An On-line Anomaly Detection Method Based on LMS algorithm

Ziyu Wang*, Jiahai Yang† and Fuliang Li‡
*Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100084, China,
Tsinghua National Laboratory for Information Science and Technology(TNList)
Email: see wangziyu11@mails.tsinghua.edu.cn
† Email: see yang@cernet.edu.cn
‡ Email: see lifuliang207@126.com

***Abstract*—Anomaly detection has been a hot topic in recent years due to its capability of detecting zero attacks. In this paper, we propose a new on-line anomaly detection method based on LMS algorithm. The basic idea of the LMS-based detector is to predict IGTE using IGFE, given the high linear correlation between them. Using the artificial synthetic data, it is shown that the LMS-based detector possesses strong detection capability, and its false positive rate is within acceptable scope.**

***Index Terms*—anomaly detection, Least Mean Square, IGTE, IGFE**

I. Introduction

Network anomalies have been serious challenges for the Internet nowadays. There are basically two classes of detection methods. The first class is called misuse detection, also known as signature-based detection [1]. The primary advantage of misuse detection is its high degree of accuracy. However, the misuse detection is incapable of detecting zero day attacks whose features are not known in advance. The second class of detection methods is called anomaly detection [2], [3]. Anomaly detection only cares about the statistical properties of network traffic rather than specific anomaly features. Hence, it is capable of detecting zero day attacks. This capability is the strong advantage of anomaly detection over misuse detection. Hence, anomaly detection has been well studied by researchers in recent years [4], [5], [6].

However, most anomaly detection methods operate in an off-line fashion, such as the famous wavelet-based detector and PCA-based detector [7], [8], etc. They do not output any results until the entire data set has been collected and stored. This batch mode would inevitably cause serious time delays.

In this paper, we propose a new on-line detection method based on the linear correlation between two metrics—IGTE and IGFE. We make use of IGFE to predict IGTE using the LMS algorithm [9]. Then we take the prediction error as the criterion for anomaly detection. The LMS-based detector can operate in a real-time fashion using only the data points collected before and at the current step instead of the entire data set.

The main contributions of this paper are: (1) validating the highly linear correlation between IGTE and IGFE; (2) proposing an on-line flow-based anomaly detection method; (3) validating the effectiveness of the LMS-based detector quantitatively using artificial synthetic data.

The remainder of this paper is organized as follows. In section 2, we introduce the procedure of generating IGTE and IGFE series and validate the high linear correlation between them. We explain the principles and rationales of the LMS-based detector in section 3. In section 4, we describe the data source used in this paper. We validate the effectiveness of the LMS-based detector and compare it with the famous wavelet-based detector in section 5. We conclude this work in section 6.

II. IGTE and IGFE

Network traffic is composed of millions of IP flows. Anomalies exists in certain number of IP flows. Theoretically, checking each IP flow is the most high-precision way to detect anomalies. However, the number of IP flows is usually extremely huge in today's network traffic. Analyzing each IP flow is high-cost and impractical. In this paper, we make a tradeoff between computation complexity and information granularity of anomalies by mapping each IP flow into different groups. The number of these groups are much smaller than the number of IP flows. Analyzing these groups rather than each IP flow can largely reduce the overhead of anomaly detection. Meanwhile, this mapping practice preserves the information of anomalies in as much detail as possible. The procedure of mapping IP flows is given as follows.

An IP flow can be characterized by its five-tuple value(Source IP address, Destination IP address, Source port, Destination port and Protocol type). Using the five-tuple value as key, we can hash the IP flows into different groups. Denote the number of the groups by p, and the number of time intervals by t. When we count the number of IP flows mapped into each group, a $t \times p$ matrix called Randomly Aggregated Flow Matrix(RAFM) is generated. The (i, j) entry of RAFM corresponds to the number of IP flows in group j at instant i. Similarly, when we calculate the overall traffic volume of the IP flows mapped into each group, a $t \times p$ matrix called Randomly Aggregated Traffic Matrix(RATM) is generated. The (i, j) entry of RATM corresponds to the overall traffic volume of group j at time instant i.

For each row of RAFM, we calculate the sample entropy of its distribution. Then we have a series with length t called IGFE(Inter-group Flow Entropy) series. From the point of view of information theory, the IGFE can be seen as a summarization tool for the distribution of RAFM. For each row of RATM, we calculate the sample entropy of the distribution. Then again we have a series with length t called IGTE(Inter-group Traffic Entropy) series. The IGTE can be seen as a summarization tool for the distribution of RATM.

Denote RAFM by F, we define the IGFE series as follows:

$$IGFE[i] = -\sum_{j=1}^{p}\left\{\frac{F(i,j)}{\sum_{j=1}^{p}F(i,j)}\ln\frac{F(i,j)}{\sum_{j=1}^{p}F(i,j)}\right\} \quad (1)$$
$$i = 1, 2, ..., t$$

Denote RATM by T, we define IGTE series as follows:

$$IGTE[i] = -\sum_{j=1}^{p}\left\{\frac{T(i,j)}{\sum_{j=1}^{p}T(i,j)}\ln\frac{T(i,j)}{\sum_{j=1}^{p}T(i,j)}\right\} \quad (2)$$
$$i = 1, 2, ..., t$$

We find that IGTE and IGFE are highly linearly correlated. From three days Netflow records collected on a border router in CERNET2(an academic network in China which will be described in detail later), we calculate the corresponding IGTE and IGFE series, which are shown in Figure 1. We observe that the IGTE and IGFE series evolve almost synchronously. The two curves are very similar in shape, which implies the high linear correlation between them. In order to validate this linear relationship rigorously, we estimate the correlation coefficient between the IGFE and IGTE series as follows:

$$\rho = \frac{\sum_{i=1}^{t}(IGTE[i]-\overline{IGTE})\times(IGFE[i]-\overline{IGFE})}{\sqrt{\sum_{i=1}^{t}(IGTE[i]-\overline{IGTE})^2\times\sum_{i=1}^{t}(IGTE[i]-\overline{IGTE})^2}} \quad (3)$$

where $\overline{IGTE} = \frac{\sum_{i=1}^{t}IGTE[i]}{t}$ and $\overline{IGFE} = \frac{\sum_{i=1}^{t}IGFE[i]}{t}$ are the sample means of IGTE and IGFE series. Apply the above formula to the three days CERNET2 data, we have $\rho = 0.976$, which means that IGTE and IGFE are indeed highly correlated.

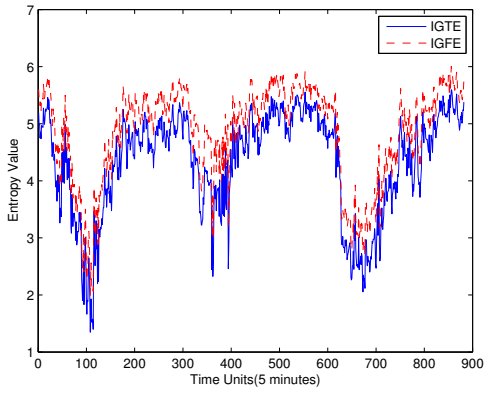

Fig. 1. IGFE series versus IGTE series from CERNET2

III. Detection methods based on LMS algorithm

Based on the high linear correlation between IGTE and IGFE, it is reasonable to predict IGTE using IGFE. In order to achieve this goal, we adopt the Least Mean Square algorithm, which is a widely used adaptive filter algorithm [9]. If the IGFE series cannot precisely predict the IGTE series, it means that anomalies may occur in the network.

A. Wiener filter

In this paper, the general framework for prediction is traverse filter. Its structure is illustrated in Figure 2 [10]. The length of the traverse filter is denoted by $N+1$. The IGFE series is used as the input signal, denoted by $x(k)$. The input signal vector of the traversal filter at time k is denoted by $X(k) = [x_0(k), x_1(k), ..., x_N(k)]^T$, where $x_0(k) = x(k), x_1(k) = x(k-1), ..., x_N(k) = x(k-N)$. $W(k) = [w_0(k), w_1(k), ..., w_N(k)]^T$ is the coefficient vector of the filter, which is to be determined. The signal $d(k)$ is the so called reference or desired signal, which is specified as the IGTE series in this paper. Denote $y(k)$ as the response signal of the traversal filter corresponding to the input signal $x(k)$.

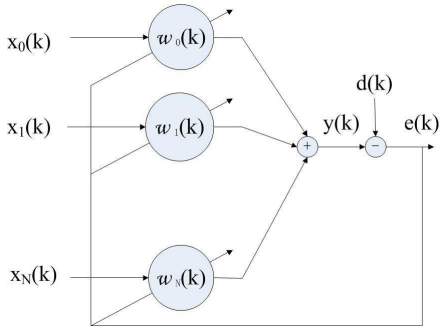

Fig. 2. Architecture of The Traverse Filter

Given these notations, we have

$$y(k) = \sum_{i=0}^{N} w_i(k)x_i(k) = W^T(k)X(k) \quad (4)$$

Note that the predicting task is equal to ensure $y(k)$ and $d(k)$ as close as possible. The difference between $y(k)$ and $d(k)$ is called the error signal, defined as follows:

$$e(k) = d(k) - y(k) \quad (5)$$

The metric used to measure the degree of the difference between $y(k)$ and $d(k)$ is MSE(Mean Square Error), which is defined as follows:

$$\xi(k) = E[e^2(k)] \quad (6)$$

Substituting equation (4) and equation (5) into equation (6), we have

$$\xi(k) = E[d^2(k) - 2d(k)y(k) + y^2(k)] \quad (7)$$

The MSE $\xi(k)$ is also known as the objective function in the field of optimization. Thus the problem of prediction is

turned into minimizing the MSE function—$\xi(k)$. When the objective of the traverse filter is to minimize certain kind of objective function, it is known as the Wiener filter.

From equation (4), the MSE function can be rewritten as [10]

$$\begin{aligned}\xi(k) =&E[d^2(k) - 2d(k)W^T(k)X(k)+\\ &W^T(k)X(k)X^T(k)W(k)]\\ =&E[d^2(k)] - 2E[d(k)W^T(k)X(k)]+\\ &E[W^T(k)X(k)X^T(k)W(k)]\end{aligned} \tag{8}$$

Assume the coefficient vector of the Wiener filter is fixed and the environment is stationary, the MSE function is given by

$$\xi = E[d^2(k)] - 2W^T P + W^T RW \tag{9}$$

where $P = E[d(k)X(k)]$ is the cross-correlation vector between the reference and input signals, and $R = E[X(k)X^T(k)]$ is the correlation matrix of the input signal. Note that in equation (9) ,the time indexes of the coefficient vector and MSE function are removed due to the previous assumptions.

In order to minimize ξ, we calculate its gradient vector g_W related to W. We have

$$g_W = \frac{\partial \xi}{\partial W} = -2P + 2RW \tag{10}$$

By equating the gradient vector to zero, the optimal coefficient vector which minimizes the MSE function ξ can be evaluated as follows:

$$W_0(k) = R^{-1}P \tag{11}$$

This solution is called the Wiener solution [9]. Choosing the Wiener solution as the coefficient vector, we can precisely predict the IGTE series with the IGFE series to the greatest extent.

Unfortunately, we can not directly obtain the Wiener solution in practice. The difficulties lie in that it is usually hard to precisely estimate R and P since the stationary prerequisite required by the Wiener solution is usually not satisfied, especially in network environment. The solution for this situation is the LMS algorithm, which will be described later.

B. Least Mean Square Algorithm

Since the Wiener solution can not be directly calculated, we update the coefficient vector according to the famous steepest-descent algorithm in optimization theory [11]:

$$W(k+1) = W(k) - \mu g_W(k) \tag{12}$$

where μ is the convergence factor. Note that the coefficient vector is no longer fixed, but varies over time. Hence, the variables in the above equation are all associated with a time index.

However, the same problems are still unsolved. Recall equation (10), the value of $g_W(k)$ depends on R and P, which are not known in advance and are usually time-varying. It means that during each step of iteration, we need to re-estimate the current values of R and P. One possible solution is to estimate them using their instant samples. Denote the two estimates by $\hat{R}$ and $\hat{P}$, we have

$$\begin{aligned}\hat{R} &= X(k)X(k)^T\\ \hat{P} &= d(k)X(k)\end{aligned} \tag{13}$$

Substituting the above equation into equation (10), we have

$$\begin{aligned}\hat{g}_W =&2X(k)(-d(k) + X^T(k)W(k))\\ =&-2e(k)X(k)\end{aligned} \tag{14}$$

where $e(k)$ is the error signal defined in equation (5).

Substituting equation (14) back into equation (12), we have

$$W(k+1) = W(k) - 2\mu e(k)X(k) \tag{15}$$

The above updating equation is called the Least Mean Square(LMS) algorithm, which forms the core of our anomaly detection algorithm. When the traverse filter shown in Figure 2 updates $W(k)$ according to equation (15), it is known as the LMS adaptive filter.

C. Detection Algorithm

The updating procedure of the LMS algorithm tends to reduce the difference between the output signal $y(k)$ and the reference signal $d(k)$ to the greatest extent. As long as $d(k)$ and the input signal $x(k)$ are highly linearly correlated, the resulting error signal $e(k)$ should be close to zero. Hence, we propose an online anomaly detection algorithm based on the LMS algorithm.

Take the IGFE series as the input signal $x(k)$, and the IGTE series as the reference signal $d(k)$. The procedure of detection is as follows: first, calculate the IGFE value $x(k)$ and IGTE value $d(k)$ in the current step. Along with previous N IGFE values, we construct the current input signal vector $X(k) = [x(k), x(k-1), ..., x(k-N)]^T$. Using the coefficient vector $W(k)$ obtained in time instant $k-1$, we calculate the output signal of the LMS filter by $y(k) = X^T(k) \times W(k)$. From equation (5), we generate the error signal $e(k)$. Once $|e(k)|$ exceed certain threshold, an alarm is triggered. Then update the coefficient vector according to equation (15) for the next step. The details is illustrated in Algorithm 1.

Note that the LMS-based detector is an on-line method in its nature. That is because it generates the error signal in each step, and reports the detection results in real time. In other words, it does not need to wait for the data to be collected completely before operation. At each time instant, it can detect anomalies using past data points which are collected in previous steps. This real time property makes the LMS-based detector suitable for on-line applications.

D. Rationale Behind LMS-based Detection Method

Anomalies usually change the number of IP flows on the link or the traffic volume of certain IP flows. Some anomalies such as port scans, would generate lots of small IP flows in the network. This leads to a large increase in the number

Algorithm 1 LMS-based anomaly detection algorithm

Input: Raw IP flow records from time instant 1 to t;
Threshold τ;
Convergence factor μ;
Output: Anomalous time intervals;

1: **for all** k such that $1 \leq k \leq t$ **do**
2: Hash the IP flows in time interval i into p different groups;
3: Generate the kth row of RATM T;
4: Generate the kth row of RAFM F;
5: $d(k) = -\sum_{j=1}^{p} \left\{ \frac{T(i,j)}{\sum_{j=1}^{p} T(i,j)} \ln \frac{T(i,j)}{\sum_{j=}^{p} T(i,j)} \right\}$;
6: $x(k) = -\sum_{j=1}^{p} \left\{ \frac{F(i,j)}{\sum_{j=1}^{p} F(i,j)} \ln \frac{F(i,j)}{\sum_{j=1}^{p} F(i,j)} \right\}$;
7: $X(k) = [x(k), x(k-1), ..., x(k-N)]^T$;
8: $e(k) = d(k) - X^T(k) \times W(k)$;
9: $W(k+1) = W(k) - 2\mu e(k) X(k)$;
10:
11: **if** $|e(k)| > \tau$ **then**
12: Output: Time interval k;
13: **end if**
14: **end for**

of IP flows, which changes the IGFE value dramatically. However, the traffic volume generated by the anomalies is very small compared to the overall traffic volume on the link, which barely changes the IGTE value. Therefore, the linear correlation between IGTE and IGFE is destroyed, which means we can not precisely predict IGTE using IGFE. This would result in large error signals, which can be detected by the LMS-based detector.

Some anomalies such as DDoS attacks, would increase the number of IP flows and the traffic volume at the same time. However, the magnitude of traffic volume change is usually much larger than the number of IP flows. Hence, the degree of change of IGTE is much larger than IGFE. This results in the breach of the linear relation between IGTE and IGFE. Thus the anomalies would be detected by the LMS-based detector.

There are also some anomalies which would increase the number of IP flows but decrease the traffic volume on the link. Take Low-rate DDoS attacks [12] as an example, the attackers would generate millions of attacking IP flows, which will definitely change the IGFE value. On the other hand, the traffic volume generated by the attacking IP flows is very low on average, since these attacks are performed in the form of pulses. At the same time, the traffic volume of normal IP flows would be reduced dramatically due to the misuse of congestion control mechanism in network. Therefore the overall traffic on the link would decrease dramatically, which would change the IGTE value. Though both IGTE and IGFE change, they change in opposite directions, which would destroy the linear correlation between them. This will generate large error signals, which can be detected by the LMS-based detector.

IV. Data Source

The data used in this paper is three-day Netflow Records collected from a border router in the Second Generation of China Education and Research Network (CERNET2). CERNET2 connects 25 PoPs including Peking University, Tsinghua University, Beijing University of Aeronautics and Astronautics(Beihang University), University of Science and Technology ,etc. The border router used for collecting data connects CERNET2 backbone and Beihang University Campus Network. Due to the high volume of traffic in backbone, the sampling rate is set to $1:1000$. The Netflow Records are exported every five minutes. The five-tuple value, the traffic volume of each IP flow are recorded. In five minutes, the average traffic volume is about 1.525×10^8 bytes, the average traffic volume of each IP flow is about 985 bytes, and the average number of IP flows is about 154730. Note that these numbers are based on the sampled data, the real values of them should be multiplied by 1000 for the original network data.

V. Validation

We choose the number of groups into which the IP flows are hashed as 1024. From the three-day CERNET2 data, we generate an IGTE series and an IGFE series, both with length 882. Note that in practice, the IGTE and IGFE values should be calculated in real time for on line detection. The reason for the "off-line" fashion here is that we can conveniently control the procedure of injecting artificial anomalies into the CERNET2 data.

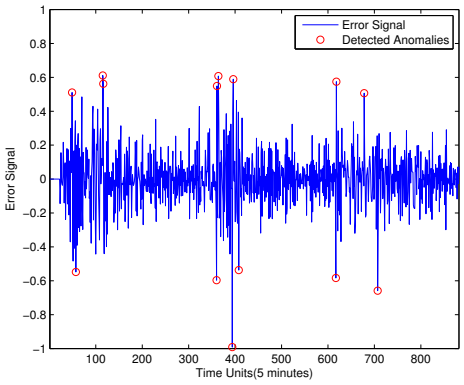

Fig. 3. Anomalies detected by the LMS-based detector for CERNET2 data

In this paper, we first apply the LMS-based detector on the CERNET2 data to pinpoint suspicious data points and generate "cleaned" data set. Then we manually inject artificial anomalies into the "cleaned" data. We apply the LMS-based detector on this synthetical data to evaluate its effectiveness quantitatively, that is, to calculate its true positive rate(detection rate) and false positive rate. The definitions of true positive rate and false positive rate in this paper originate from the introductory document about ROC analysis [13]. We also apply the widely applied wavelet-based detector [7] on the synthetic data for the purpose of comparison.

A. Generation of "Clean" Traffic

We apply the LMS-based detector on the three-day CERNET2 Netflow data, the resulting error signal is illustrated in Figure 3. The detected anomalies are marked with red circles. Totally 14 data points are reported as anomalies. Since the CERNET2 Netflow Records are not benchmark data, we do not know in advance which points are true anomalies and which are not. Thus we can not claim that the data points pinpointed by the LMS-based detector are true anomalies. The only thing we can be sure is that these points are different from the others in the point of view of statistics. Therefore, we can not quantitatively evaluate the effectiveness of the LMS-based detector. We bypass this difficulty by using artificial synthetic data. In order to generate the synthetic data, we need to generate the so-called "clean" data set first. We simply achieve this by removing the 14 detected anomalous points to eliminate their impacts on calculating the true and false positive rates of the LMS-based detector. We can not be sure whether the "clean" data is really "clean". However, we manually check the log files of a large number of servers in CERNET2, they show no obvious trails of attacks during the three days when the Netflow Records are collected. It means that even if the "clean" data contained some anomalies, the number would be too small to impact the results of the validation process.

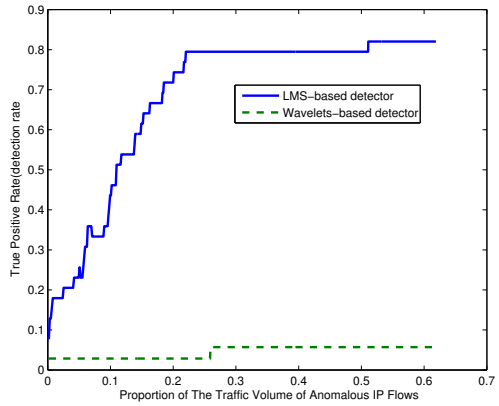

Fig. 4. True Positive Rate for A Small Number of Anomalous IP Flows

B. Experiments on Synthetic Data

In order to evaluate the effectiveness of LMS-based detector quantitatively and rigorously, we manually injected anomalies into the "clean" data. First, we inject certain number of anomalous IP flows every 22 time intervals into the "clean" data. There are totally 35 artificial anomalies contained in the synthetic data. We do not inject any anomaly in the first 30 time intervals. The reason is that it needs some time for the LMS-based detector to converge. After that, we apply the LMS-detector on the synthetic data and calculate the true positive rate and false positive rate respectively.

According to the number of IP flows related to the injected anomalies, we evaluate the effectiveness of the LMS-based detector in two cases:

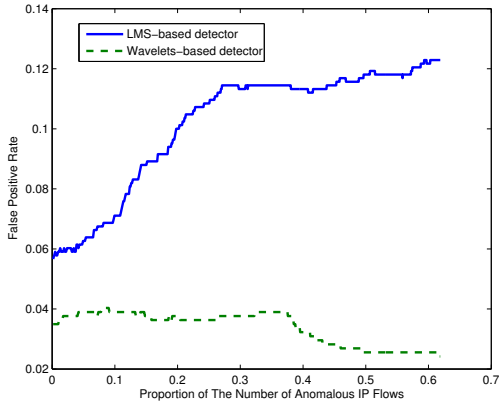

Fig. 5. False Positive Rate for A Small Number of Anomalous IP Flows

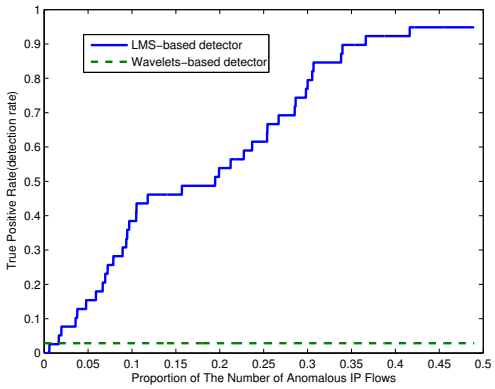

Fig. 6. True Positive Rate for A Large Number of Anomalous IP Flows

- Anomalies involving a small number of IP flows.
- Anomalies involving many small IP flows.

Note that we ignore the case where the anomalies involve many large IP flows on purpose. Because in this case, the volume of the network traffic would change so much that the anomalies can be identified by volume-based methods or even by the naked eyes. There is no need to show the experiment results in this case for the sake of brevity.

In the first case, we focus on the impact of the traffic volume of injected anomalies for the LMS-detector. We inject 22 anomalous IP flows and gradually increase their traffic volume. The resulting true positive rate curve and false positive rate curve are shown in Figure 4 and 5. It can be seen that the true positive rate rises sharply at first, and converges to around 80% when the proportion of the anomalous traffic reaches 21% of the total traffic volume on the link. The false positive rate of LMS-based detector rises slowly, and converges to around 12% eventually. These results are not excellent but within acceptable scope.

We also apply the wavelet-based detector to the same synthetic data. The results are shown in Figure 4 and 5 as well. The true positive rate of the wavelet-based detector

keeps around 2%, and shows no sign of growth when the anomalous traffic increases. This low level of true positive rate is unacceptable in practice. In the mean time, the false positive rate of the wavelet-based detector keeps between 2% and 4%, which is smaller than the LMS-based detector. However, this advantage can not make up its bad performance in detection capability.

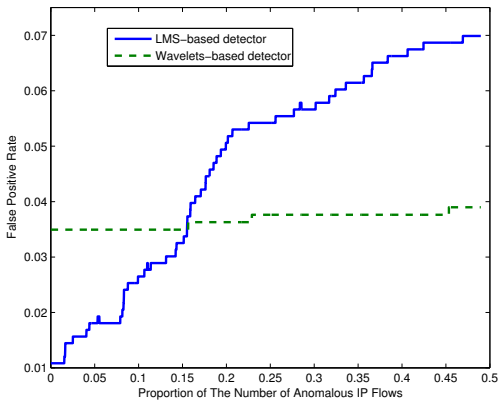

Fig. 7. False Positive Rate for A Large Number of Anomalous IP Flows

In the second case, we focus on the impact of the number of anomalous IP flows. We set the traffic volume of each injected IP flow as 50 bytes. Considering the average traffic volume of each IP flow in sampled CERNET2 data is around 985 bytes, the traffic volume per anomalous IP flow is very small. Then we gradually increase the number of injected IP flows. The results for the LMS-based detector are shown in Figure 6 and 7. The true positive rate rises sharply and converges to 100% eventually, which means the LMS-based detector can detect all the injected anomalies when the proportion of anomalous IP flows reaches 40%. The false positive rate of LMS-based detector keeps below 7%, which is acceptable for practical applications.

As comparison, the results of wavelet-based detector are also shown in Figure 6 and 7. We can see that the wavelet-based detector barely detect any injected anomalies, even its false positive rate is lower than 4%. We figure the reason for the bad performance of the wavelet-based detector as follows. It takes the variance of the network traffic as the only criterion for detection. It ignores the fact that the variance of network traffic is usually proportional to the absolute volume of network traffic, and high traffic volume usually corresponds to massive normal users rather than network anomalies. We leave the validation of this guess for future work.

In conclusion, based on the synthetic CERNET2 data, our LMS-based detector performs well at detecting both anomalies involving a few large IP flows and anomalies involving many small IP flows.

VI. Conclusions

In this paper, we propose a new on-line anomaly detection method. It makes use of two highly correlated metrics—IGTE and IGFE. Its basic idea is to use IGFE values to predict the current IGTE value with the LMS algorithm, and to use the magnitude of the prediction error(error signal) as the criterion for detection. It does not need to wait for the entire data set to be collected before detecting anomalies. Instead, it can output detection results in real time, and is suitable for on-line applications.

Using the artificial synthetic data, it is shown that the LMS-based detector possesses strong detection capability, and its false positive rate is within acceptable scope. We compare the LMS-based detector with the wavelet-based detector, it turns out that the former performs much better than the latter.

Acknowledgments

We are grateful to Lujing Sun for providing us with the Netflow data from CERNET2, and to Kun Wen for many helpful discussions. This work is supported by the National Basic Research Program of China under Grant No. 2012CB315806, the National Natural Science Foundation of China under Grant No. 61170211, 61202356, 61161140454, Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20110002110056, 20130002110058.

References

[1] M. Roesch *et al.*, "Snort: Lightweight intrusion detection for networks." in *LISA*, 1999, pp. 229–238.

[2] F. Silveira, C. Diot, N. Taft, and R. Govindan, "Astute: Detecting a different class of traffic anomalies," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 267–278, 2010.

[3] T. Andrysiak, Ł. Saganowski, and M. Choraś, "Ddos attacks detection by means of greedy algorithms," in *Image Processing and Communications Challenges 4*. Springer, 2013, pp. 303–310.

[4] F. Soldo and A. Metwally, "Traffic anomly detection based on the ip size distribution," in *INFOCOM, 2012 Proceedings IEEE*, 2012, pp. 2005–2013.

[5] B. Zhang, J. Yang, J. Wu, D. Qin, and L. Gao, "Mcst: Anomaly detection using feature stability for packet-level traffic," in *Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific*. IEEE, 2011, pp. 1–8.

[6] B. Zhang, J. Yang, J. Wu, and Z. Wang, "Mbst: detecting packet-level traffic anomalies by feature stability," *The Computer Journal*, vol. 56, no. 10, pp. 1176–1188, 2013.

[7] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*. ACM, 2002, pp. 71–82.

[8] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4. ACM, 2004, pp. 219–230.

[9] S. Haykin, *Adaptive Filter Theory*, T. Kailath, Ed. Tom Robbins, 2001.

[10] P. S. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Springer, 2013.

[11] C. D. Publications, *Optimization theory with applications*. Pierre, Donald A, 2012.

[12] A. Kuzmanovic and E. W. Knightly, "Low-rate tcp-targeted denial of service attacks: the shrew vs. the mice and elephants," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 2003, pp. 75–86.

[13] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.