An effective cyber attack detection system based on an improved OMPCA

ELKHADIR Zyad
LASTID Research Laboratory
Ibn Tofail University
Kenitra, Morocco
Email: zyad.elkhadir@gmail.com

CHOUGDALI Khalid
GEST Research group
National School of Applied Sciences (ENSA)
Ibn Tofail University
Kenitra, Morocco
Email: chougdali@gmail.com

BENATTOU Mohammed
LASTID Research Laboratory
Ibn Tofail University
Kenitra, Morocco
Email: mbenattou@yahoo.fr

***Abstract*—Countering network threats, particularly intrusions, is a challenging area of research in the field of information security. Intruders use sophisticated mechanisms to hide the attack payload and break the detection techniques. To overcome that, various unsupervised learning approaches from the field of machine learning and pattern recognition have been employed. The most popularly used method is Principal Component Analysis (PCA). It proposes to extract the critical features of a network connection, then, it exploit them to identify the intrusion. However, PCA approach is prone to outliers due to the square ℓ_2-norm based objective function. As a solution to that, many PCA variants such R1-PCA and ℓ_1-PCA were proposed. Nevertheless, They still work with the square ℓ_2-norm distance based mean, which is not the optimal mean. This paper introduces a new variant of PCA namely QR-OMPCA. Firstly, this method integrates the mean calculation into the feature extraction function, such that the optimal mean can be obtained to enhance the intrusion detection accuracy. Secondly, it incorporates a fast QR decomposition. Experiments on KDDcup99 and NSL-KDD datasets confirm the superiority of the proposed method over many PCA variants in terms of intrusion detection accuracy and CPU time reduction.**

***Index Terms*—PCA, R1-PCA, OMPCA, QR-OMPCA, Network Anomaly Detection, KDDcup99, NSL-KDD.**

I. Introduction

Intrusion detection system (IDS) is a process of monitoring, detecting, and analyzing the events that are seen as violation to the security policies of a networked environment. This concept saw the light of day thanks to Denning [1]. Attackers are always looking for disrupting network traffic and degrade it performance with numerous types of attacks and intrusions. According to network security jargon, intrusion can be depicted as actions that compromise the confidentiality, integrity, or availability of information resource. Therefore, it is necessary to take different measures to minimize such risks.

There are in general two preferences for IDS. Host-based IDS and network-based IDS. Correspondingly, the detection methods used in IDS are anomaly detection and misuse detection. Each one has their own advantages and shortcomings. In misuse-based detection, data gathered from the system is compared to a set of rules or patterns already stored in a database. If there is any matching, the IDS sends an alarm that confirm the existence of an attack. This approach is very expensive, because we have to assemble regularly the all possible variations of the intrusion to decrease the false-negative and false-positive alarms. As a result, this concept does not contribute enough in zero-day attack detection. The other concept uses a collection of examples that represent normal behavior and constructs a model of familiarity. Therefore, any action that deviates from the model is considered suspicious and it is classified as an intrusion. It main advantage is the ability to detect novel attacks. Nonetheless, this approach produces a high false alarm rate. A potential cause of this phenomenon is the manipulation of large network traffic data with useless features.

To tackle with this restriction, many data dimensionality reduction techniques have been exploited. The most employed one is Principal Component Analysis (PCA), used in [2] [3] [4]. It projects the original high-dimensional feature space to a low-dimensional space, wherein the important features are well preserved.

Although PCA [5] is popular, it is sensitive to the outliers due to the square ℓ_2-norm based objective function. In the field of network security, the data outliers often appear in the network connections, thus PCA may not give the optimal results. To address this problem, multiple robust PCA methods have been presented, such as the rotational invariant PCA (R1-PCA) [6], convex robust PCA [7] and ℓ_1-PCA [8]. In R1-PCA, the authors solve the problem by reducing a ℓ_2,ℓ_1-norm error. That was done mainly by adopting the ℓ_2-norm on the feature dimension and the ℓ_1-norm on the data points dimension. After that, this concept was generalized to robust tensor factorization [9]. Consequently, with convex relaxation objectives, the global solutions can be obtained. ℓ_1-PCA adopts a greedy strategy to maximize a ℓ_1-norm dispersion instead of the euclidean norm. In the same manner, the paper [10] maximizes a more general Lp-norm using the conjugate gradient algorithm. Nevertheless, all the aforementioned PCA methods do not propose an algorithm for mean calculation. They still work with the square ℓ_2-norm distance based mean, which is a biased mean.

Optimal mean PCA [11] tries to handle this problem by introducing an iterative re-weighted method which remove

978-1-5386-2123-3/17/$31.00

the optimal mean automatically from the objective function. However, this method includes a singular value decomposition (SVD) to find the principal features. However, it is known in literature that The computational complexity of SVD method is high.

In this paper, to solve the aforementioned drawback, we propose a new variant of PCA namely QR-OMPCA. The method improves Optimal mean PCA by replacing the SVD decomposition with a more faster and a stable decomposition. The rest of this paper is organized as follows. In Section II, we revisited PCA , after that, Section III introduces Optimal mean PCA (OMPCA). In Section IV, we describe the proposed approach. Section V reports the experimental results and demonstrate the effectiveness of QR-OMPCA and show it superiority compared to other PCA variants. Finally, section VI summaries the principal obtained results.

II. Principal Component Analysis Revisited

Suppose we have $X = [x_1, \ldots, x_n] \in \mathbb{R}^{d\times n}$ a centered data matrix, where d and n represent the dimension and total number of the samples x_i respectively. Principal Component Analysis (PCA) looks for a new matrix Z which approximate the given matrix in a reduced dimension space. Mathematically speaking, PCA tries to solve the following problem:

$$\min_{rank(Z)=k} ||X - Z||_F^2 \tag{1}$$

To do that, we try to find the projection matrix W that satisfies:

$$\min_{W\in R^{d\times k}, V\in R^{n\times k}, W^TW=I} ||X - WV^T||_F^2 \tag{2}$$

Then the problem becomes:

$$\max_{W\in R^{d\times k}, W^TW=I} Tr(W^TXX^TW) \tag{3}$$

The matrix W represents the k eigenvectors of XX^T corresponding to the $k$largest eigenvalues. These vectors are called principal components (PCs).

We observe from (1) that samples with large norms (outliers) will dominate due to the ℓ_2-norm. As a result, the process of finding W will not be accurate. In order to solve this, many papers proposed to replace the ℓ_2-norm by the ℓ_1-norm which is more insensitive to outliers. These works gave good results but still incomplete. In fact, they do not consider the optimal mean calculation. To center the data X they remove the square ℓ_2-norm distance based mean from it. That mean will not be the optimal mean due to the ℓ_1-norm used in the objective functions.

III. optimal mean PCA (OMPCA)

Motivated by the above mentioned issue, the paper [11] introduces the calculation of the correct mean inside the objective function. Then equation (1) will be reformulated as:

$$\min_{b, rank(Z)=k} ||X - b1^T - Z||_2 \tag{4}$$

Denote 1 as a column vector with all the elements being one and b is also a variable to be optimized. This can be rewritten as a sum:

$$\min_{b, W\in R^{d\times k}, W^TW=I} \sum_{i=1}^{n} ||X - b - W(v^i)^T||_2 \tag{5}$$

Including $v^i = (x_i - b)^TW$ inside the above equation gives:

$$\min_{b, W\in R^{d\times k}, W^TW=I} \sum_{i=1}^{n} ||(I - WW^T)(x_i - b)||_2. \tag{6}$$

This problem can be approximated by:

$$\min_{b, W\in R^{d\times k}, W^TW=I} \sum_{i=1}^{n} ||d_{ii}(I - WW^T)(x_i - b)||_2^2. \tag{7}$$

S.t d_{ii} represents the diagonal elements of a weighted matrix D. After many steps the problem in (7) becomes :

$$\max_{W\in R^{d\times k}, W^TW=I} Tr(W^TXH_dX^TW). \tag{8}$$

where $H_d = D - \frac{D11^TD}{1^TD1}$.

Then, the authors use an iterative re-weighted method to solve (8). In every iteration, to find W, they apply SVD on:

$$M = X(D^{1/2} - \frac{D11^TD^{1/2}}{1^TD1}). \tag{9}$$

S.t $XH_dX^T = MM^T$. The detailed algorithm is below outlined:

Algorithm 1 : Optimal mean PCA algorithm

1) Initialize D as an identity matrix
2) Update W by the k right singular vectors of M corresponding to the k largest singular values.
3) Update b by $b = \frac{XD1}{1^TD1}$.
4) Update D by $d_{ii} = \frac{1}{2||d_{ii}(I-WW^T)(x_i-b)||_2}$.
5) If W does not converge goto Step 2. Else, return W and Stop iteration.

Since this algorithm uses SVD decomposition, it will be called SVD-OMPCA in the rest of paper.

IV. QR-OMPCA method

The previous algorithm has an important computation deficiency. If the dimensionality is extremely large ($d >> n$), then the computation of SVD(M) in step2 becomes really slow [12], it requires around $14dn^2 - 2n^3$ flops. To overcome that, we propose a new QR-OMPCA which can extract eigenvectors and eigenvalues of MM^T in a numerically stable manner, moreover, its computational complexity is inferior than the SVD.

Let the rank of $MM^T \in R^{d\times d}$ be r, where $1 \leq r \leq n$. The matrix M can be decomposed into orthogonal matrix $Q \in R^{d\times r}$ and upper triangular matrix $R1 \in R^{r\times n}$ using economic QR decomposition as:

$$M = QR1. \tag{10}$$

After that, MM^T will equals $QR1R1^TQ^T$. The matrix $R1^T$ can be decomposed by SVD and written as:

$$R1^T = U1D1V^T. \quad (11)$$

where $U1 \in R^{n\times r}$ and $V \in R^{r\times r}$ are orthogonal matrices and $D1 \in R^{r\times r}$ is a diagonal matrix. Substituting Eq. 11 in Eq. 10, we get :

$$MM^T = QVD2V^TQ^T. \quad (12)$$

S.t $D2 = D1^2$. Since $(QV)^T(QV) = I$, we conclude that (QV) is an orthogonal matrix. The latter also diagonalizes MM^T. These two facts confirm that QV is an eigenvector matrix and $D2$ is an eigenvalue matrix of MM^T.

Therefore, the projection matrix would be $W = QV_k$ S.t V_k are the k eigenvectors corresponding to the largest k diagonal entries of D2. Having said that, the QR-OMPCA algorithm is summarized as follow:

Algorithm 2 : QR-OMPCA algorithm

1) Initialize D as an identity matrix
2) Compute Q and $R1$ using economic QR decomposition of M.
3) Compute $D1$ and V with the help of economic SVD on $R1^T$.
4) Obtain V_k the k eigenvectors corresponding to the largest k diagonal entries of $D2$.
5) $W = QV_k$.
6) Update b by $b = \frac{XD1}{1^TD1}$.
7) Update D by $d_{ii} = \frac{1}{2||d_{ii}(I-WW^T)(x_i-b)||_2}$.
8) If W does not converge goto Step 2. Else, return W and Stop iteration.

To compare the computational complexity of QR-OMPCA and OMPCA, it is sufficient to compare the complexity of step 2 in Algorithm 1 with the complexity of steps 2-5 in Algorithm 2. The economic QR decomposition M needs $2dn^2$ flops (if modified Gram-Schmidt QR method is used) or $2dn^2 - 2n^3/3$ flops (if fast Givens QR method is used) [12]. The economic SVD of $R1^T$ in equation (11) requires $4nr^2 + 8t^3$ flops [12]. Lastly, the product of Q and V_k requires $2drk$ flops. The total estimation of QR-OMPCA is $2dn^2 + 2drk$ flops, which is inferior than the SVD one.

V. Experiments and Results

A. Data sets

In the experiments, we work on 2 popular benchmark intrusion datasets. Namely KDDcup99 and NSL-KDD. We give some brief descriptions about them in the below paragraphs.

1) KDD Cup 99 data set [13] from UCI repository is widely used as the benchmark data set for IDS evaluation. it contains 4,898,431 and 311,029 records in the training set and test set, respectively. Each of record contains 41 features and is labeled as either normal or an attack, with exactly one specific attack type. In our experiments, we apply its 10% data set consisting of 494021 connection records for training. All attacks fall in one of the following four categories:
   a) Denial of Service (DoS): Is an attempt to consume network resources in such a way that their services become limited or unavailable for the legitimate users.
   b) User to Root (U2R): Is an attack in which the attacker starts accessing a normal user account on a machine and gains root access to the machine by exploiting vulnerabilities.
   c) Remote to Local (R2L): Occurs when an attacker does not have an account on a remote system, but who has the ability to send packets to a system over a network and exploits vulnerabilities to gain local access as a user of that system.
   d) Probing: An attack in which the attacker scans network to collect information about its systems for the apparent purpose of circumventing its security controls.

   KDD Cup 99 features can be classified into three groups:
   a) Basic features: this category encapsulates all the attributes that can be extracted from a TCP/IP connection. Most of these features leading to an implicit delay in detection.
   b) Traffic features: this category includes features that are computed with respect to a window interval.
   c) Content features: Most of the DoS and Probing attacks have many intrusion frequent sequential patterns, this is due to the fact that these attacks establish many connections to the host(s) in a very short period of time. Unlike these attacks, the R2L and U2R attacks do not have any intrusion frequent sequential patterns. The R2L and U2R attacks are embedded in the payload of the packets, and normally include only a single connection. To identify these kinds of attacks, some relevant features are needed to identify suspicious behavior in the packet payload. These features are called content features.
2) NSL-KDD [14] proposes to solve some of KDDcup99 problems. The principal improvements are as follow. The number of NSL-KDD records in the train and test sets is more reasonable. This advantage makes it a good choice to run the approaches on the complete KDD Cup 99 data set without the need to randomly select a small portion. It should be mentioned that the test set is not from the same probability distribution as the training set, and it includes unknown attack types that do not exist in the training set that makes it more realistic. The data sets contain a total number of 22 training attack types, with an additional 17 types in the test data set.

B. Performance Measures

To estimate the reliability of an IDS, the community of network security use 2 popular measures:

1) The False Positive Rate (FPR) is defined as the number of normal records that are incorrectly detected as intrusions divided by the total number of normal records.
2) The Detection Rate (DR) is defined as the number of intrusion records classified by the IDS divided by the total number of intrusion records present in the test data set.

These are good performance measures, since they measure what percentage of intrusions the system is able to detect and how many incorrect classifications are made in the process. They are defined by the following equations:

$$DR = \frac{TP}{TP + FN} * 100 \tag{13}$$

$$FPR = \frac{FP}{FP + TN} * 100 \tag{14}$$

where TP, TN, FP, FN are the numbers of true positives, true negatives, false positives and false negatives, respectively.

C. Experimental Results

In this section, we perform many experiments to evaluate the performance of the proposed approach using the K-NN classifier. In Figs 1 to 4, we compare PCA, R1-PCA and QR-OMPCA in term of DR and FPR. Since these methods can share the common reduced dimensionality, they can be compared under the same dimension. In the experiment we choose to work on 4 dimensions $(k = 4)$. Then, we increase the size of the training dataset and illustrate it effect on DR and FPR behaviors. We change the size of training samples with the following manner: We began with 1350 training samples composed of 1000 normal samples, 100 DOS attacks, 50 U2R, 100 R2L and 100 PROBE attacks. After that, we work on 2000 normal samples, 200 DOS, 50 U2R, 100 R2L and 200 PROBE attacks.etc. U2R and R2L still unchanged due to the rarity of these kind of attacks in the datasets. We fix test dataset with 100 normal samples, 100 DOS, 50 U2R, 100 R2L and 100 PROBE attacks. It should be noted that the samples were selected randomly from the datasets.

Fig.1 shows that QR-OMPCA overcomes R1-PCA and PCA in the most of the times in term of DR. Fig.2 asserts that the proposed approach and R1-PCA produce the lowest FPR. Those were the results which concern KDDcup99.

When we reproduce the same experiment on NSL-KDD as depicted in Fig.3 and Fig. 4, the proposed approach surpasses R1-PCA and PCA continuously. Nevertheless, we observe that increasing training data leads to a DR diminution. The explanation of this fact resides in training data nature. The latter is closer to normal than malicious nature. Consequently, creating a model that detects attacks becomes really difficult once the number of training samples reaches a high value. That what causes DR deterioration. Fig.4 confirms that QR-OMPCA gives the fewest FPR. This trivial result flows directly from the important resistance of QR-OMPCA against outliers.

In the next experiment, we compare QR-OMPCA with SVD-OMPCA in term of detection rate and time consuming. To do that, we generate a training data composed of 1000

Fig. 1. The number of training samples vs. DR(%) for KDDcup99

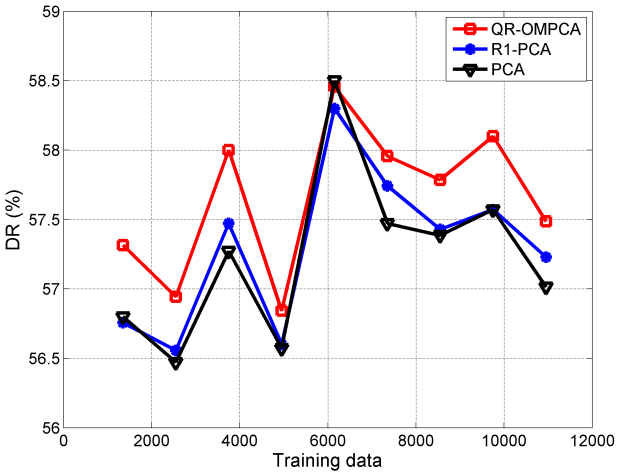


Fig. 2. The number of training samples vs. FPR(%) for KDDcup99

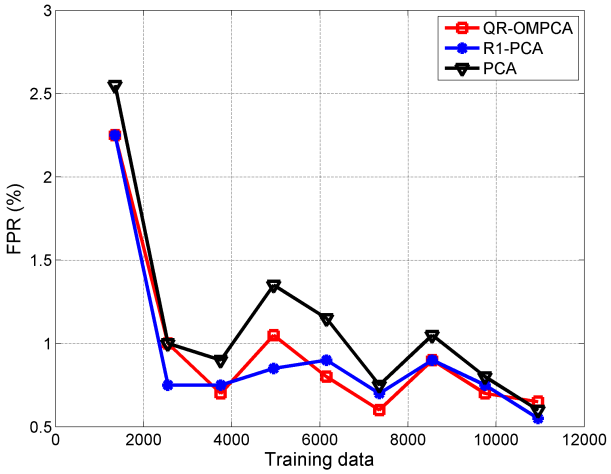


Fig. 3. The number of training samples vs. DR (%) for NSL-KDD

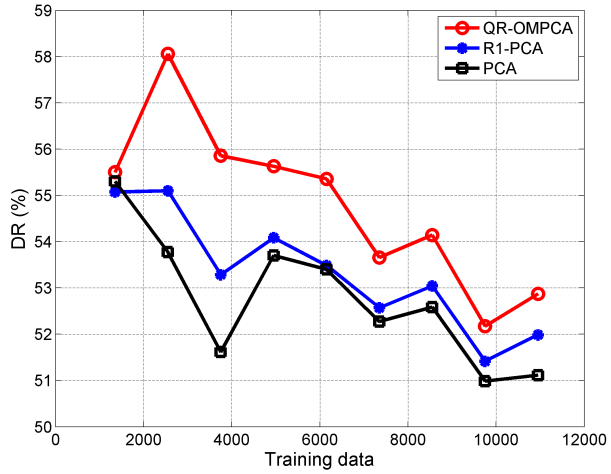


Fig. 4. The number of training samples vs. FPR (%) for NSL-KDD

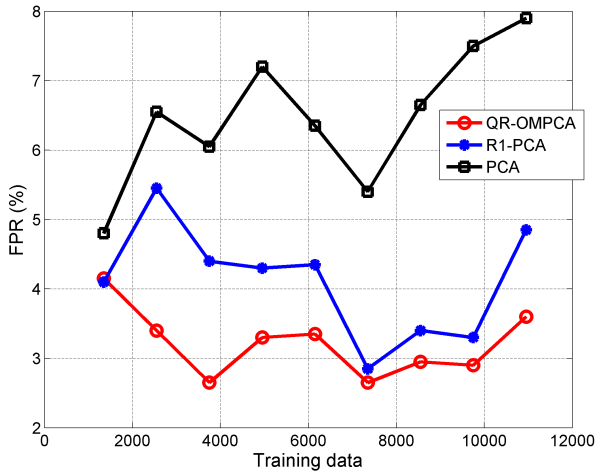


Fig. 5. The number of principal components vs. DR(%) for KDDcup99

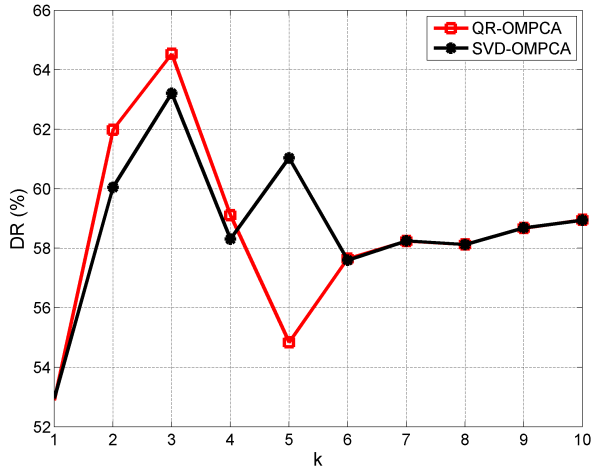
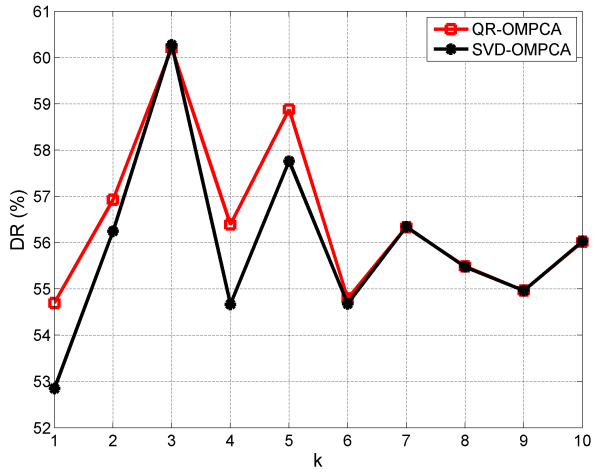


Fig. 6. The number of principal components vs. DR(%) for NSL-KDD



normal samples, 100 DOS attacks, 50 U2R, 100 R2L and 100 PROBE attacks. Then, we varied the number of principal components k from 1 to 10 and visualize the consumed CPU time. Figs 5 and 6 show the DR of the two methods when applied on the two databases. It can be seen from these figures that the proposed method is more accurate than the SVD-OMPCA when k is inferior than 4. Once the latter is exceeded, the results become similar. The last figures depict the relationship between the number of principal components and CPU time. On the one hand we observe that increasing k leads to a high consuming time. On the other hand, the figures show that the proposed method is computationally faster than SVD-OMPCA. The rapidity of QR-OMPCA mainly comes from the exploitation of QR decomposition. The latter decomposes the matrix M into orthogonal matrix and upper triangular matrix. After that, SVD decomposition can be applied just to the upper triangular matrix. Following this path will be less greedy compared to using SVD directly on the rectangular matrix M.

Fig. 7. The number of principal components vs. CPU time (s) for KDDcup99

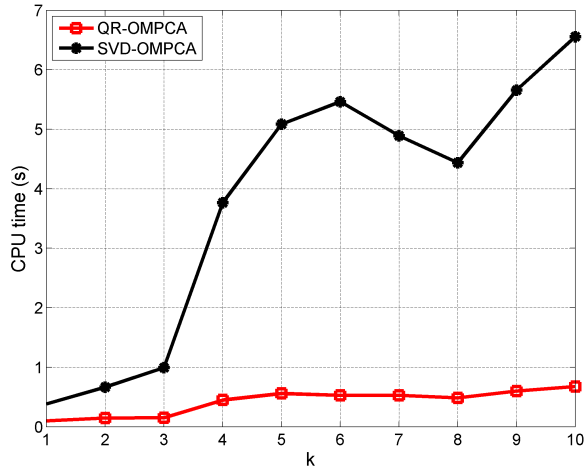
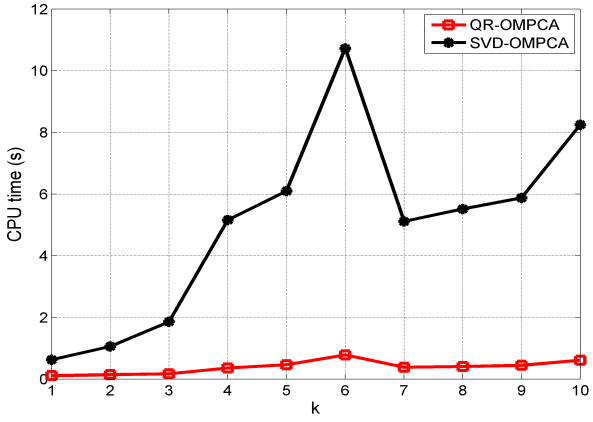


Fig. 8. The number of principal components vs. CPU time (s) for NSL-KDD



VI. Conclusion

In this paper, we have proposed a PCA-based dimension reduction algorithm, called QR-OMPCA. This method incorporates the mean calculation such that the optimal mean can be obtained. In addition, it applies QR decomposition rather than SVD. Integrating QR-OMPCA into the intrusion detection system (IDS) makes the latter more effective and resistant against outliers. Furthermore, it speeds up the IDS process. Such a facts were verified by numerous experiments on KDDcup99 and NSL-KDD. Meanwhile, The results show that QR-OMPCA outperforms R1-PCA, PCA and OMPCA.

References

[1] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering,*, no. 2, pp. 222–232, 1987.

[2] C. Callegari, L. Gazzarrini, S. Giordano, M. Pagano, and T. Pepe, "Improving pca-based anomaly detection by using multiple time scale analysis and kullback–leibler divergence," *International Journal of Communication Systems*, vol. 27, no. 10, pp. 1731–1751, 2014.

[3] K. K. Vasan and B. Surendiran, "Dimensionality reduction using principal component analysis for network intrusion detection," *Perspectives in Science*, vol. 8, pp. 510–512, 2016.

[4] Z. Elkhadir, K. Chougdali, and M. Benattou, "Intrusion detection system using pca and kernel pca methods," in *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015*. Springer, 2016, pp. 489–497.

[5] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[6] C. Ding, D. Zhou, X. He, and H. Zha, "R 1-pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 281–288.

[7] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[8] N. Kwak, "Principal component analysis based on l_1-norm maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1672–1680, 2008.

[9] H. Huang and C. Ding, "Robust tensor factorization using r 1 norm," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[10] Z. Elkhadir, K. Chougdali, and M. Benattou, "Network intrusion detection system using pca by lp-norm maximization based on conjugate gradient," *International Review on Computers and Software (IRECOS)*, vol. 11, no. 1, pp. 64–71, 2016.

[11] F. Nie, J. Yuan, and H. Huang, "Optimal mean robust principal component analysis," in *Proceedings of the 31st international conference on machine learning (ICML-14)*, 2014, pp. 1062–1070.

[12] G. H. Golub and C. F. Van Loan, "Matrix computations. 1996," *Johns Hopkins University, Press, Baltimore, MD, USA*, pp. 374–426, 1996.

[13] [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/

[14] [Online]. Available: http://nsl.cs.unb.ca/KDD/NSLKDD.html