Hindawi Publishing Corporation
Discrete Dynamics in Nature and Society
Volume 2014, Article ID 323764, 10 pages
http://dx.doi.org/10.1155/2014/323764

Research Article

A Large-Scale Network Data Analysis via Sparse and Low Rank Reconstruction

Liang Fu Lu,[1] Zheng-Hai Huang,[1,2] Mohammed A. Ambusaidi,[3] and Kui-Xiang Gou[1]

[1] *Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, China*
[2] *Center for Applied Mathematics of Tianjin University, Tianjin 300072, China*
[3] *Faculty of Engineering and IT, University of Technology, Sydney, NSW 2007, Australia*

Correspondence should be addressed to Kui-Xiang Gou; gkxiang@tju.edu.cn

Received 11 April 2014; Accepted 28 April 2014; Published 26 May 2014

Academic Editor: Xiang Li

With the rapid growth of data communications in size and complexity, the threat of malicious activities and computer crimes has increased accordingly as well. Thus, investigating efficient data processing techniques for network operation and management over large-scale network traffic is highly required. Some mathematical approaches on flow-level traffic data have been proposed due to the importance of analyzing the structure and situation of the network. Different from the state-of-the-art studies, we first propose a new decomposition model based on accelerated proximal gradient method for packet-level traffic data. In addition, we present the iterative scheme of the algorithm for network anomaly detection problem, which is termed as NAD-APG. Based on the approach, we carry out the intrusion detection for packet-level network traffic data no matter whether it is polluted by noise or not. Finally, we design a prototype system for network anomalies detection such as Probe and R2L attacks. The experiments have shown that our approach is effective in revealing the patterns of network traffic data and detecting attacks from large-scale network traffic. Moreover, the experiments have demonstrated the robustness of the algorithm as well even when the network traffic is polluted by the large volume anomalies and noise.

1. Introduction

The rapid growth of data communication through the Internet and World Wide Web has led to vast amounts of information available online. In addition, business and government organizations create large amounts of both structured and unstructured information which need to be processed, analyzed, and linked. The large-scale network data plays a popular and important role in network operation and management. Consequently, high-dimensional data and multivariate data are becoming commonplace as the number of applications increases, such as statistical and demographic computation and digital libraries. Though it can provide flexible and cost-saving IT solutions for the end users, it is much easier in causing a great deal of problems such as network and system security issues due to its sharing and centralizing computing resources.

In general, network managers consider that the packet-level and flow-level data constitute the traditional network traffic data. On one hand, the packet-level data analysis performs successfully in maintaining simple, scalable, highly available, and robust networks [1]. On the other hand, flow-level data analysis has become popular in recent studies because the data can describe the network-level status and behavior of communication networks from origin nodes to destination nodes (OD) effectively [2]. However, there is no absolute way to secure the data and data transformations in large-scale networking systems. The existing techniques and tools of securing a network system still rely heavily on human experiences. Most of them require human involvement in analyzing and detecting anomalies and intrusions. Moreover, the existing networked-data analysis techniques are mainly based on the complete data, which limits the application of them. Unveiling the anomalies is a crucial task, especially nowadays, as big data acquisition and storage become increasingly difficult with the increasing amount of data due to the sampling bandwidth and storage space constraints [3]. Some approaches have been achieved in flow-level network data [2, 4, 5]. They reconstructed all origin-destination flows via compressive sensing methods by leveraging the low intrinsic-dimensionality of OD flows and the sparse nature of anomalies. Meanwhile, due to the difficulties of collecting and processing the large-scale packet-level data, to the best of our knowledge, few researchers pay attention to analyzing the incomplete packet-level data for managing and controlling the whole network.

Intrusion detection systems are security managements systems developed to find inconsistency with expected patterns in network traffic data, which is termed as well in literature [3] as novelty detection, anomaly mining, and noising mining. They play an important role in detecting different types of network attacks including Denial of Service (DOS), surveillance and other probing (Probe), unauthorized access to local super user (root) privilege (U2R), and unauthorized access from a remote machine (R2L) attacks. Intrusion detection approaches can be categorized into two main categories: signature-based and anomaly-based detection. Signature-based or misuse-based detection systems detect on-going anomalies by looking for a match with any predefined attack signature [6]. Anomaly-based detection, on the other hand, makes an assumption that intruders' behaviors are different from that of normal network traffic. Therefore, any deviation from the normal flow can be considered as an attack [7].

To enhance the human perception and understanding of different types of network intrusions and attacks, and inspired by the literature [5], approaches on network traffic data analysis and network anomalies detection based on compressed sensing in big data are put forth in this paper. As pointed out in literature [2, 3], on one hand, the number of normal data instances is much more than the number of anomaly data ones, which exactly meets the sparsity requirements of compressed sensing theory. On the other hand, traffic matrices usually have low effective dimensions because they can be well approximated by a few principal components that correspond to the largest singular values of the matrices, which are introduced by Lakhina et al. [8] by using of Principal Component Analysis (PCA) method to traffic matrix analysis.

Therefore, at the first stage, we propose a new decomposition model for packet-level traffic matrix. Then, we present the iterative algorithm based on accelerated proximal gradient method for network anomaly detection problem, which is termed as NAD-APG. Based on the approach, we carry out the intrusion detection for network traffic data no matter whether it is polluted by noise or not. Finally, we design a prototype system for network anomalies detection such as Probe and R2L attacks and so on. The experiments have shown that our approach is effective in revealing the patterns of network traffic data and detecting attacks from large-scale network traffic. In addition, the experiments have demonstrated the robustness of the algorithm as well even when the network traffic is polluted by the large volume anomalies and noise.

The rest of the paper is organized as follows. Section 2 gives an overview of existing methods on structural analysis of network traffic via compressed sensing techniques. Section 3 presents our approach on anomaly detection in network traffic based on accelerated proximal gradient line research method (APGL). The experimental evaluation of our new approaches is explored in Section 4. Finally, conclusions and future work are presented in Section 5.

2. Related Work

It has become popular in recent studies that considering the traffic matrix analysis as the main flow-level data because the traffic matrix can describe the network-level status and behavior of communication networks from origin nodes to destination nodes (OD) effectively and it is a combination of different classes of network traffic to represent how much data is transmitted during different time intervals [8].

As one of the most widely used methods to analyze traffic matrix, PCA was put forth in [9] by Lakhina et al. They calculated the principal component that corresponds to the largest singular value of the matrix and utilized these principal components to approximate the original traffic matrix. Moreover, they improved this method and proposed volume anomaly detection approach based on PCA-subspace [8]. In the following approaches, researchers improved the classical PCA method and proposed distributed PCA [10], network anomography [11], and traffic matrix evaluation from adaptivity and bias perspectives [12]. However, as mentioned by the literatures [5, 8, 13], there are some limitations when we utilize the PCA method to deal with the traffic matrix in order to analyze and manage the whole network, such as the fluctuation of estimation error with the volume change of anomalies, the sensitivity to the choice of parameters, and failure on exploiting the sparsity of anomalies.

Therefore, due to the increasing complexity and amounts of internet applications, the acquisition and storage of big data becomes more and more difficult. Moreover, to overcome the limitations of PCA, researchers have obtained some approaches for analyzing end-to-end network traffic in recent years. To solve the problem that PCA performs poorly in polluted traffic matrix by large volume anomalies, Lakhina et al. [8] proposed structural analysis by decomposing the network traffic matrix into deterministic traffic, anomaly traffic, and the noise traffic matrix. They analyzed that the decomposition problem is equivalent to the relaxed principal component pursuit method. A distributed estimation method to unveil the anomalies presented in OD flows using proximal gradient method was proposed by Mardani et al. [5]. A centralized solver and the in-network processing of link-load measurements were analyzed in their work as well. While Nie et al. found that the size of OD flows obeys the power laws [4]. By using this characteristic and restricted isometric property in compressed sensing theory, they reconstructed all OD flows with the help of partial observed samples from backbone network traffic data.

However, all the current researches pay too much attention on the network traffic in flow-level network. Therefore, we propose a novel approach based on the latest method from compressed sensing to reveal the abnormal patterns by dealing with the packet-level network data. Firstly, we propose to apply the most popular accelerated proximal gradient line search method (APGL) [14] to recover the low rank matrices with network traffic data. Moreover, to get a more accurate and robust approximation to reconstruct traffic matrix, motivated partly by the literature [15], we propose a traffic matrix decomposition method based on the APGL algorithm. Finally, the simulation results and analysis describe the effectiveness and robustness of our approaches in network traffic data.

3. Overall Approach

3.1. Principal Component Analysis Method. Principal component analysis (PCA), as a widely used method in high dimensional data analysis, can be viewed as a coordinate transformation process which transforms the redundant data points to a low dimensional system. As pointed out in literature [2, 9], each row vector x_i of the traffic matrix $X \in \mathbb{R}^{m\times n}$ is considered as a data point. PCA is performed by calculating the principal component vectors of X, which are represented as $\{v_i, i = 1, 2, \ldots, n\}$. The first principal component vector enjoys the property of the maximum variance of the original matrix X. Similarly, the tth principal component vector v_t, $t = 2, 3, \ldots, m$ captures the maximum variance of the residual traffic matrix as follows:

$$v_t = \arg\max_{\|v\|=1} \left\| \left(X - \sum_{i=1}^{t-1} X v_i {v_i}^T \right) v \right\|. \tag{1}$$

Corresponding to v_i, we denote another unit vector u_i as

$$u_i = \frac{X v_i}{\|X v_i\|}, \quad i = 1, 2, 3, \ldots, m. \tag{2}$$

It is noted that vectors u_i and v_i ($i = 1, 2, 3, \ldots, m$) form the orthogonal basis of $\mathbb{R}^n$, respectively. Therefore, the traffic matrix can be decomposed by the following formula:

$$X = \sum_{i=1}^{m} \|X v_i\| u_i {v_i}^T. \tag{3}$$

If we denote $\sigma_i := \|X v_i\|$, the traditional PCA method can be recited by the famous singular value decomposition (SVD) method in the research field of matrix computation as follows:

$$X = \sum_{i=1}^{m} \sigma_i u_i {v_i}^T, \tag{4}$$

where v_i can be achieved by calculating the eigenvectors of the matrix $X^T X$, while $\sigma_i = \sqrt{\lambda_i} \cdot \lambda_i$ is the corresponding eigenvalue of the matrix $X^T X$. That is to say,

$$X^T X v_i = \lambda_i v_i := {\sigma_i}^2 v_i. \tag{5}$$

SVD plays an important role for its revealing, interesting, and attractive algebraic properties and conveys important geometrical and theoretical in-sights about transformations. The entries of each matrix obtained by the SVD algorithm have their special physical significances. According to the rationale of Eckart-Young theorem, $\sum_{i=1}^{r} \sigma_i u_i {v_i}^T$ ($1 \le r \le m$) is considered to be the best rank-r approximation of X, that is, $\sum_{i=1}^{r} \sigma_i u_i {v_i}^T = \arg\min_{\operatorname{rank}(Y)\le r} \|X - Y\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm.

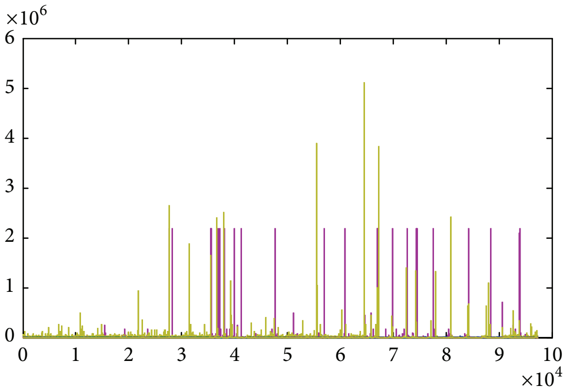


Figure 1: Visualization for original normal traffic with 97278 data items.

To the matrix expression for the PCA, it seeks an optimal estimate of A via the following constrained optimization:

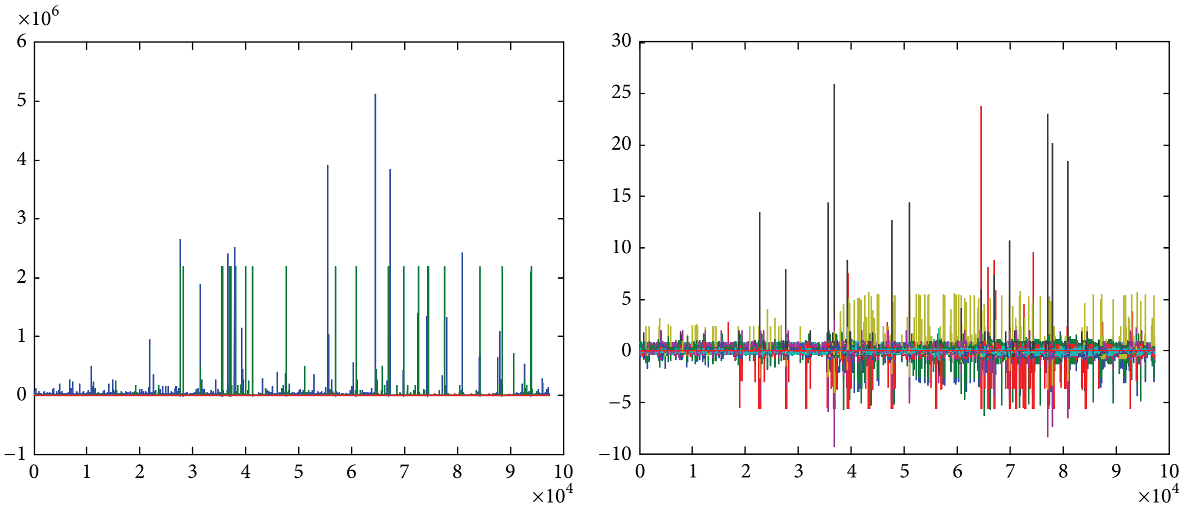
$$\begin{aligned} \min_{A,E} \quad & \|E\|_F, \\ \text{subject to} \quad & \operatorname{rank}(A) \le r, \\ & X = A + E, \end{aligned} \tag{6}$$

where $A, E \in \mathbb{R}^{m\times n}$, $r \ll \min(m, n)$. In fact, the optimal estimate of A is the projection of the columns of X onto the subspace spanned by the r principal left singular vectors of X [16].

3.2. Network Anomaly Detection Algorithm Based on APG. Though classical PCA method processes the data with the corruption of small Gaussian noise effectively, it always breaks down under large corruption [16]. Therefore, to recover a low-rank matrix A from a corrupted data matrix $X = A + E$, where some of the matrix E may be of arbitrarily large magnitude, Wright et al. [17] proposed a method termed as Robust PCA (RPCA) which can exactly recover the low-rank matrix in the presence of gross errors. Based on their analysis for the above optimization problem, the Lagrangian reformulation of it is

$$\begin{aligned} \min_{A,E} \quad & \operatorname{rank}(A) + \lambda \|E\|_0, \\ \text{subject to} \quad & X = A + E, \end{aligned} \tag{7}$$

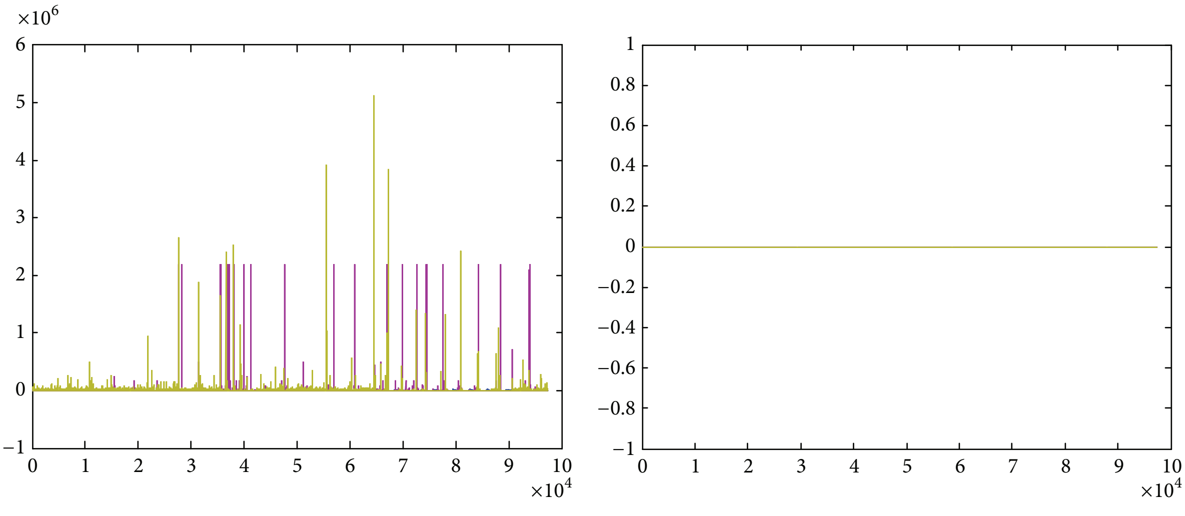
where λ is a positive parameter that balances the two terms. Unfortunately, the above optimization problem is NP-hard in general due to the nonconvexity and discontinuous nature of the rank function. Moreover, the nuclear norm $\|\cdot\|_*$ (the sum of singular values of a matrix) is well known as a convex surrogate of the nonconvex matrix rank function. Therefore, in the literature [17], researchers proposed to solve



(a)

(b)

FIGURE 2: (a) Principal components of original data in PCA. (b) Residual matrix visualization in PCA.



(a)

(b)

FIGURE 3: (a) Normal traffic in NAD-APG method. (b) Visualization for abnormal traffic using NAD-APG.

the following convex optimization problem by replacing the l^0-norm with l^1-norm and $\operatorname{rank}(A)$ with $\|A\|_*$:

$$\begin{aligned} \min_{A,E} \quad & \|A\|_* + \lambda\|E\|_1, \\ \text{subject to} \quad & X = A + E. \end{aligned} \tag{8}$$

To develop faster and more scalable algorithms associated with the robust PCA, one popular method among state-of-art research approaches [16, 18–21] dubbed accelerated proximal gradient algorithm (APG) is widely exploited to seek an optimal solution of a soft constrained version of the convex problem (8). In this paper, we adopt the algorithm partially in the literature [16]. The main model is the following unstrained minimization optimization problem:

$$\min_{A,E} \mu\|A\|_* + \lambda\mu\|E\|_1 + \frac{1}{2}\|X - A - E\|_F^2. \tag{9}$$

Moreover, they summarized the convergence of the algorithm theoretically as follows.

Theorem 1 (see [16]). *Suppose that $F(A, E) = \mu\|A\|_* + \lambda\mu\|E\|_1 + (1/2)\|X - A - E\|_F^2$. For all $k > \log(\mu_0/\mu)/\log(1/\eta)$, any solution X^* of the problem* (9), *we have $F(X) - F(X^*) \leq 4\|X_{k_0} - X^*\|_F^2/(k - k_0 + 1)^2$.*

The APG algorithm solves the optimization problem (9) by iteratively updating A, E, and other parameters. At last, in

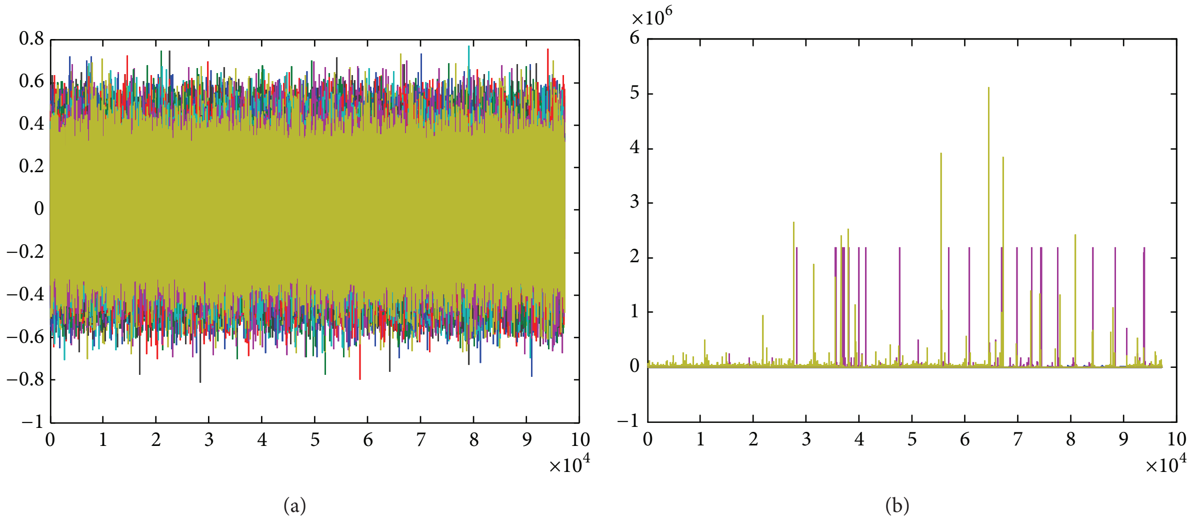


Figure 4: (a) Gaussian white noise matrix. (b) Anomaly-free traffic matrix with Gaussian white noise.

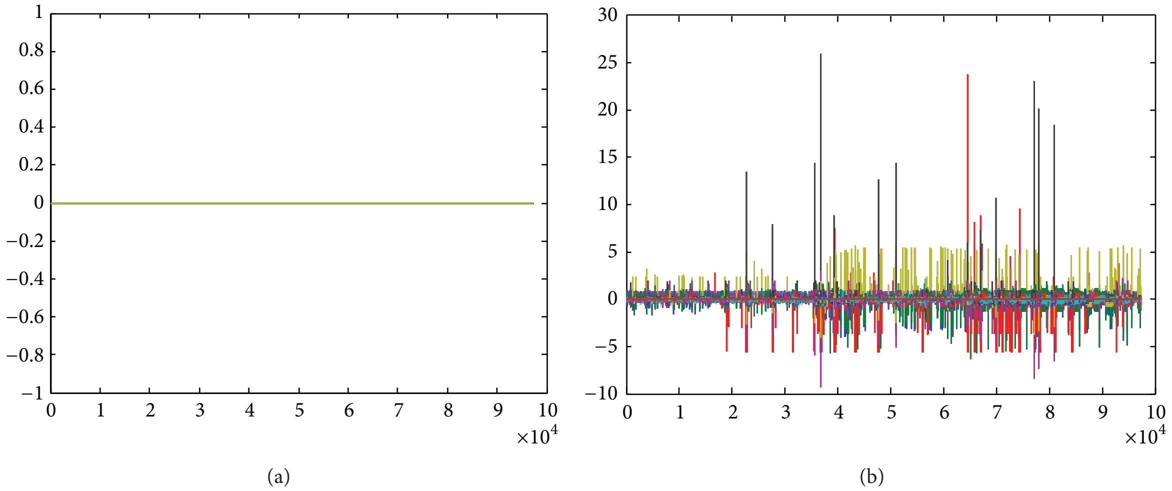


Figure 5: (a) Abnormal traffic using NAD-APG. (b) Residual matrix with Gaussian white noise in PCA.

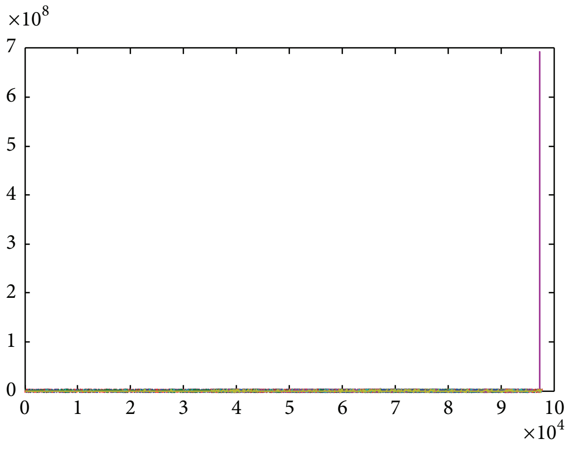
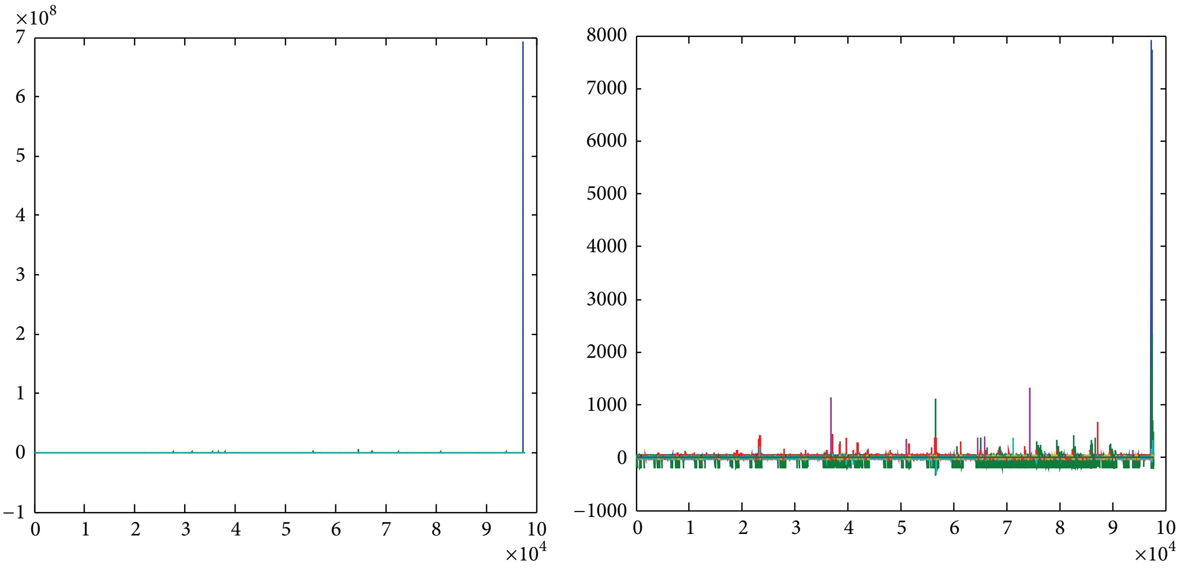


Figure 6: Original normal traffic mixed with 500 Probe attack data.

the kth iteration, we update A_{k+1} and E_{k+1} as the following iterative scheme:

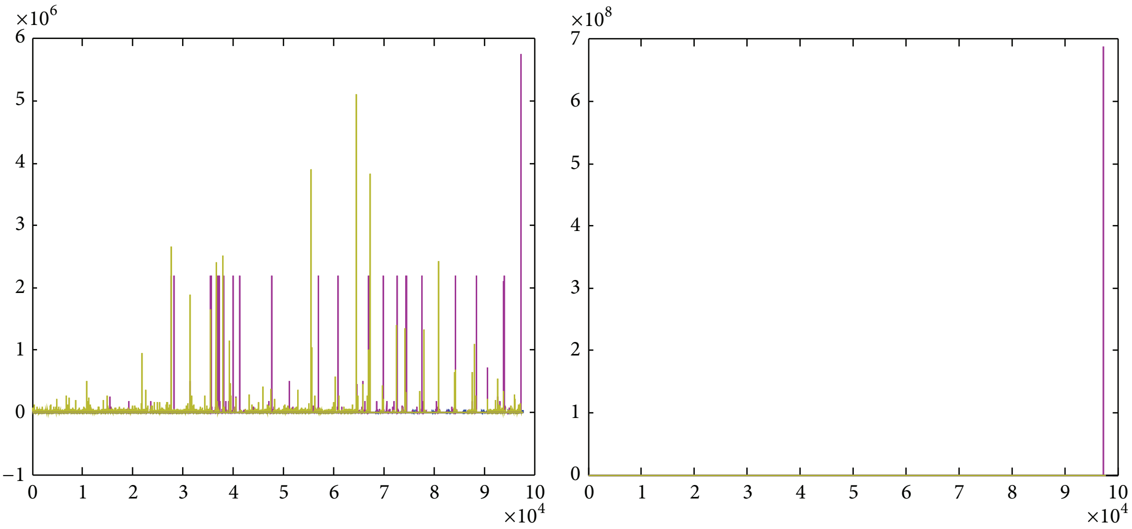
$$A_{k+1} = \mathbb{S}_{\mu/2}\left(Y_k^A + \frac{X - Y_k^A - Y_k^E}{2}\right),$$
$$E_{k+1} = \mathbb{S}_{\mu/2}\left(Y_k^E + \frac{X - Y_k^A - Y_k^E}{2}\right),$$
$$\text{where } \mathbb{S}_{\mu/2}(x) = \begin{cases} x - \dfrac{\mu}{2}, & \text{if } x > \dfrac{\mu}{2}; \\ x + \dfrac{\mu}{2}, & \text{if } x < \dfrac{\mu}{2}; \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$



(a)

(b)

Figure 7: (a) Principal components of hybrid data in PCA. (b) Residual matrix visualization in PCA.



(a)

(b)

Figure 8: (a) Normal traffic in NAD-APG method. (b) Abnormal patterns shown in NAD-APG.

Y_{k+1}^{A}, Y_{k+1}^{E} and t_{k+1} are updated in the same way as [16] as follows:

$$
\begin{aligned}
Y_{k+1}^{A} &= A_{k+1} + \frac{t_k - 1}{t_{k+1}} \left(A_{k+1} - A_k\right), \\
Y_{k+1}^{E} &= E_{k+1} + \frac{t_k - 1}{t_{k+1}} \left(E_{k+1} - E_k\right), \\
t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}.
\end{aligned}
\tag{11}
$$

Here we summarize the main procedure for solving our network anomaly detection problem by APG algorithm, which is called NAD-APG (see Algorithm 1). As pointed out in [19], the algorithm has a convergence rate of $O(1/k^2)$.

4. Experiments and Results

In this section, we conduct several experiments on different attack types to show the effectiveness of our proposed approach.

4.1. The Data Set. Currently, there are only few public datasets for intrusion detection evaluation. According to the literature review by Tsai et al. [22], the majority of the IDS experiments

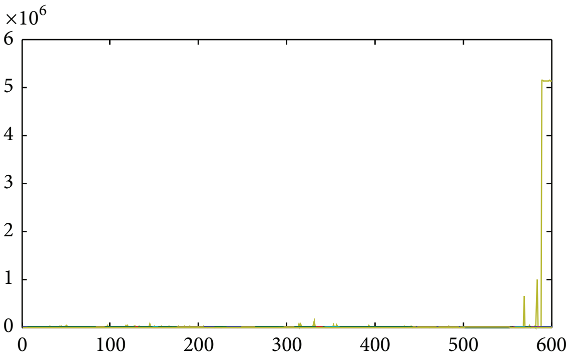


Figure 9: Visualization for 500 normal and 100 R2L network traffic.

are performed on the KDD Cup 99 datasets. It is the most comprehensive dataset that is still widely applied to compare and measure the performance of IDSs. Therefore, in order to facilitate fair and rational comparisons with other state-of-the-art detection approaches, we select the KDD Cup 99 dataset to evaluate the performance of our approach for detection. This dataset was derived from the DARPA 1998 datset. It contains training data with approximately five million connection records and test data with about two million connection records. Each record in this dataset is unique with 41 features. KDD Cup 99 dataset includes normal traffic and our different types of attacks, namely, Probe, Denial of Service (DOS), User o Root (U2R), and Remote to User (R2U). More details about these attacks are given by [23].

During our experiments, we use 10% KDD Cup 99 for training and testing. Literature review shows that a significant number of state-of-the-art IDSs, such as [23, 24], were evaluated using 10% KDD Cup 99 data. Therefore, training and testing our system on the 10% KDD Cup 99 data can help to provide a fair comparison with those approaches. The 10% KDD Cup 99 consists of 494,021 TCP/IP connection records simulated in a military network environment, US Air Force LAN. Each record is labeled as either normal or an attack, and it has 41 different quantitative and qualitative features. These features are generally categorized into three main groups. The first group is the basic features (i.e., attributes 1 to 9) that can be extracted from a TCP/IP connection. The second group refers to features 10 to 22 that are named as content-based features presenting the information derived from network packet payloads. The third group corresponds to the traffic-based features, which are carried by the features 23 to 41 of each record. A complete list of the set of features and the detailed description is available in [25].

4.2. Experimental Results. To demonstrate our method for managing the internet network, especially detecting the abnormal behaviors from the normal network traffic, we conduct PCA and NAD-APG methods for normal traffic mixed with some attacks in our experiments.

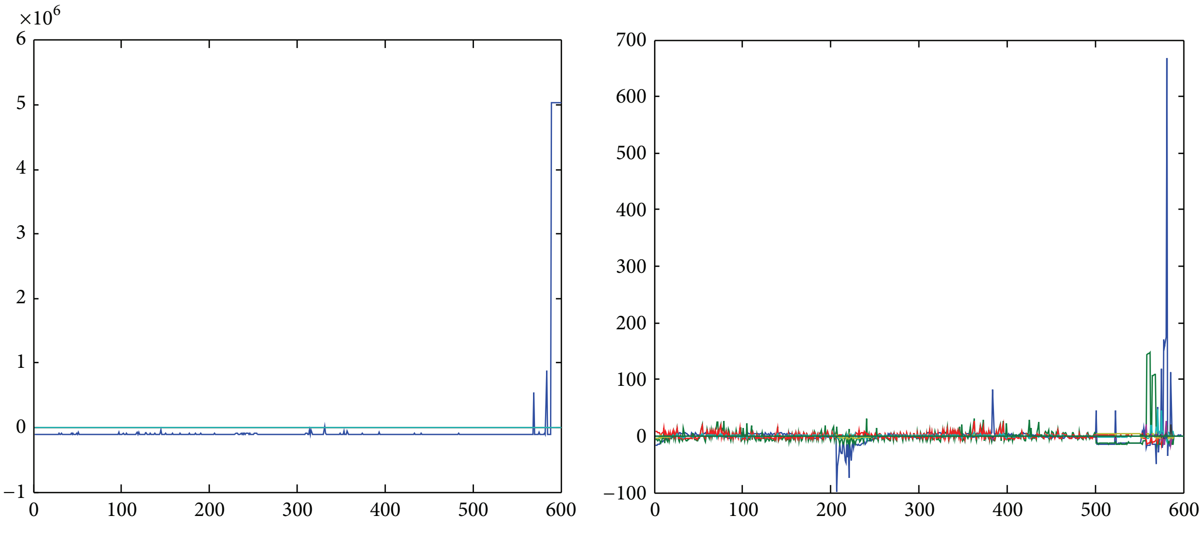
Firstly, we randomly choose some normal traffic data which consists of 97278 data items (see Figure 1). In this paper, we consider these data as the normally-free traffic to test the performance of our method. After using PCA, we can find that the principal components almost enjoy the whole property of the original normal traffic. Here, the sum of the variances of the first ten principal components is near 100% of the total variance of the original data. Figure 2 shows us the details of two matrices after decomposing the original matrix using PCA, where Figure 2(a) shows the visualization for matrix which is composed of top ten principal components and Figure 2(b) shows the details of residual matrix. However, it is very difficult for us to find any pattern from them. Moreover, there are still some data characteristics left in residual matrix after we apply PCA to the original normal data, which confuse the network analysts greatly. If we use our proposed method to the above normal traffic, two matrices with low-rank and sparse properties can be obtained to show the normal and abnormal traffic, respectively. Figure 3 visualizes the details of the two matrices, where Figure 3(a) represents the low-rank matrix with main properties of the original normal traffic. In our paper, we term this matrix as normal traffic matrix. While Figure 3(b) displays the matrix with all zero elements, which is termed as abnormal traffic matrix. Therefore, if we decompose the unique normal traffic into two matrices, we can obtain normal matrix uniquely. To sum up, Figure 3 shows the correctness and effectiveness of our proposed NAD-APG. Furthermore, if the traffic was polluted by noise (in this paper, we refer the noise to be Gaussian white noise with 0 mean and 1 variance), NAD-APG can identify the anomaly-free traffic accurately.

Figures 4 and 5 show the robust property of our proposed NAD-APG method, where (a) is the visualization of Gaussian white noise added to the normal traffic and (b) is the recovery of anomaly-free traffic. In fact, we obtained the all zero matrix as well in this decomposition, which means the traffic is anomaly-free. Figure 5 compares the effectiveness of NAD-APG method with PCA. In the visualization of residual matrix of PCA, it is obvious that we cannot find any pattern of attacks.

To evaluate the effectiveness of our method for detecting attacks in the whole internet, we add 500 "Probe" data items to the normal traffic, which means that the amount of the total data items is 97778 (see Figure 6). As we all know, Probe attacks refer to attackers that typically probe the victim's network or host by searching through the network or host for open ports before they launch an attack on a given host [26]. Therefore, there may be a large volume of traffic in a short time interval. Figure 6 displays the details of hybrid traffic. Firstly, we try to use PCA to detect the Probe attack. Figure 7(a) shows us the principal components of the hybrid traffic data which occupy the 99% contribution to the whole data. Though only 1% of energy of the whole data is left in the residual matrix, we find that there are still some intrinsic properties of the original data set which does not show any valuable pattern for attacks. However, if we test the hybrid data using NAD-APG method, the sparse traffic matrix obtained from the algorithm shows the attack pattern apparently. This enables the network manager to identify the anomalous packets from the normal traffic and can improve the accuracy of attacks detection. Figures 8(a) and 8(b) show the patterns of normal and abnormal traffic decomposed by the NAD-APG scheme.

```
Input: Network traffic matrix $X \in \mathbb{R}^{m\times n}$, $\lambda$ and tolerance $\varepsilon$.
Initialize: $A_0 \leftarrow 0; E_0 \leftarrow 0; t_0 \leftarrow 1$;
Repeat
  Step 1. Update $A_{k+1}, E_{k+1}$ as $A_{k+1} = \mathbb{S}_{\mu/2}(Y_k^A + ((X - Y_k^A - Y_k^E)/2))$ and $E_{k+1} = \mathbb{S}_{\mu/2}(Y_k^E + ((X - Y_k^A - Y_k^E)/2))$.
  Step 2. Let $t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$.
  Step 3. Update $Y_{k+1}^A$ and $Y_{k+1}^E$: $Y_{k+1}^A = A_{k+1} + ((t_k - 1)/(t_{k+1}))(A_{k+1} - A_k)$, $Y_{k+1}^E = E_{k+1} + ((t_k - 1)\,/\,(t_{k+1}))(E_{k+1} - E_k)$
Until $\|A_{k+1} - A_k\| \le \varepsilon$; $\|E_{k+1} - E_k\| \le \varepsilon$; $A \leftarrow A_{k+1}, E \leftarrow E_{k+1}$.
Analyze: normal traffic pattern matrix $A$; abnormal traffic matrix $E$.
Output: Is there any abnormal activity or not in the whole traffic?
```

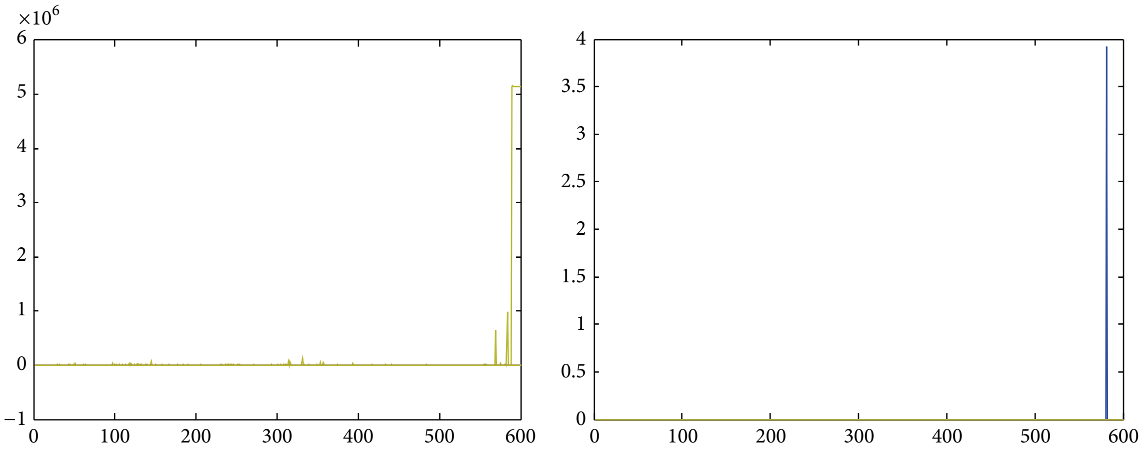
Algorithm 1: Anomaly detection using APG.



(a)

(b)

Figure 10: (a) Principal components of NR data in PCA. (b) Residual matrix of NR data in PCA.



(a)

(b)

Figure 11: (a) Normal traffic of NR in NAD-APG method. (b) Abnormal patterns of NR in NAD-APG.

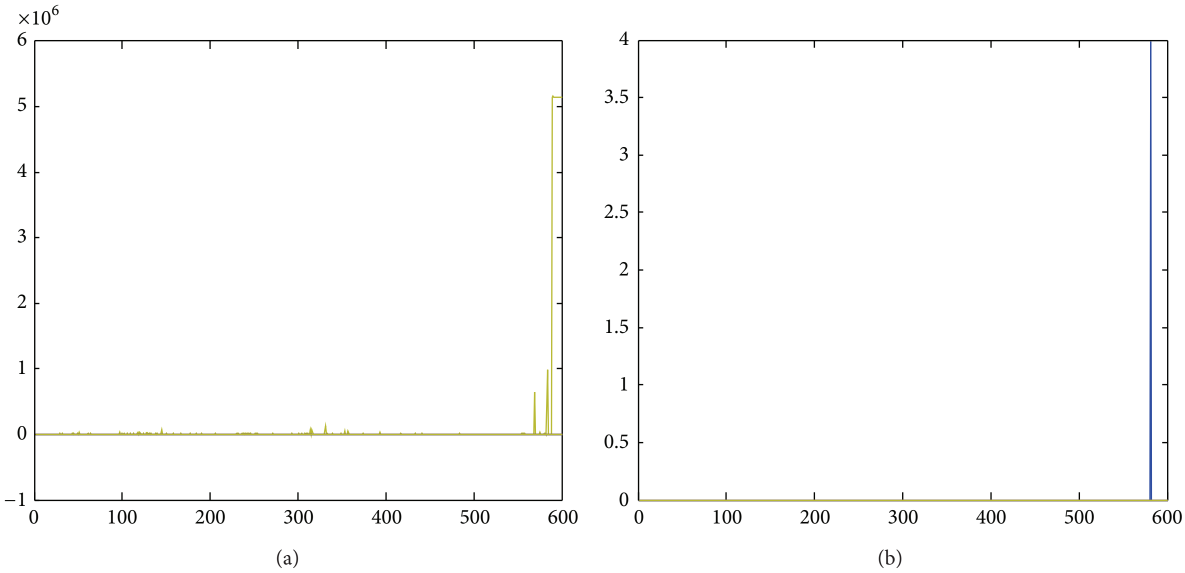


Figure 12: (a) Normal traffic of NRN in NAD-APG method. (b) Abnormal patterns of NRN in NAD-APG.

Table 1: Rank of low-rank matrix in the decomposition of NAD-APG method for attacks detection.

Different sampling data	Data items	Rank (sampling data)	Rank (low-rank matrix)
Normal traffic	97378	38	10
Normal traffic + Probe	97778	38	4
500 Normal traffic + 100 R2L traffic	600	36	4
Normal traffic + 100 R2L traffic	97478	38	4

To further demonstrate the advantages of our proposed scheme in small sampling data, we randomly choose 500 normal data items and 100 R2L attacks traffic as our test data set (here we term this data set as NR). R2L attacks always reveal the unauthorized local access from a remote machine. Moreover, the data values in R2L attacks are always much larger than the common data traffic. Therefore, we can find that the fluctuation in the scale of the data values occurs from Figure 9. Figures 10 and 11 represent the details of the normal and abnormal traffic matrices obtained from the two different decomposition algorithms, where the residual matrix in PCA is still confusing. Therefore, it is very difficult for network analysts to find the R2L attacks pattern from it. On the contrary, the attack pattern can apparently be found from the sparse matrix as shown in Figure 11(b). Even sometimes the real network traffic data are polluted by the noise (here we term this data set as NRN), especially the Gaussian white noise; the NAD-APG method can separate the normal and R2L attack patterns apparently. Figures 12(a) and 12(b) reveal the different patterns hidden in the network traffic, respectively.

To sum up, the experiments implemented above show that the low-rank matrix represents the normal traffic and the sparse matrix can always reveal the patterns of different attacks when we use our proposed NAD-APG scheme to detect anomalies. Table 1 describes the different ranks of the normal traffic matrix in detecting different attacks. There is no doubt that the low-rank matrices in processing different sampling data have much lower ranks than the original ones. However, the rank of low-rank matrix is ten when we deal with the pure normal network traffic, which may be caused by the pure type of data.

5. Conclusion

This paper introduced a new decomposition model for packet-level network traffic data no matter whether it was polluted by large-scale anomalies and noise or not. We presented the iterative algorithm based on accelerated proximal gradient method, which was termed as NAD-APG. Moreover, we designed a prototype system for network anomalies detection such as Probe and R2L attacks and so on. The experiments have shown that our approach is effective in revealing the patterns of network traffic data and detecting attacks from a variety of networking patterns. In addition, the experiments have demonstrated the robustness of the algorithm as well when the network traffic is polluted by the large volume anomalies and noise.

Though it is effective in detecting the attacks from the large-volume network traffic, it is difficult to classify the abnormal activities. Therefore, leveraging some feature selection and classification methods to our approaches to enhance the efficiency of intrusion detection is considered as our near future work. On the other hand, we will do more researches on APG algorithm itself and make our method more powerful and practical.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (no. 11171252 and no. 61202379) and the Doctoral Fund of Ministry of Education of China (no. 20120032120041).

References

[1] C. Fraleigh, S. Moon, B. Lyles et al., "Packet-level traffic measurements from the sprint IP backbone," *IEEE Network*, vol. 17, no. 6, pp. 6–16, 2003.

[2] Z. Wang, K. Hu, K. Xu, B. Yin, and X. Dong, "Structural analysis of network traffic matrix via relaxed principal component pursuit," *Computer Networks*, vol. 56, no. 7, pp. 2049–2067, 2012.

[3] W. Wang, D. Lu, X. Zhou, B. Zhang, and J. Mu, "Statistical wavelet-based anomaly detection in big data with compressive sensing," *EURASIP Journal on Wireless Communications & Networking*, vol. 2013, no. 1, article 269, pp. 1–6, 2013.

[4] L. Nie, D. Jiang, and L. Guo, "A power laws-based reconstruction approach to end-to-end network traffic," *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 898–907, 2012.

[5] M. Mardani, G. Mateos, and G. B. Giannakis, "Unveiling anomalies in large-scale networks via sparsity and low rank," in *Proceedings of the Conference Record of the 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR '11)*, pp. 403–407, November 2011.

[6] G. Vigna and R. A. Kemmerer, "NetSTAT: a network-based intrusion detection approach," in *Proceedings of the 14th Annual Computer Security Applications Conference*, pp. 25–34, 1998.

[7] A. Hassanzadeh and B. Sadeghian, "Intrusion detection with data correlation relation graph," in *Proceedings of the 3rd International Conference on Availability, Security, and Reliability (ARES '08)*, pp. 982–989, March 2008.

[8] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," *ACM SIGMETRICS Performance Evaluation Review*, vol. 32, no. 1, pp. 61–72, 2004.

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

[10] L. Huang, M. I. Jordan, A. Joseph, M. Garofalakis, and N. Taft, "In-network PCA and anomaly detection," in *Advances in Neural Information Processing Systems*, pp. 617–624, 2006.

[11] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan, "Network anomography," in *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement (IMC '05)*, p. 30, 2005.

[12] A. Soule, A. Lakhina, N. Taft et al., "Traffic matrices: balancing measurements, inference and modeling," *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, pp. 362–373, 2005.

[13] H. Ringberg, A. Soule, J. Rexford, and C. Diot, "Sensitivity of PCA for traffic anomaly detection," *ACM SIGMETRICS Performance Evaluation Review*, vol. 35, no. 1, pp. 109–120, 2007.

[14] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 3, pp. 615–640, 2010.

[15] H. Yao, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2117–2130, 2013.

[16] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast algorithms for recovering a corrupted low-rank matrix," in *Proceedings of the 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP '09)*, pp. 213–216, December 2009.

[17] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *Proceedings of the Neural Information Processing Systems*, 2009.

[18] R. He, W. S. Zheng, T. Tan, and Z. Sun, "Half-quadratic based iterative minimization for robust sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 261–275, 2014.

[19] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[20] Y.-F. Li, Y.-J. Zhang, and Z.-H. Huang, "A reweighted nuclear norm minimization algorithm for low rank matrix recovery," *Journal of Computational and Applied Mathematics*, vol. 263, pp. 338–350, 2014.

[21] M. Zhang, Z.-H. Huang, and Y. Zhang, "Restricted p-isometry properties of nonconvex matrix recovery," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4316–4323, 2013.

[22] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: a review," *Expert Systems with Applications*, vol. 36, no. 10, pp. 11994–12000, 2009.

[23] S. Mukkamala, A. H. Sung, and A. Abraham, "Intrusion detection using an ensemble of intelligent paradigms," *Journal of Network and Computer Applications*, vol. 28, no. 2, pp. 167–182, 2005.

[24] A. Mitrokotsa, M. Tsagkaris, and C. Douligeris, "Intrusion detection in mobile Ad Hoc Networks using classification algorithms," *IFIP International Federation for Information Processing*, vol. 265, pp. 133–144, 2008.

[25] F. Amiri, M. Rezaei Yousefi, C. Lucas, A. Shakery, and N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1184–1199, 2011.

[26] K. Labib and V. R. Vemuri, "Detecting and visualizing denial-of-service and network probe attacks using principal component analysis," in *Proceedings of 3rd Conference on Security and Network Architectures (SAR '04)*, La Londe, France, June 2004.