Robust Traffic Anomaly Detection with Principal Component Pursuit

Atef Abdelkefi[1], Yuming Jiang[1], Wei Wang[2], Arne Aslebo[3], Olav Kvittem[3]
[1] Q2S,Norwegian University of Science and Technology, Norway
[2] University of Luxembourg, Luxembourg
[3] UNINETT, Norway

1. INTRODUCTION

Principal component analysis (PCA) is a statistical technique that has been used for data analysis and dimensionality reduction. It was introduced as a network traffic anomaly detection technique firstly in [1]. Since then, a lot of research attention has been received, which results in an extensive analysis and several extensions. In [2], the sensitivity of PCA to its tuning parameters, such as the *dimension of the low-rank subspace* and the *detection threshold*, on traffic anomaly detection was indicated. However, no explanation on the underlying reasons of the problem was given in [2]. In [3], further investigation on the PCA sensitivity was conducted and it was found that the PCA sensitivity comes from the inability of PCA to detect temporal correlations. Based on this finding, an extension of PCA to Kalman-Loeve expansion (KLE) was proposed in [3]. While KLE shows slight improvement, it still exhibits similar sensitivity issue since a new tuning parameter called *temporal correlation range* was introduced. Recently, in [4], additional effort was paid to illustrate the PCA-*poisoning* problem. To underline this problem, an evading strategy called *Boiled-Frog* was proposed which adds a high fraction of outliers to the traffic. To defend against this, the authors employed a more robust version of PCA called PCA-GRID. While PCA-GRID shows performance improvement regarding the robustness to the outliers, it experiences a high sensitivity to the threshold estimate and the k-dimensional subspace that maximizes the dispersion of the data. The purpose of this work is to consider another technique to address the PCA poisoning problems to provide robust traffic anomaly detection: The *Principal Component Pursuit*.

ACM CoNEXT Student Workshop, November 30, Philadelphia, USA.

2. THE PRINCIPAL COMPONENT PURSUIT

Reason behind PCA Poisoning: A foundation of PCA is that it seeks the best rank-k estimate E of a data matrix X by solving:

$$\min \ \|X - E\|_2, \ \textit{subject to} \ \ \textit{rank}(E) \leqslant k,$$

where $\| \ \|_i$, $(i = 0, 1, 2)$ denotes the ℓ_i-norm.

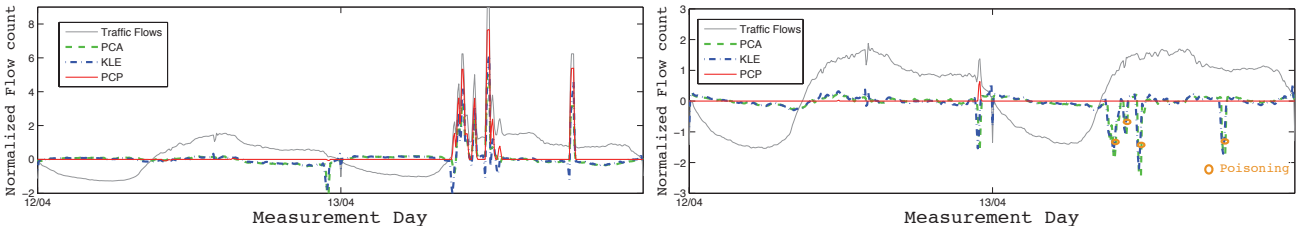
It is worth highlighting that here, the rank k has to be known a priori. However, this requirement can hardly be met in real network environments. In addition, the problem is a *least square* problem. It is well-known that, while the least square gives optimal results in the presence of Gaussian noise (e.g., low and medium intensity attacks), it has been proven that it's very sensitive to the outliers (known as high intensity attacks). This is mainly due to the fact that it is an ℓ_2-norm fitting problem [7], leading to the *low-rank subspace poisoning* phenomenon. This phenomenon depicts the deviation of the principal components from the true distribution of the data toward even a low fraction of outliers. As a consequence, shifted principal components conduct to a perturbation in the low-rank subspace, resulting in inaccurate detection and high false positive rates.

Principal Component Pursuit (PCP): In order to make the detection robust to the poisoning phenomenon and without a priori requirement for the calibration setting, we advocate employing *Sparse and Low-Rank Matrix Decomposition* approach.

It is well-known that real traffic data, X, includes an unknown fraction of normal traffic, N, while the remaining fraction is corrupted by different types of anomalies, A. Since *normal traffic* is known to be *low rank* [1] and *anomalies* to be temporally localized or *sparse* in time, an accurate anomaly detector aims to decompose the observed traffic data into these two components and essentially to solve:

$$\min_{N,A} \|A\|_0, \ \text{subject to } X = N + A, \ \text{rank } (N) \leq k \quad (1)$$

This optimization problem is NP-hard [5]. However, based on the recent advances in convex optimization theory, it has been proven that the nuclear norm, i.e, the sum of singular values, exactly recovers the low rank



(a) PCA, KLE & PCP residuals on Link 4

(b) PCA, KLE & PCP residuals on Link 9

Figure 1: PCA, KLE and PCP residuals for traffic per link data

component [5] while the ℓ_1 norm, i.e, the sum of absolute values, exactly recovers the sparse component with a remarkable robustness to the outliers in comparison to the ℓ_2 norm [7]. Accordingly, Equation 1 can be solved using the *Principal Component Pursuit* [5] defined as:

$$\min_{N,A} \|N\|_* + \lambda\|A\|_1, \ \textit{subject to}\ X = N + A, \qquad (2)$$

where X denotes the traffic data $X \in \mathbb{R}^{n_1 \times n_2}$, $\| \ \|_*$ denotes the nuclear norm of the normal data traffic matrix N and λ is a weighting parameter. In the rest, we call the term $\|A\|_1$ the PCP residual.

To solve such a convex problem, different solvers have been proposed. We opt for the one proved to converge Q-linearly to the optimal solution using the inexact version of the *Augmented Lagrange Multiplier*(ALM) solver [6]. Interestingly, one might expect that the scalar λ has to be properly chosen to balance the two terms $\|N\|_*$ and $\|A\|_1$. However, our experiment results show that the value of $\lambda = \frac{2}{\sqrt{max(n_1,n_2)}}$, suggested in [5], provides appealing performance, making the detection free of tuning parameters.

3. EXPERIMENT RESULTS

Our experiment used Netflow data captured in the second week of April 2010 at the Norwegian Academic and research Network (Uninett). To evaluate the performance of PCA, KLE and PCP, we analyzed the amount of traffic per link data matrix. Common traffic feature such as the flows amount were studied for the biggest router in the Norwegian backbone that has 9 links. All studied metrics are obtained by aggregating traffic at 5 minutes time bin. The studied traffic data involve a variety of real and injected anomalies, identified by using visual inspection and several existing detection techniques. Fig. 1 compares the performance of PCA and KLE versus PCP-based anomaly detection. It is observed that the PCP residual is sparse in time, which makes unnecessary to estimate a detection threshold. Only the non-zero elements in the residual vector are considered to be anomalous events. Additionally, PCA and KLE show performance degradation: Link 9 experiences false non-zero elements for the same

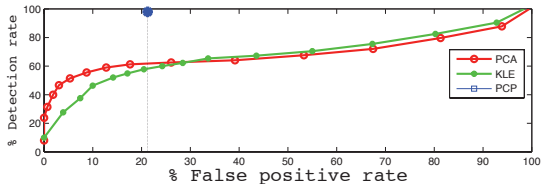


Figure 2: ROC curves: PCA, KLE and PCP

time bins where high intensive attacks (outliers) are encountered on Link 4. This illustrates the poisoning phenomenon propagating false positives in the observed data. However, unlike PCA and KLE approaches, PCP shows robustness against poisoning and has better outliers detection and less false positive rate. To look more closely at the robustness behavior, the Receiver Operating Characteristic (ROC) curves of the three different approaches are presented in Fig. 2. It is clear from the figure that PCP shows a stable better performance. Particularly PCP experiences a stable performance of 100 % detection rate at a false alarm rate of 21%, compared to 61% detection rate for PCA and 55% for KLE at the same false positive rate. We hence believe that PCP is a more robust traffic anomaly detector.

4. REFERENCES

[1] A. Lakhina, et al. "Diagnosing Network-Wide Traffic Anomalies". *ACM SIGCOMM*, 2004.

[2] H. Ringberg, et al. "Sensitivity of PCA for Traffic Anomaly Detection". *ACM SIGMETRICS*, 2007.

[3] D. Brauckhoff, et al. "Applying PCA for Traffic Anomaly Detection: Problems and Solutions". *IEEE INFOCOMM*, 2009.

[4] B. Robinstein, et al. "ANTIDOTE: Understanding and Defending Against Poisoning of Anomaly Detectors". *ACM IMC*, 2009.

[5] E. Candes, et al. "Robust Principal Component Analysis ?". *preprint*, 2009.

[6] Z. Lin, et al. "The Augented Lagrange Multiplier Method for Exact Recovery of Corrupted Low Rank Matrices". *preprint*, 2009.

[7] D. Donoho. "Compressed Sensing". *IEEE ToIT*, 2006.