandwidth, router processing capacity, or network stack resources, breaking network connectivity to the victim. In addition, the strategies of hierarchical attack and IP spoofing of the DDoS attacks make attackers difficult to be traced [1, 14]. Now DDoS attacks are getting more sophisticated, spreading faster, and causing more damages [2]; and there have not been full defense solutions of DDoS attacks since these attacks have appeared in June 1998 for the first time [3].

Although there have been many studies on detecting DDoS attacks, the proposed solutions cannot always distinguish flooding attacks from abrupt changes of legitimate activity [4]. In 1994, Will E. Leland et al. demonstrated that the normal Ethernet Local Area Network (LAN) traffic is statistically self-similar [5].This theory helps the network analysts greatly in detecting the abnormalities from the legitimate network traffic. Because of the distribution and concealment of DDoS attacks, they can produce noticeable influence on the self-similarity of network traffic.

Therefore, some studies have been conducted on the detection of DDoS attacks via the decrease of Hurst parameter in the self-similarity of network traffic [1, 3, 4, 6-9]. These methods require statistics of self-similarity of normal network traffic before the attack occurs. Ren and Li et al. commented that the wavelet analysis method is much more useful than other existing methods, such as variance time plot (VTP), periodogram, R/S analysis, and Whittle's estimator etc. in computing the Hurst parameter. Although the wavelet analysis method can detect DDoS attacks. during the busy traffic times, it needs to collect too many samples which will increase the computational complexity of the method significantly. This will in turn affect the real-time response time and the accuracy of the detection process.

To overcome the above weaknesses of the traditional methods, we propose a new method called "Modified Wavelet Analysis" that is based on dimension reduction and wavelet analysis. The new algorithm can make the attributes of the data highly abstracted. It enhances the relevance of the attributes of the data, and enlarges the change of the Hurst parameter between the normal and abnormal network traffic patterns. Thus the new approach can avoid the false-negative and positive-negative findings effectively. Moreover, the new method can reduce the dimension of the network data greatly, so the processed data can then be used for analysis more efficiently. The experiments show that the efficiency of our method can be achieved in many different perspectives.

The rest of the paper is organized as follows. We first briefly introduce the related work in Section 2. Then we propose our new method in Section 3. In Section 4, through experimental evaluation, attack traffic is studied with our new method. Finally, we conclude the paper with a brief summary and discussion of future work in Section 5.

II. RELATED WORK

When a network is flooded by DDoS attacks, the normal data transformations of TCP packets in that network will be jammed and then the self-similarity of the network traffic will be decreased noticeably. If the network is overwhelmed with DDoS attack packets, the traffic model tends to be a

Poisson distribution [4]. As stated in [4, 5], the main reason of causing self-similarity of the traffic is the heavy-tailed distributions of file sizes. The Hurst value (H) is the only parameter of the self-similarity, whose range is between $(0.5,\ 1)$. The larger the value of H , the higher the possibility of self-similarity in the network traffic. So, the differentiation of the value of H in the abnormal and the normal network traffic can help in identifying DDoS attacks, which is much easier than the common feature matching based approaches. Different from the traditional statistical detection methods, the Hurst value analysis can distinguish between busy traffic and DDoS attacks.

Several DDoS detection methods [1, 4, 6] have been proposed based on wavelet analysis. These methods can distinguish between busy traffic and DDoS attacks, and are sensitive to the change of the Hurst parameter. Ren Xunyi et al. [8] have compared the Rescales Range Analysis and the Wavelet Analysis method and the result shows that the latter is more fit and more sensitive than R/S in detecting DDoS type of the network attacks. Yongli, et al. [7] have analyzed and examined the existing wavelet analysis methods that are based on the energy method, such as variance of wavelet coefficient method, spectrum estimation method and energy method.

When we calculate the Hurst value in wavelet analysis, there are two things must be considered carefully: the choice of different wavelet functions and the influence of vanishing moments on the estimation of the Hurst parameter. It is still an open problem to decide which wavelet function would be the best for estimating H .

Ren Xunyi et al. [9] have carried out a large number of experiments with three different wavelet bases (dbN, symN, coifN) and showed that the average change of Hurst estimations by dbN is maximized when DDoS attacks take place. Therefore we can describe the traditional DDoS attack detection method based on wavelet analysis as follows:

Algorithm 2.1 DDoS attack detection method based on wavelet analysis [4]:

Step 1: Extract the data from the character library and generate a matrix N .

Step 2: Choose the Db(3) as the analysis wavelet and compute its maximum scale.

Step 3: Use the wavelet function to decompose the signal N , and obtain the wavelet coefficient matrix $[C, L]$.

Step 4: When ($i \leq$ maximum scale), extract high frequency wavelet coefficient $L(i)$ from $[C, L]$ and compute the value of $\log_2 Var[d_l(i)]$, where $Var[d_l(i)]$ is the variance of $d_l(i)$.

Step 5: Use the value of $[i, \log_2 Var[d_l(i)]]$ to draw a straight line, and then calculate the slope k of the line. We then get the Hurst parameter: $H = (k-1)/2$.

Step 6: For a given $\Delta h = H_n - H_a$, and a threshold θ, if $\Delta h > \theta$, then we believe that a DDoS attack is occurring, otherwise no DDoS attacks have occurred.

DDoS attacks always generate enormous data packets to send to the target through a large number of agents. They can easily exhaust the computing and communication resources of the victim within a short period of time. The attributes of data packets usually appear in multi-dimensional data format, in which there is a large amount of less important and meaningless information. Therefore, it is necessary to use some dimension reduction algorithms to extract more relevant and important information that can be presented in low-dimensional structures for easy analysis and detection of DDoS attacks.

The classical techniques for manifold learning, such as Principal Components Analysis (PCA) and Multi-Dimensional Scaling (MDS), are designed to operate when the sub-manifold is embedded linearly, or almost linearly, in the observation space. However, these algorithms often fail due to the nonlinear structures of the network packets. Since the computation of the Hurst parameter in the self-similarity of network traffic rely heavily on the global structure, and analysts therefore wish to keep more information of the data packages through the dimension reduction in which the geometrical structures can be preserved.

Thus, we use a nonlinear global dimension reduction method called Isomap [10] to recover the intrinsic geometric structure of the data manifolds. Isomap is a global algorithm which attempts to preserve the geometry at all scales, mapping nearby points on the manifold to nearby points in a low-dimensional space, and the faraway points to faraway points[11].

The four steps of the Isomap algorithm is described below[10]:

Step 1: Calculate the k nearest neighbors of data X_i from the observed data set $\{X_i\}$. For a given nonnegative real number k , if the Euclidean length of the two data items $\left|X_i - X_j\right| < k$, we consider X_j as one of the k nearest neighbors of data X_i (and vice versa).

Step 2: Determine a neighborhood graph G of the data set $\{X_i\}$. If G contains the edge ($X_i X_j$), then X_j

is one of the k nearest neighbors of data X_i (and vice versa).

Step 3: Compute the shortest paths in graph G for all pairs of data points. Each edge ($X_i X_j$) in the graph is weighted by the Euclidean length $\left|X_i - X_j\right|$.

Step 4: Apply MDS to the resulting shortest-path distance matrix $D_G = \{d_G(i, j)\}$ to find a new embedding of the data in Euclidean space.

III. OUR PROPOSED METHOD-BASED ON ISOMAP DIMENSION REDUCTION AND WAVELET ANALYSIS

The wavelet analysis with the calculation of the Hurst parameter is always carried out with a threshold to determine whether a DDoS attack is occurring or not [1, 4]. Usually with normal network traffic, the Hurst value of self-similarity should be around *0.75.* If the Hurst value is less than *0.5*, we will consider that the network traffic is out of self-similarity. Thus, we always set the threshold θ in Algorithm 2.1 to less than *0.25*. So if we choose an inappropriate θ the detection methods will cause a large volume of false-negative and positive-negative findings. The sample capacity of wavelet analysis will be large, and the process of calculation will become very slow, the requirements of calculation ability and storage capacity are high. Therefore, the current methods do not work very well in detecting weak DDoS attacks. Fig.1 shows the framework of our new approach. The new method is implemented by Algorithm 3.1 given below:

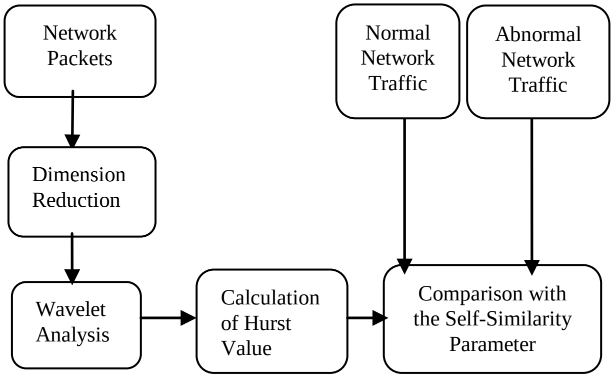

Figure 1. Framework of the new method

The new algorithm 3.1 DDoS attack detection method based on wavelet analysis [4] and Isomap dimension reduction [10]**:**

Step 1: Extract the data from the character library and make up the matrix $M = [m_1, m_2, \cdots, m_n]$, where m_i is a network data item.

Step 2: Select the nonnegative real number k and use the Isomap algorithm to reduce the dimension of M . The corresponding new data after the dimension reduction is: $X = \left[X_1, X_2, \cdots, X_n\right]$:

- Calculate the k nearest neighbors of data X_i from the observed data sets $\left\{X_i\right\}$. For a given nonnegative real number k, if the Euclidean length of the two data satisfies $\left|X_i - X_j\right| < k$, we then consider X_j as one of the k nearest neighbors of X_i (and vice versa).
- Determine a neighborhood graph G of data sets $\left\{X_i\right\}$. G contains the edge ($X_i X_j$) only if X_j is one of the k nearest neighbor of X_i (and vice versa).
- Compute the shortest paths in G for all pairs of data points. Each edge $X_i X_j$ in the graph is weighted by the Euclidean length $\left|X_i - X_j\right|$.
- Apply MDS to the resulting shortest-path distance matrix $D_G = \{d_G(i, j)\}$ to find a new embedding of the data in Euclidean space.

Step 3: Choose the Db(3) for wavelet analysis and compute its maximum scale.

Step 4: Use wavelet function to decompose the signal X , and obtain the wavelet coefficient matrix $[C, L]$.

Step 5: When $i \le$ maximum scale , extract the high frequency wavelet coefficient $L(i)$ from $[C, L]$ and then calculate $\log_2 Var[d_l(i)]$, where $Var[d_l(i)]$ is the variance of $d_l(i)$.

Step 6: Using $[i, \log_2 Var[d_l(i)]]$ to draw a straight line, and calculate the slope k . We then have: Hurst $= (k-1)/2$.

Step 7: For a given $\Delta h = H_n - H_a$, and a threshold θ. We assume a DDoS attack occurs if $\Delta h > \theta$.

Isomap is a dimension reduction algorithm that can preserve the global geometric structure of the data. This method takes the advantages of both PCA and MDS including 1) computational efficiency, 2) few free parameters, 3) non-iterative global optimization of a natural cost function—with the ability to recover the intrinsic geometric structure of a broad class of nonlinear data manifolds[11].

From the above discussion we can see that the reason of self-similarity of network traffic is the heavy-tailed distributions of file sizes. Self-similarity means that the whole and parts (or between different parts) of the traffic possess the similar characteristics, i.e., they all have the invariance in geometric or nonlinear transformations. Isomap is a nonlinear method that can better preserve the original properties of the data than the traditional linear algorithms. Thus we choose Isomap for the reduction of the dimensions of network data to decrease the complexity and increase the relevance of analytical data. This method enlarges the differentiation of the Hurst value between normal and abnormal network traffics. This alleviates the dependence of the wavelet analysis on the threshold θ and avoids the false-negative and positive-negative findings greatly.

IV. CASE STUDIES

To evaluate our new method, we have selected a set of data with different volumes from the KDD99 Cup dataset randomly. The KDD99 dataset used for The Third International Knowledge Discovery and Data Mining Tools Competition [13] contains 5 weeks of *tcpdumps* from a simulated network from DARPA. Although the KDD99 dataset has been shown to have certain deficiencies, it is still the only publicly available large benchmark dataset which may assist in the baseline comparison of different methods[12]. The attack types include teardrop and neptune. We choose 0.15 as the value of threshold θ, which means that if the change of the Hurst value becomes larger than 0.15, we then consider that a DDoS attack may be occurring. Table 1 given below provides the comparison between the results of our new approach and the traditional approach. In Table 1, the result for the old method is labeled as “H-Wavelet” and the result for the new method is labeled as “H-Combined”.

Table 1 shows the results of comparison among three sets of data. The first data set includes 2,436 data items, where 921 data items are neptune attacks, the others are normal network traffic. The change of the Hurst parameter in traditional Wavelet Analysis method is 0.0347, which is far less than 0.15, while the change of the Hurst parameter in our new method is 0.1556 which is larger than 0.15.This indicates that by using the new method, we can detect the DDoS attacks while the traditional method does not perform well. The situation of the third data set is the same as the first one, in which there are 448 data items including 155 neptune data items.

A set of 6,110 data items, which consist of 921 neptune attack data items and 459 ipsweep attack data items, makes up the second group. Unfortunately, both methods cannot detect DDoS attacks directly because of the change of the Hurst value is always less than 0.15. This could be because these two different types of attacks have been mixed up into a new pattern which is very similar to that of the normal traffic. On the other hand, our method still enlarges the change of the Hurst value. Fig.2(a) and Fig.2(b) show the oscillograms of the normal and the abnormal network traffic of the second data set respectively. We have also conducted some other experiments to evaluate the new approach. All of the experimental results show that the new approach can detect the weak DDoS attacks very well.

TABLE 1. THE COMPARISONS BETWEEN THE NEW METHOD AND THE TRADITIONAL METHOD BY USING THREE SEPARATE DATA SETS.

	Group 1	Group 2	Group3
H-Wavelet	0.0347	0.0042	0.1206
H-Combined	0.1556	0.0778	0.2093

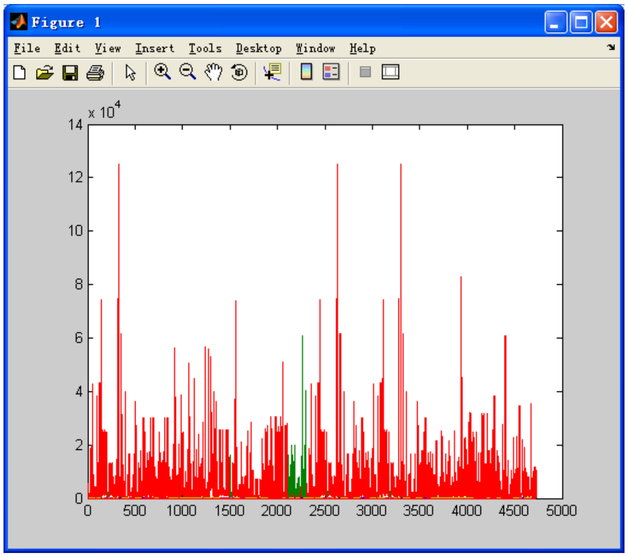

Figure 2. (a) Normal network traffic

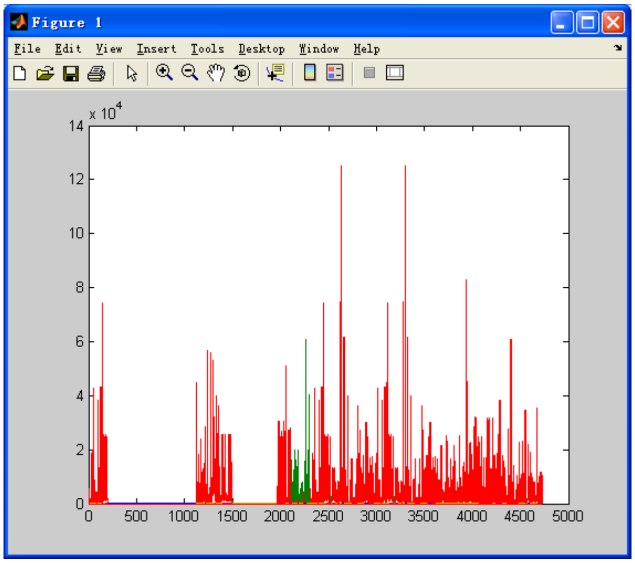

Figure 2. (b) Abnormal network traffic

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an improved DDoS attack detection method that combines the traditional Wavelet Analysis method and the Isomap dimension-deduction method. The Isomap algorithm can be used to not only reduce the dimensions of the network traffic data, but also to enhance the relevance of the analytical network data. The new method can help us to detect the weak DDoS attacks by enlarging the differentiation of the Hurst parameter. We then can use the Wavelet Analysis method to calculate the self-similarity of the network traffic data. The experiments have shown that the new method can detect weak DDoS attacks better than the traditional method. The experiments also show that the new approach can avoid the false-negative and positive-negative findings more effectively. On the other hand, the iteration of the Isomap algorithm is an extra step of computation in the new method which will have an impact on its performance. Thus in the future we will attempt to use parallel algorithms or GPU methods to tackle the real-time detection of DDoS attacks and to validate our method.

REFERENCES

[1]. G.L. Lan Li, “DDoS attack detection and wavelets,” Telecommunication Systems, Volume 28, Numbers 3-4 / March, 2005, pp. 435-451.

[2]. X. Zhou and C.-Z. Xu, “Distributed denial-of-service and intrusion detection,” Journal of Network and Computer Applications, vol. 30, no. 3, 2007, pp. 819-822.

[3]. K. Lee, et al., “DDoS attack detection method using cluster analysis,” Expert Syst. Appl., vol. 34, no. 3, 2008, pp. 1659-1665.

[4]. Ren Xunyi, et al., “Wavelet analysis method for detection of DDoS attack on the basis of self-similarity,” Frontiers of Electrical and Electronic Engineering in China, vol. 2, no. 1, 2007, pp. 73-77.

[5]. M.S.T. Will E. Leland, W alter Willinger,Daniel V. Wilson, “On the self-similar nature of Ethemet traffic,” IEEE/ACM Trans on Networking[Extended Version], vol. 2, no. 1, 1994, pp. 1-15.

[6]. Gu Junjia and L. Ning, “Detection of DDoS Attack Flow in Web Traffic Based on Wavelet Analys,” Computer engineering and applications(In Chinese), vol. 42, no. 5, 2006, pp. 127-130.

[7]. Li Yongli, et al., “On wavelet -based methods for hurst index estimation of self-similar traffic,” Journal of ELECTRONICS AND INFORMATION TECHNOLOGY, vol. 25, no. 1, 2003, pp. 100-105.

[8]. Ren Xunyi, et al., “Study and comparison of R/S and wavelet analysis for DDoS attack detection,” Journal of Nnajing University of Posts and Telecommunications(Natural Science)(In Chinese), vol. 26, no. 6, 2006, pp. 48-51.

[9]. Ren Xunyi, et al., “Wavelet Choice for Detection of DDoS Attack Based on self一Similar Testing,” Journal of Nanjin Universlty of Aeronautics and Astronautics(In Chinese), vol. 39, no. 5, 2007, pp. 588-592.

[10]. Joshua B. Tenenbaum, et al., “A Global Geometric Framework for Nonlinear Dimensionality Reduction,” Science, vol. 290, 2000, pp. 2319-2323.

[11]. Vin de Silva and J.B. Tenenbaum, “Global versus local methods in nonlinear dimensionality reduction,” Advances in Neural Information Processing Systems 15 vol. 15, no. 2, 2003, pp. 705-712.

[12]. Huy Anh Nguyen, Deokjai Choi: Application of Data Mining to Network Intrusion Detection: Classifier Selection Model. In the Proceedings of 11th Asia-Pacific Network Operations and Management Symposium, APNOMS 2008, Beijing, China, October 22-24, 2008. Lecture Notes in Computer Science 5297 Springer 2008 , pp. 399-408

[13]. KDD99 Cup dataset; http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (last accessed 4 May 2010)

[14]. Shui Yu, Wanlei Zhou and Robin Doss, "Information Theory Based Detection Against Network Behavior Mimicking DDoS Attack," IEEE Communications Letters, vol. 12, no. 4, April 2008, pp. 319-321.