A Dictionary Learning Based Anomaly Detection Method for Network Traffic Data

Taha Yusuf Ceritli * YUSUF.CERITLI@BOUN.EDU.TR
Barış Kurt † BARIS.KURT@BOUN.EDU.TR
Çağatay Yıldız † CAGATAY.YILDIZ1@BOUN.EDU.TR
Bülent Sankur ‡ BULENT.SANKUR@BOUN.EDU.TR
Ali Taylan Cemgil † TAYLAN.CEMGIL@BOUN.EDU.TR

* Bogazici University, Department of Computational Science and Engineering, 34342 Bebek, Istanbul, TURKEY

†Bogazici University, Department of Computer Engineering, 34342 Bebek, Istanbul, TURKEY

‡Bogazici University, Department of Electrical & Electronics Engineering, 34342 Bebek, Istanbul, TURKEY

Abstract

In this paper we propose a dictionary learning scheme to extract network traffic pattern templates for different types of anomalies and the normal traffic via nonnegative matrix factorization. We employ Bayesian change point models on the representation of the running network traffic in terms of those templates to detect network anomalies. Our proposed methods are tested and evaluated on a simulated SIP network with attacks generated by a commercial network vulnerability scanning tool.

1. Introduction

In stationary time series data, such as large Voice Over IP (VoIP) network traffic, significant deviations from the usual data pattern over a period of time is considered as anomaly (Chandola et al., 2009). One of the first requirements in providing a secure network service is to detect such anomalies, since they may be network attacks such as the most commonly occurring attacks, Distributed Denial of Service (DDoS). The DDoS attack aims to consume network server resources to make the system unresponsive to valid user requests (Keromytis, 2012), and therefore should be detected as early as possible.

In this work we focus on detecting several types DDoS attacks in VoIP networks running with the Session Initiation Protocol (SIP). The SIP (Rosenberg et al., 2002) has become a very popular text-based request-response type protocol to initialize, audit, modify and end VoIP sessions between the clients. Its lightweight nature, simplicity and the ease of implementation made SIP a commonly used VoIP protocol but also a direct target for cyber-attacks.

Along with the classical DDoS attacks, such as SYN flooding, SIP servers are prone to the attacks that try to exploit the SIP protocol. For example in one type of DDoS attack, called *REGISTER* attack, the server is flooded with invalid register requests, and the server is kept busy, requesting passwords from non-existing users. In another type of attack, called *INVITE* attack, invalid call requests are sent from unregistered users.

The patterns of the different types of attacks and the normal traffic can be represented in terms of a dictionary of network traffic patterns. In this work, we propose a dictionary learning mechanism, based on Nonnegative Matrix Factorization (NMF) to create a dictionary of traffic patterns and to express the running network traffic data as a linear combinations of the dictionary atoms.

Previously, using NMF for dimensionality reduction prior to the anomaly detection has been proposed by (Xiaohong et al., 2009). Furthermore, detection of DDoS attacks on networks via Bayesian Change Point Models (BCPM) has been shown to be a robust method with low latency and high accuracy (Wang et al., 2004). In this work, we define a methodology that combines NMF with BCPM, whose details are given in the next section. We tested our methodology in a simulated environment as described in Section 3. The last section is dedicated to the evaluation of current results and future research directions.

2. Methodology

In our methodology, we reduce the anomaly detection task on streaming data into an online change point detection problem. A change point model in general, keeps track of a feature vector and calculates the probability of a change for any given vector. In the context of SIP networks, the input may consist of the histogram of incoming and outgoing SIP packet types since the concentration of the feature vector is expected to be widely different under the normal traffic and DDoS attacks. The features that NMF generates via dictionary of traffic patterns can also be used as input to models.

2.1. Bayesian Change Point Model

BCPM's are a special form of hierarchical Markov models (Fearnhead, 2006). In our notation, observations and latent variables are denoted by x_t and (π_t, r_t), respectively. As can be seen from the graphical model in Figure 1, π_t is a Markov chain conditioned on r_t, which is a binary variable indicating a change in the regime. Furthermore, observations x_t are conditioned on latent states π_t.

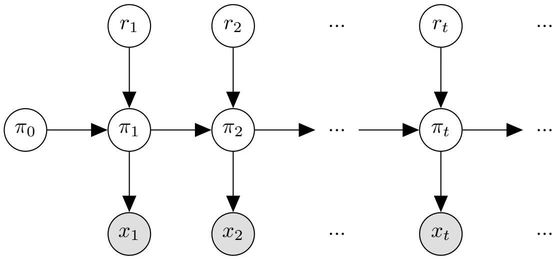

Figure 1. Bayesian change point graphical model.

The observation at time t, x_t, is assumed to be a random variable sampled from a $\Theta(x;\pi)$ distribution with an unknown parameter π_t. The model allows π_t to change multiple times during the process. Initially, π_0 is drawn from a $\Omega(\pi)$ distribution. Afterwards, at each time step t, π_t is either re-drawn from the same distribution or set to the previous value π_{t-1}. The decision for change is given by a Bernoulli random variable r_t. Hence, the complete generative model is given as:

$$\pi_0 \sim \Omega(\pi_0) \tag{1}$$

$$r_t \sim p^{r_t}(1-p)^{(1-r_t)} \tag{2}$$

$$\pi_t | r_t, \pi_{t-1} \sim [r_t = 0]\delta(\pi_t - \pi_{t-1}) + [r_t = 1]\Omega(\pi_t) \tag{3}$$

$$x_t | \pi_t \sim \Theta(x_t; \pi_t) \tag{4}$$

where δ is Dirac delta function.

Posterior probability of a change at any time t can be calculated using Forward-Backward Algorithm. Models in which r_t is a two-valued variable and Ω is the conjugate prior of Θ, recursions can be implemented in linear complexity with time(Barber & Cemgil, 2010). When the inference task is online, only forward recursion is implemented and $p(r_t|x_{1:t})$ is calculated. The degree of belief on the posterior can be improved by fixed lag smoothing, if a certain amount of delay is admitted.

2.2. Nonnegative Matrix Factorization

Nonnegative matrix factorization (NMF) is a decomposition method for multivariate data. Formally, NMF aims to find positive matrices T and V matrices given a nonnegative matrix X such that:

$$x_{\nu,\tau} \approx [TV]_{\nu,\tau} = \sum_i t_{\nu,i} v_{i,\tau} \tag{5}$$

where i =1:I, ν =1:W and τ =1:K. T is called the template matrix with $W \times I$ and $I \times K$ dimensional matrix V is referred as the excitation matrix. Due to the nonnegativity constraint, NMF is widely used as a dictionary learning technique where the goal is to find a more sparse representation of data as linear combinations of certain patterns (Lee & Seung, 1999).

NMF can be described as a hierarchical model where original multiplicative update rules appears as an expectation-maximization (EM) algorithm for estimation of Poisson model via data augmentation. Variational methods can be used for inference in the resulting models. This leads to the full Bayesian treatment of Nonnegative Matrix Factorization for model selection (Cemgil, 2009).

Figure 2 shows the decomposition of the SIP traffic histograms. The X matrix represents the count of packet types over time; one can see that 5 different DDoS attack types occur, interspersed normal traffic intervals.The excitation and template matrices inferred by the NMF model are shown on the right side. We can see that the first column of the template matrix represents the normal traffic, and the other columns correspond to different types of attacks. For example, the second column represents an *OPTIONS* attack, with high concentration of *OPTIONS* requests and its corresponding server response: *200* messages.

2.3. Anomaly Detection Models

In this paper, we extended the Multinomial-Dirichlet (MD) model proposed in (Yıldız et al., 2016) by feeding the model with excitation vectors extracted by NMF. We also proposed another change point model with Poisson-Gamma (PG) observations and tested it with both the raw network data and the features extracted by NMF.

2.3.1. Model 1: MD

In the Multinomial-Dirichlet (MD) model, feature vectors collected from the network data are used as input to the change point model. Also, observations are assumed to be generated by a Multinomial distribution, whose parameters

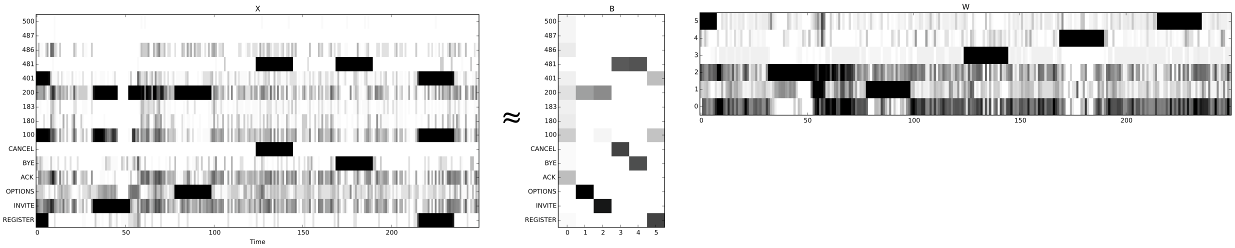

Figure 2. The figures represent the data, template and corresponding excitation matrices, respectively from left to right. The (x,y) axes of the matrices are (time,message types), (template indices,message types), (time, template indices) respectively.

are drawn from a Dirichlet distribution:

$$\Omega(\pi;\alpha) \equiv \mathcal{D}ir(\pi;\alpha) \tag{6}$$

$$\Theta(x;\pi) \equiv \mathcal{M}ult(x;\pi) \tag{7}$$

2.3.2. Model 2: PG

In Poisson-Gamma (PG) model, the feature vector is assumed to have independent elements which are given to separate change point models with Poisson-Gamma observations as input. Hence, the observation distribution of MD model is replaced by a Poisson distribution with parameters drawn from a Gamma distribution:

$$\Omega(\pi;\alpha,\beta) \equiv \mathcal{G}(\pi;\alpha,\beta) \tag{8}$$

$$\Theta(v;\pi) \equiv \mathcal{PO}(v;\pi) \tag{9}$$

2.3.3. Model 3: NMF-MD

In the NMF-MD model, the network features are expressed as a conical combination of template vectors, T, representing different network behaviors, which are calculated from the data prior to the experiment via NMF. Each observation x_t is transformed to the D dimensional representation v_t by running a few NMF iterations for the approximation

$$x_t \approx T v_t \tag{10}$$

2.3.4. Model 4: NMF-PG

The NMF-PG model is a variant of NMF-MD model, where the Multinomial-Dirichlet change point model is replaced by Poisson-Gamma change point models.

3. Experiments & Results

3.1. Experiment Design

Collecting real-world VoIP network traces and annotating them without violating the privacy of the users is a tedious task. Therefore, for the proof of our concept, we conducted our experiments in a simulated environment. We used an Asterisk-based PBX software, named *Trixbox* that serves as a SIP server (Trixbox). To mimic the normal traffic on a SIP server, we have built a probabilistic SIP network simulation system (Kurt et al., 2016), which initiates calls between a number of users in real-time. During the simulation, different types of DDoS attacks were generated by a vulnerability scanning tool named *NOVA V-Spy* (V-Spy).

The normal network traffic generation was controlled by two parameters of the SIP network simulation system: (1) Network traffic intensity and (2) the number of users. In our simulations, the traffic intensity was either low or high, and the number of users was either 50 or 250. In total, we had four different simulation settings, which are referred as *Low50*, *Low250*, *High50* and *High250* in Tables 1 and 2.

In each simulation setting, five types of DDoS attacks were realized by flooding one of *REGISTER*, *INVITE*, *OPTIONS*, *CANCEL* and *BYE* packets. In addition, *NOVA V-Spy* allows to tune the intensity and fluctuation levels of an attack. We set the attack intensity to be either of *low*, *medium* and *high*, and the fluctuation to be either *on* or *off*. At the end, 30 DDoS attacks per a simulation setting were performed.

We monitored the incoming and outgoing network traffic at the SIP server by capturing each network packet and generated histograms of the counts of 28 different types of SIP response and request messages within 1 second intervals. SIP request and response packets which were not generated by the network simulation system and the *NOVA V-Spy* were eliminated so that the 28 dimensional observation vectors were reduced to 16 dimensional vectors and provided as input to our anomaly detection system. Models 1 and 3 raise an alarm when the probability of change is greater than a certain threshold, which we took 0.95 during the experiments. In PG models, if the posterior probability calculated by any of the separate change point models is greater than the threshold, an alarm is raised.

	Low50	Low250	High50	High250
MD	0.77	0.68	0.81	0.60
NMF4-MD	0.93	0.66	**0.94**	**0.83**
NMF5-MD	**0.95**	**0.70**	0.83	0.67
NMF6-MD	0.84	0.63	0.82	0.60
NMF7-MD	0.83	0.62	0.85	0.58

Table 1. F-Scores of the Multinomial-Dirichlet Change Point Model when prior of probability of change $p = 10^{-5}$.

	Low50	Low250	High50	High250
PG	0.30	0.29	0.31	0.28
NMF4-PG	0.52	**0.56**	**0.50**	**0.49**
NMF5-PG	**0.53**	0.51	0.43	0.45
NMF6-PG	0.49	0.46	0.42	0.43
NMF7-PG	0.47	0.46	0.43	0.41

Table 2. F-scores of the Poisson-Gamma Change Point Model when prior of probability of change $p = 10^{-4}$.

3.2. Results

We measure the performance of models using F-measure (F), or the harmonic mean of precision (P) and recall (R) measures. Good performance is achieved when both P and R are close to 1. As the rates of false alarm and/or undetected change points increase, P and R, respectively, gets closer to zero.

$$\text{F-Measure (F)} = 2 \times \frac{P \times R}{P + R} \tag{11}$$

$$\text{Precision (P)} = \frac{\text{\# true alarms}}{\text{\# alarms}} \tag{12}$$

$$\text{Recall (R)} = \frac{\text{\# true alarms}}{\text{\# all changes}} \tag{13}$$

Results are provided in Tables 1 and 2. Tables exhibit the performance with the highest F measure among different settings of p, the prior probability of a change at any time, L, the lag of smoothing and $maxComponent$, the maximum component number for pruning.

Fixed-lag smoothing improves the performance using later observations in a fixed interval due to the information of more observations. Although higher values of L yields more information for inference, we fixed L to an acceptable value as it does not compensate the complexity cost.

The same trade-off comes with the $maxComponent$ as well. Since the complexity of used Bayesian change point models increases linearly with time t, we limit complexity by setting $maxComponent$ to an acceptable value.

Table 1 shows that NMF-MD model with rank 4 is superior to the rest of the models. The same is also valid for the NMF-PG model with rank 4. In addition, the F-scores point that the higher the rank, the lower the performance.

It can be further deduced that when the number of simulated users are increased and the traffic intensity is kept as it is, F-scores tend to decrease. This is justified since changes are more abrupt under low network traffic. Yet, we cannot observe a similar trend when the traffic intensity is altered.

4. Conclusion and Future Work

In this work, we proposed a dictionary learning based anomaly detection method where the network traffic is represented in terms of traffic patterns extracted by nonnegative matrix factorization and the change points are inferred by two Bayesian change point models that differ by their observation models. We tested our methods in a simulated SIP network traffic with DDoS attacks generated by a commercial network vulnerability scanning tool. Our current findings show that the NMF does improve the precision and recall of the change point models, hence the dictionary based network representation is promising for further investigations.

We are currently working on the use of our dictionary learning based method for network attack and user behavior classifications. Since each traffic anomaly is generated by a different combination of traffic templates, the excitation vector output by the NMF can also be used for anomaly classification. Especially, the individual change point detectors in NMF-PG model can be used to signal the type of the anomaly.

In case of a network anomaly, the ideal server behavior is to keep giving service to registered and harmless users while blocking the malicious traffic. We also speculate that the traffic generated by a specific user expressed in terms of known network templates can be used to classify the user as harmless or malicious.

Finally, online NMF inference to learn traffic patterns alongside with change point detection in real time is another possible future work of this study. A nonparametric NMF model may be used to detect new traffics pattern the first time they are observed, and insert them into the dictionary of known templates.

Acknowledgements

This study is a Bogazici University - NETAS collaboration and it is funded with TEYDEB project number 3140701, ”Realization of Anomaly Detection and Prevention with Learning System Architectures, Quality Improvement, High Rate Service Availability and Rich Services in a VoIP Firewall Product”, by the Scientific and Technological Research Council Of Turkey (TUBITAK).

References

Barber, D. and Cemgil, A. T. Graphical models for time-series. *Signal Processing Magazine, IEEE*, 27(6):18–28, 2010.

Cemgil, A. T. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009, 2009.

Chandola, V., Banerjee, A., and Kumar, V. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15:1–15:58, 2009.

Fearnhead, P. Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and computing*, 16(2):203–213, 2006.

Keromytis, A. D. A comprehensive survey of voice over ip security research. *Communications Surveys & Tutorials, IEEE*, 14(2):514–537, 2012.

Kurt, B., Yıldız, Ç., Ceritli, Y. T., Yamaç, M., Semerci, M., Sankur, B., and Cemgil, A. T. A probabilistic sip network simulation system. In *24th IEEE Conference on Signal Processing and Communications Applications (Accepted, in Turkish)*, 2016.

Lee, D. D. and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–91, 1999.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E, Schooler. RFC 3261: SIP: Session Initiation Protocol. Technical report, IETF, 2002. URL www.ietf.org/rfc/rfc3261.txt.

Trixbox. Trixbox. http://www.fonality.com/trixbox, 2016. [Online; accessed 29-April-2016].

V-Spy. Nova V-SPY. http://novacybersecurity.com/en/nova-vspy.html, 2016. [Online; accessed 29-April-2016].

Wang, Haining, Zhang, Danlu, and Shin, Kang G. Change-point monitoring for detection of dos attacks. *IEEE Transactions on Dependable and Secure Computing*, 1: 2004, 2004.

Xiaohong, G., Wang, W., and Zhang, X. Fast intrusion detection based on a non-negative matrix factorization model. *Journal of Network and Computer Applications*, 32(1):31 – 44, 2009.

Yıldız, Ç., Semerci, M., Ceritli, Y. T., Kurt, B., Cemgil, A. T., and Sankur, B. Change point detection for monitoring sip networks, (accepted). In *European Conference on Networks and Communications, EuCNC 2016*, 2016.