Anomaly Detection Based on a Multi-class CUSUM Algorithm for WSN

Xiao Zhenghong
School of Information Science and Engineering, Central South University, Changsha (410083), China
Email: huasxzh@126.com

Chen Zhigang
School of Information Science and Engineering, Central South University, Changsha (410083), China
Email: czg@mail.csu.edu.cn

Deng Xiaoheng
School of Information Science and Engineering, Central South University, Changsha (410083), China
Email: advarl@gmail.com

Abstract— Security is one of the most important research issues in wireless sensor networks (WSN) applications. Given that the single detection threshold of the cumulative sum (CUSUM) algorithm causes longer detection delays and a lower detection rate, a multi-class CUSUM algorithm is hereby proposed. Firstly a maximum and minimum thresholds, which sensor nodes are able to reach during sending packet, are set to eliminate abnormal flow to enhance the detection efficiency. Secondly, CUSUM algorithms of different thresholds, all of which are selected according to the mean of traffic sequences, are applied to detect anomalous nodes. This study aims to optimize threshold parameters, the size of which increases with the number of traffic sequence. Using the NS2 tool, the different values of network traffic sequence are generated and simulated. Based on these values, the detection rates of the CUSUM algorithm and multi-class CUSUM algorithms, as well as their false positive rates, are then evaluated. Results show that the proposed algorithm achieves a higher and more accurate rate of detection and lower false positive rates than do the current important intrusion detection schemes of WSN.

Index Terms—WSN, CUSUM algorithm, multi-class CUSUM algorithm, threshold, anomaly detection

I. INTRODUCTION

Wireless Sensor Networks (WSN) have been widely applied in biomedicine, hazardous environment exploration, and military tracking; these are applications requiring higher safety requirements. However, WSN has limited power and has the disadvantages of bandwidth constraint, limited computational capabilities of nodes, a large number of nodes (which are also widely distributed), open communication medium, network dynamics, and enormous sensed data streams[1,2]. These characteristics make WSNs more vulnerable to malicious attacks than other networks. Therefore, developing an intrusion detection system with distributed, lightweight, and rapid detection availability is one of the major challenges of WSN applications.

Currently, in the field of WSN security, more focus has been given on key management, authentication, privacy, security routing protocols and high-level security services, among others.[3,4]. However, these mechanisms may become invalid once the sensor nodes are compromised. Thus, developing an anomaly detection scheme based on the characteristics of WSN is essential.

The remaining part of this paper consists of the following: related works; description of a CUSUM algorithm and the proposal for a multi-class CUSUM algorithm for anomaly detection of WSN; experimental results compared with those of similar research; and conclusions.

II. RELATED WORK

Anomaly detection has been the focus of research on intrusion detection system. In [5], when DDOS attacks occur, that is, when the traffic of the core router ports’ ingress and egress change, a modified CUSUM algorithm is applied to trace their statistics characteristic in real time as well as to detect network traffic abnormality. This method has higher detection speed and accuracy in relation to DDOS attacks. In [6], to make the defense at the source-end more practical, a lightweight detection method is proposed, where network information is extracted and stored in the Bloom filter, after which CUSUM is applied to detect abnormal changes. This method yields the most accurate detection result with less computation cost. In [7], an information-sharing model for distributed intrusion detection systems is presented. In this model, each individual ID in the network uses CUSUM along with its local data and compares it with a local
threshold to check if an anomaly has been detected. One of the advantages of using the CUSUM algorithm is that it can consume less computing resources, making it very suitable for the characteristics of resource-constrained WSN. In [8], a multi-chart for detecting an unknown shift in the mean of an identically distributed process is proposed, where two advantages are validated: on the one hand, it can much reduce computational complexity compared with the GLR (Generalized Likelihood Ratio) and GEWMA (Generalized Exponentially Weighted Moving Average) control charts when the in-control ARL (Average Run Length) is large; on the other hand, it can quickly detect the size of the mean shift. Thereby, the multi-chart can be wholly superior to a single CUSUM in detecting the various mean shifts when the in-control ARL is not large. In [9], an improved non-parameter recursive CUSUM algorithm is used to detect efficiently attack on line. During this procedure, the suspicious packets are also recorded. As for DDoS attacks based on TCP, many unacknowledged segments will be observed in victim ends. In every time period, the ratio of the number of unacknowledged segments and the number of all segments is calculated, then, the time based statistical sequence comes into being. In conclusion, its algorithm is fast and efficient, and it has low false positive rate and could adapt to more complex network environments. In addition, it is helpful to attack analysis and tracing. In [10], an adaptive non-parametric CUSUM control chart algorithm is proposed. Firstly, a fixed threshold is set to eliminate abnormal flow and simplify the detection of obvious anomalies; Secondly, the filtered data is smoothed and transformed based on the simple moving average method and the Chebyshev inequality; Lastly, an adaptive threshold is set according to the transformed results, and the decision-making process will continue monitoring the anomaly for its possible ending after an alarm is raised. In [11], an algorithm is proposed to adjust the values of network traffic and alarm threshold in order to adapt the change of network environments. Whilst, the influence on setting parameters, which consist of the fail alarm and the number of outgoing packets from this node, as well as comparing these values with the predefined thresholds, the attacks can be detected. Nevertheless, these studies present a model without any evaluation on wireless sensor networks.

III. ANOMALY DETECTION SCHEME BASED ON MULTI-CLASS CUSUM ALGORITHM

A. Traffic Pre-Processing

The normal scope of the traffic is decided by the values that sensor node is able to reach during sending packet, in general, these values are limited between the related minimum \(T_{\text{min}}\) and the maximum \(T_{\text{max}}\). When it is out of the scope, the WSN traffic is considered as abnormal data. For the input traffic \(X_n\), if \(X_n > T_{\text{max}}\) or \(X_n < T_{\text{min}}\), these cases are directly eliminated by the system, otherwise, they are handled by the multi-class CUSUM algorithm. The parameter \(A_0\) is set to denote the number of abnormal traffic, and let \(K_0\) denote alarm control parameter. When \(A_0\) increase to \(K_0\), the system thinks the remarkable abnormality happens, so it will trigger an alarm and reset parameters. The traffic pre-processing procedure is shown in Figure 1.

1. input \(X_n\)
2. if \(X_n > T_{\text{max}}\) or \(X_n < T_{\text{min}}\) then
3. \(A_0 = A_0 + 1\)
4. end if
5. if \(A_0 > K_0\) then
6. \(A_0 = 0\)
7. reset parameters
8. else
9. call multi-class CUSUM
10. return
11. end if

Fig 1. Pseudo-Code of the traffic pre-processing

B. CUSUM Algorithm

The CUSUM algorithm with less computation, accurate detection, and non-parametric characteristic is commonly used in anomaly detection. It has been widely used in anomaly detection because it can detect the change of means of a statistical process.

Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed $N(0,1)$ variables, and let $X_{n+1}, X_{n+2}, \ldots, X_{n+3}$ be independent and identically distributed $N(\delta,1)$ variables, where $h$ is the unknown change-point for a given observation sequence $x_1, x_2, \ldots, x_n$. $\Phi(*)$ indicates the distribution density function of standard normal distribution $N(0,1)$. The Likelihood ratio test between $x_1, x_2, \ldots, x_n$ and $x_{n+1}, x_{n+2}, \ldots, x_{n+3}$ is:

$$
L_{n,v} = \frac{\prod_{i=1}^{v} \Phi(x_i) \prod_{j=v+1}^{n} \Phi(x_j - \delta)}{\prod_{i=1}^{n} \Phi(x_i - \delta)} = \frac{\prod_{i=1}^{n} \Phi(x_i - \delta)}{\prod_{i=v+1}^{n} \Phi(x_i)} = \exp \left\{ \delta \sum_{i=1}^{n} \left[ x_i - \frac{\delta}{2} \right] \right\},
$$

(1)

where $\prod_{i=1}^{n} \Phi(x_i) = 1$, $\sum_{i=1}^{n} x_i = 0$.

The logarithm of (1) is:

$$
\Lambda_{n,v} = \ln L_{n,v} = \delta \sum_{i=1}^{n} \left[ x_i - \frac{\delta}{2} \right].
$$

(2)

therefore, the log-likelihood ratio test between the supposition that there is an offset and the original assumption that there is no offset is:

$$
\Lambda_n = \max_{1 \leq v < n} \Lambda_{n,v} = \max \left\{ \delta \sum_{i=1}^{n} x_i - \frac{\delta}{2} \right\}.
$$

(3)

If an upward shift is detected and $\delta > 0$, then the log-likelihood test is equivalent to the following test:

$$
Z_n = \max_{1 \leq v < n} \left\{ \delta \sum_{i=1}^{n} \left[ x_i - \frac{\delta}{2} \right] \right\}.
$$

(4)

The above represents the CUSUM values. Supposing that the observation values of n-1 are not mean offset during use, and $Z_n \leq 0, i=1, 2, \ldots, n-1, \text{ and } \theta$ is the threshold, if $x_1 + x_2 + \ldots + x_k + n\theta > \theta$ is satisfied at time n, then the mean offset is therefore considered to occur.

Let $\overline{x}_i = x_i - \frac{\delta}{2}, \overline{x}_0 = 0$.

Then the formula $Z_n$ can be presented as:

$$
Z_n = \max_{1 \leq v < n} \left\{ \sum_{i=0}^{v} \overline{x}_i - \sum_{i=0}^{v-1} \overline{x}_i \right\} = \overline{S}_n - \min_{0 \leq v < n} \overline{S}_v
$$

$$
Z_n = Z_{n-1} = \left( \overline{S}_n - \overline{S}_{n-1} \right) - \left( \min \overline{S}_v - \min \overline{S}_v \right)_{0 \leq v < n-1}
$$

(5)

(6)

for $\overline{x}_n = \overline{S}_n - \overline{S}_{n-1}$, we have

$$
\min \overline{S}_v = \min \left\{ \overline{S}_n, \min \overline{S}_v \right\}.
$$

(7)

thus,

$$
Z_n - Z_{n-1} = \overline{x}_n - \min \left\{ 0, \overline{S}_n - \min \overline{S}_v \right\} = \max \left\{ \overline{x}_n - \overline{x}_n - \overline{S}_n + \min \overline{S}_v \right\}
$$

$$
= \max \left\{ \overline{x}_n, \min \overline{S}_v - \overline{S}_{n-1} \right\} = \max \left\{ \overline{x}_n - Z_{n-1} \right\}
$$

(8)

(9)

The recursive formula of $Z_n$ is presented as:

$$
Z_n = \max \left\{ 0, Z_{n-1} + x_n - k \right\}, n=1,2,\ldots
$$

(10)

If the alarm threshold is set $0$, and the conditions $Z_n > 0 (Z_n < 0, i=1,2,\ldots,h-1)$ are satisfied at the observation points $h$, then an alarm is reported.

C. Multi-Class CUSUM Algorithm

It can be gleaned from the derivation process previously discussed that the key to using the CUSUM algorithm effectively is that an appropriate $(k, \theta)$ parameter should be chosen. Generally, a rather small $(k, \theta)$ parameter should be selected so as to identify attacks, give an alarm, adopt corresponding measures, and prevent attacks. Sybile, wormholes, sinkhole, selective forwarding, and hello flooding cause the network traffic to deviate from normal conditions. The variations in types and strengths of attacks result in great change in abnormal traffic. Although the CUSUM algorithm possesses the advantages of being light-weight, fast and the capability of accurate detection, there is still a possibility that the single threshold will cause a longer delay for an alarm. This could occur if the same $(k, \theta)$ parameter for different attacks is used when the
strength of attacks is very weak or if the types of attacks switch fast. As a solution to this problem, a multi-class CUSUM algorithm is proposed.

**Definition 1.** Let \( x_{i,1}, x_{i,2}, \ldots, x_{i,m}(i=1,2,\ldots,m) \) be independent and identically distributed \( \mathcal{N}(0,1) \) variables. Let \( x_{i,h+1}, x_{i,h+2}, x_{i,h+3} \) be independent and identically distributed \( \mathcal{N}(\delta,1) \) variables, where \( h \) is the unknown change-point and \( m \) indicates \( m \) different traffic sequences for sensor networks. The mean of flow sequence is:

\[
T_i = \frac{1}{n} \sum_{j=0}^{n} x_{i,j}, i=1,2,\ldots,m
\]  

(11)

**Definition 2.** Suppose that the traffic sequences of \( n-1 \) are not mean offset during use, \( Z_{i-1} \leq \theta, i=2,\ldots,n-1, \) and \( \theta \) is threshold, then mean shift could occur at any of the following situations: if \( x_n + x_{n-1} + \ldots + x_1 - \delta > \theta \) is satisfied at time \( n \), if \( x_{i,n} + x_{i,n-1} - \delta > \theta \) is satisfied, or if \( x_{i,n} + x_{i,n-2} - \delta > \theta \) is satisfied.

From (10) of section 3.1, we extend \( Z_n \) to \( m \) different flow sequences; therefore, the recursive formula of \( Z_i,n \) becomes:

\[
Z_{i,n} = \max \{ 0, Z_{i,n-1} + x_{i,n} - \theta_i \},
\]

(12)

\[
S_{i,n} = \sum_{j=0}^{n} Z_{i,j}, S_{i,0} = 0
\]

(13)

\[
Y_{i,n} = S_{i,n} - \min_{k<i} S_{i,k}
\]

(14)

**Definition 3.** For the traffic sequence \( i \), where \( \theta \) indicates the different thresholds of different flow sequences, if anomaly is not detected in the \( h-1 \) time period, the sufficient and necessary conditions of detected abnormality are:

\[
Y_{i,n} \leq \theta_i, i=1,2,\ldots,m; n=1,2,\ldots,h-1,
\]

(15)

where \( Y_{i,n} = (Y_{i,n} + Z_{i,n})^+, Y_{i,0}=0 \). \( X^+ \) is defined as:

\[
X^+ = \begin{cases} 
  x, & x > 0 \\
  0, & x \leq 0 
\end{cases}
\]

**D. Threshold Optimization**

When the traffic intensity of attacks is weak, improving the detection rate at the beginning of time and optimizing the parameter \( \theta \) becomes difficulties for the CUSUM algorithm. One of the possible solutions for this is that the threshold increases with the number of flow sequence. In [15], a method is presented by improving the parameters \( \theta, h \). This method is better with weak traffic intensity of attack.

**Definition 4.** For the combination of parameters \( \theta, h \), if \( x_{n} - \delta > \theta \) is satisfied, or if \( x_{n} + x_{n-1} + x_{n-2} - \sqrt{3} \delta > \theta \) is satisfied at time \( n \), then an attack occurs.

**Definition 5.** Suppose that the new process inspection begins at the sequences of \( h+1 \), we have:

\[
Z_n = \sum_{i=h+1}^{n} x_i - \sqrt{n-h} \cdot d
\]

(16)

The recursive formula of \( Z_n \) can be expressed as:

\[
Z_n = \max \{ 0, z_{n-1} + x_n - (\sqrt{n-h} - \sqrt{n-h-1})d \}
\]

(17)

Where \( d \) is constant. It can be obtained by a training sample, where \( h \) is defined as:

\[
h = \max \{ i: j < n, Z_i = 0 \}, Z_0 = 0
\]

(18)

### TABLE I. PARAMETERS CONFIGURATION

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network size (m²)</td>
<td>1000 x 1000</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>100</td>
</tr>
<tr>
<td>Data packet size(Byte)</td>
<td>56</td>
</tr>
<tr>
<td>Rate (kbps)</td>
<td>19.2</td>
</tr>
<tr>
<td>Traffic type</td>
<td>cbr traffic flow</td>
</tr>
<tr>
<td>Route protocol</td>
<td>DSR</td>
</tr>
<tr>
<td>MAC protocol</td>
<td>IEEE 802.11</td>
</tr>
</tbody>
</table>

**IV. EXPERIMENTAL RESULTS AND ANALYSIS**

**A. Parameters Configuration**

We used NS2 as a simulation platform. We also used it to analyze and evaluate the multi-class CUSUM algorithm proposed in this paper, through the features of traffic, the CUSUM values, the true positive rate, and the false positive rate of the sensor node. During the experiment, all sensor nodes were randomly deployed in a specific region, and the associated parameters were set (see Table 1).

**B. Analysis of the CUSUM Algorithm**

![Traffic of nodes](Figure2.jpg)

The traffic of WSN under experimental conditions
defined in Table 1 is shown in Figure 2. The CUSUM values of WSN nodes are shown in Figure 3. As can be seen, the simulation time is 256 seconds, the size of traffic is 19.2 (Kbps), the CUSUM values are zero. That is to say, the current system runs normally.

The traffic of nodes attacked is shown in Figure 4. The CUSUM values obtained using the CUSUM algorithms to the flow sequence are shown in Figure 5, in which the window size is set 10 seconds. As can be seen from Figure 4, there exist attacks at 56, 128 and 192 seconds, respectively. Figure 5 shows the mean CUSUM values of 0 before the 56 second mark, i.e. the system is normal. The mean value undergoes significant change when the nodes are attacked after about 8 seconds, and the CUSUM values reach maximum at the moment of attack stop. Henceforth, they gradually decrease until the CUSUM values become zero for the system comes back normal.

The traffic under different parameter $\delta$ is shown in Figure 6, the size of traffic is $(1+\delta)\times 19.2$(Kbps), where $\delta$ values can range from 0.1 to 0.5. The CUSUM values of varying flow sequences with different parameters are shown in Figure 7.

Meanwhile, Figure 7 shows that the mean CUSUM values vary in size under the conditions of different flow sequences. The higher the parameter $\delta$, the greater the value of flow sequences and the CUSUM values. When the CUSUM algorithm is applied to flow sequences, it gives rise to one of the main problems facing the CUSUM algorithm, that is, the CUSUM values greatly vary in size, a single threshold causes a lower detection rate.

The Comparsion Between Multi-class CUSUM Algorithm and CUSUM Algorithm

In order to verify the performance difference between the multi-class CUSUM algorithm and CUSUM algorithm, the related comparsion is conducted by a series of experiments, and the associated results is illustrated in the following figures. The one of the advantages of multi-class CUSUM algorithm is that it is easy to select a threshold, compared with CUSUM algorithm. The
traffic under different attack strengths is shown in Figure 8. As seen in Figure 9, the CUSUM values obtained using CUSUM algorithm to the flow sequence (as above mentioned in Figure 8), vary so as to decide hardly a proper detection threshold. As seen in Figure 10, one of the CUSUM values by using multi-class CUSUM algorithm is single and easy to decide a proper threshold, that is to say, the following results are obtained: detection delays become shorter and a detection rate increases higher. It is obvious that the multi-class CUSUM algorithm is better than the CUSUM algorithm when the WSN traffic is various.

Via the results from Figure 9, we can see that if  \( \gamma / 2 \) is used for a detection threshold, the detection rate is lower for attacked traffic T1 in Figure 8, while detection delays are longer for attacked traffic T3 in Figure 8, however, it is difficult to get a balance between detection rate and detection delay.

By analyzing Figure 10, we select the multi-class CUSUM algorithm according to different flow sequences, as long as \( \eta / 2 \) is used for the value of detection threshold, it is obviously easy to bring the system into the balance between detection rate and detection delay.

**D. Intrusion Detection Analysis**

True Positive Rate (TPR) and False Positive Rate (FPR) are the two main evaluation indicators of the intrusion detection system. The former refers to the ratio of the correct number of attacks detected to the total number of attacks, while the latter refers to the ratio of the number of a normal measurement identified as anomalies to the number of actual normal measurements. To analyze this, we decided to prolong the time of attacks to 100 seconds, which was also the normal working time. The detection rate of the multi-class CUSUM algorithm, the CUSUM algorithm, and the TPDD [13] scheme based on traffic prediction were compared and evaluated under the same experimental conditions. We found that the packet replay rate increased from 10% to 50%. Figures 11 and 12 show the TPR of the three schema and the FPR of three schema, respectively.
From Figure 11, we can see that when the attack strength is weak, the TPR of the multi-class CUSUM algorithm is the best, the TPR of the CUSUM algorithm is better, and the TPR in [13] worst. This is because the CUSUM algorithm requires a longer delay time to achieve the alarm threshold 0 when the attack strength is weak. The intrusion detection scheme proposed in this paper is relatively shorter delay, because the multi-class CUSUM algorithm selects different thresholds for different mean flow sequences, hence improving TPR effectively. However, the TPR of three methods is roughly the same when the attack strength is stronger.

Figure 12 shows that when the attack strength is weak, the FPR in [13] is highest, the CUSUM algorithm higher, and the multi-class CUSUM algorithm the lowest. For the most part, when the attacks stop, the threshold of CUSUM algorithm is larger than that of multi-class CUSUM algorithm. Therefore, the time when traffic mean goes back to 0 for the CUSUM algorithm is longer than that of the multi-class algorithm. Comparatively, the FPR of the CUSUM algorithm is higher, while that of the multi-class algorithm is relatively lower. When the attack strength is stronger, the FPR of the CUSUM algorithm is the lowest, that of the multi-class CUSUM algorithm is higher, while that of [13] is the highest. The reason for this is that the detection delay is shorter, causing a decrease in the number of wrong judgments.

V. CONCLUSION

With the development of wireless sensor networks, the related applications are becoming more wide. These conditions bring the WSN security into more urgent status, and the current security focuses are put on key management, authentication, privacy, security routing protocols, etc. Among the emerging security issues of WSN, these mechanisms may become invalid once the sensor nodes are compromised. Thus, aiming at the characteristic of WSN, the research on an anomaly intrusion detection is essential.

This paper proposes a multi-class CUSUM algorithm based on the single threshold of CUSUM algorithm resulting in longer detection delay and lower detection rate. In this proposed algorithm, a traffic pre-processing mechanism is designed to deal with the remarkable abnormality via the decided threshold scope, then different thresholds are selected according to the mean of traffic sequences. These thresholds are applied to detect anomalous nodes, hence optimizing threshold parameters, the size of which increased with the numbers of traffic sequence. Through the NS2 simulation tool, the WNS traffic of the attacked nodes is simulated. Based on this, a series of performance parameters are thus evaluated, the results of which are compared with the TPDD. Experimental results demonstrate that the scheme based on the multi-class CUSUM algorithm is an ideal intrusion detection method for WSN since it provides higher accuracy rate of detection and lower false positive rate than do existing intrusion detection schemes.

ACKNOWLEDGMENT

This work was partly supported by the National Natural Science Foundation of China under Grant Nos. 60573127 and 60773012, and by the China Postdoctoral Science Foundation under Grant No. 20070420782.

REFERENCES


**XIAO Zheng Hong** was born in 1965. He received his bachelor’s degree from the Nanjing University of Science and Technology in 1986 and received the M. S. from Beijing Institute of Technology in 2001. He is a Ph. D. candidate at the School of Information Science and Engineering, Central South University. His research interests include wireless sensor network, network security and intrusion detection.

**CHEN ZhiGang** was born in 1964. He is a professor and Ph. D. supervisor. His research interests include computer network and distributed computing.

**DENG XiaoHeng** was born in 1974. He is an associate professor, who received his Ph.D. from the Central South University in 2006. His research interests include network modeling and security, network computing, and network processing.