SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2014; **7**:904–913
Published online 17 May 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.801

RESEARCH ARTICLE

A data mining system for distributed abnormal event detection in backbone networks

Yingjie Zhou[1,2]*, Guangmin Hu[1] and Dapeng Wu[3]

[1] University of Electronic Science and Technology of China (UESTC), Chengdu, Sichuan 611731, China
[2] Columbia University, U.S.A.
[3] University of Florida, Gainesville, FL 32611, U.S.A.

ABSTRACT

Detecting distributed abnormal events has become an increasingly significant task for efficient network management and operation. However, it is still challenging to uncover these distributed behaviors in backbone networks because of the voluminous amount of noisy, high-dimensional traffic data. In this paper, we present a novel system for detecting distributed abnormal events in backbone networks. The proposed system emphasizes on detecting distributed correlated abnormal events, which are caused by the same reason. In contrast, existing methods are not able to distinguish correlated abnormal events from the independent abnormal events. In our proposed system, a set of data mining techniques is used for modeling and detecting distributed correlated abnormal events by analyzing the traffic features. Specifically, traffic behavior representation is constructed to define and select traffic features for describing the traffic behaviors of interest, feature clustering is performed to group together similar transformations in each feature, behavioral data mining is employed to discover the most significant patterns in network interactions with respect to typical behavior, and behavior classification is used to expose the behaviors of interest. Experiment results using real traffic data present the effectiveness of our proposed methods for detecting distributed correlated abnormal events in the backbone network. Copyright © 2013 John Wiley & Sons, Ltd.

KEYWORDS

distributed correlated abnormal event; anomaly detection; network data mining; unsupervised learning; supervised learning; backbone networks

***Correspondence**
Yingjie Zhou, UESTC, China
E-mail: yjzhou@uestc.edu.cn; yjzhou09@gmail.com

1. INTRODUCTION

As network structures and applications continue to grow in complexity, detecting distributed abnormal events has become an increasingly significant task for efficient network management and operation. On the basis of the real-time detection results, the network administrator could achieve a more comprehensive understanding of the whole network and make optimal response to emergencies. However, it is still challenging for uncovering these distributed behaviors in the backbone networks.

Detecting the distributed abnormal events in backbone networks faces two major challenges: massive data processing and the noisy, high-dimensional nature of traffic data. As the network is oriented toward the large-scale, complex and distributed recently, the data transmitted by Internet increasingly present the characteristics of high-speed data streams, especially in the backbone network. This means that if we want to conduct real-time detection in network-wide range, we need to process with massive data quickly. However, there are thousands of millions of records of traffic information, make it difficult to perform precise user behavior analysis and application behavior analysis in large-scale network. The second challenge of detecting distributed abnormal events is the high-dimensional and noisy nature of network-wide traffic data [1]. This makes it difficult to extract meaningful information about abnormal events from any kind of traffic statistics, especially for those events that span multiple links in a network, and have a low average rate in each link.

To detect the distributed abnormal events from the voluminous amount of noisy, high-dimensional traffic data [2], existing methods use principal components analysis (PCA) [1] or other techniques [3,2,4,5] for distributed abnormal events detection. Although these methods take consideration of distributed abnormal events

in network-wide range, they are not able to distinguish distributed correlated abnormal events, which are caused by the same reason, from independent abnormal events (examples shown in Figure 1).

In this paper, we present a novel data mining system for detecting distributed abnormal events in backbone networks. Different from network-wide traffic anomaly detection that focuses on analyzing the OD flows [3,2,4,5], which needs global traffic matrix statistics and a lot of communication cost, our objective is to detect distributed correlated abnormal events based on the measurements from point of presences (POPs). By finding these abnormal events based on the measurements from POPs, our system is able to know which POP is with respect to special behaviors of interest so that the network administrators could take relevant actions to response. Note that the abnormal events are the emergence of special behaviors of interest; we also regard the abnormal events as anomalies in the following paper.

The contribution of this paper lies in a collection of data mining techniques that are used for modeling and detecting distributed correlated abnormal events by analyzing the traffic features and their multiple abnormal emergencies over logical topology. Traffic behavior representation is first constructed to address the problem of massive data processing by using a summarization tool to define traffic features in a higher semantic level and selecting meaningful traffic features with respect to the traffic behaviors of interest. Second, to predigest the complexity of the following data mining process, we perform feature clustering to group together similar transformations in each feature and create a new symbolic representation of them. Then, to address the high-dimensional and noisy nature of network-wide traffic data, we then employ behavioral data mining to discover the most significant patterns in network interactions with respect to typical behavior. In particular, we present and compare two learning techniques, one supervised method and one unsupervised method, for behavioral data mining. Finally, behavior classification that using the results of behavioral data mining is developed to expose the behaviors of interest.

We implement a prototype of our system and evaluate it using real traffic data collected from an ISP backbone network [6]. The experiment results show that the proposed methods are effective and scalable for detecting distributed correlated abnormal events in the backbone network.

The remainder of the paper is organized as follows. Section 2 introduces the proposed system that facilitates the detection of distributed correlated abnormal events and the data mining techniques for modeling and detecting distributed correlated abnormal events by analyzing the traffic features. The experiment results are presented and discussed in Section 3. Section 4 describes the related work. Section 5 concludes the paper.

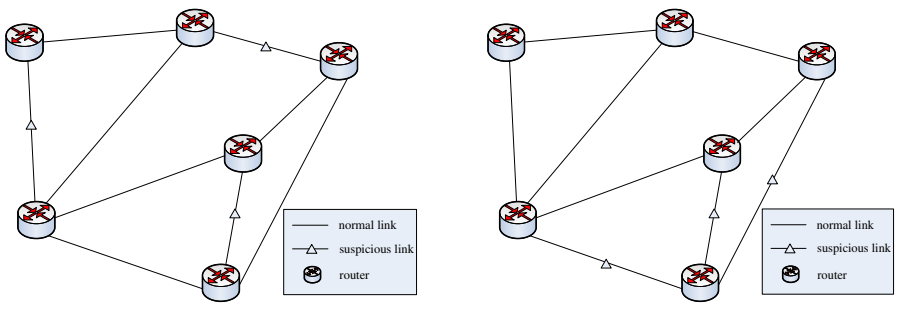
2. PROPOSED DATA MINING SYSTEM

In this section, we present the principles and design of our system that facilitate the detection of distributed correlated abnormal events in backbone networks and the data mining techniques for modeling and detecting them by analyzing the traffic features and their multiple abnormal emergences over logical topology.

2.1. Traffic behavior representation

The objective of traffic behavior representation is twofold. First, we should find some measurable variables to define traffic features from the large amount of network traffic. Second, we should compare and select the most meaningful traffic features that could characterize traffic behaviors.

We use entropy as a summarization tool to express traffic features based on packet header traces collecting from the large amount of network traffic. We regard each attribute in packet header traces as a sequence of random events. Entropy measures the uncertainty of random events. Thus, there is an entropy value for each traffic feature in a time interval; different entropy values over time consist of a time series for corresponding traffic



(a) independent abnormal events (b) distributed correlated abnormal events

Figure 1. Example of independent abnormal events and distributed correlated abnormal events. (a) independent abnormal events and (b) distributed correlated abnormal events.

feature. The entropy measurement is highly sensitive for a wide range of anomalies [2].

The entropy of attribute X is defined as follows:

$$\textbf{\textit{Entropy}}(X) = -\sum_{i=1}^{k} p_i Inp_i \quad (1)$$

where p_i is the frequency of a certain value in attribute X during a time interval, k is the number of possible values and $\sum_{i=1}^{k} p_i = 1$. The entropy expresses the distribution for an attribute. The smaller the entropy is, the more the attribute concentrates.

By investigating the characteristics of common network abnormal events, we consider eight attributes in packet header traces to be the candidates for traffic features that characterize traffic behaviors. They can describe the volume information of flows, the flows' end hosts and services, protocols and more, listed as follows: packet size (PZ, denotes the number of bytes that a packet contains in a flow on average), source IP address (SIP), destination IP address (DIP), source port (SP), destination port (DP), protocol (PROT), type of service (TOS), and TCP flag (TCP_flag). We demonstrate the usage and selection of these features in two steps.

First, we investigate the ability for single attribute to suggest abnormal events. We evaluate this ability by two measurements: (i) percent of anomalies that significantly change on this attribute and (ii) percent of significant changes that are anomalies. The significant change is defined as the change, whose instant changed value is higher than the average. As shown in Table I, we evaluate the ability on a week's flow data from the LOSA node of Abilene [6].

On the basis of the result in Table I, we found that no single attribute is valid for suggesting abnormal events. Both the percent of significant changes that are anomalies and percent of anomalies that significantly change measurements are unacceptable. Thus, it is meaningless to inspect the relationship between single attribute and abnormal events. In this paper, we use behavior patterns, which combine several attributes to characterize traffic behaviors. We show the efficient of using behavior patterns in Section 3.2.

Second, we demonstrate how we select the most meaningful traffic features to characterize traffic behaviors. By using the same data as we investigate the ability for single attribute to suggest abnormal events, we calculate the correlation coefficients to measure the relationship between different traffic features. The correlation matrix is as follows.

	PZ	SIP	DIP	DP	SP	PROT	TOS	TCP_flag
PZ	1.0000	0.1629	0.1625	0.1658	0.1122	0.0962	−0.0111	0.0045
SIP	0.1629	1.0000	0.6770	0.6037	0.3821	0.0845	0.0071	0.0300
DIP	0.1625	0.6770	1.0000	0.5769	0.3620	0.0646	0.0041	0.0359
DP	0.1658	0.6037	0.5769	1.0000	0.4844	0.0902	0.0032	0.0484
SP	0.1122	0.3821	0.3620	0.4844	1.0000	0.0628	−0.0064	0.0574
PROT	0.0962	0.0845	0.0646	0.0902	0.0628	1.0000	0.0066	−0.0013
TOS	−0.0111	0.0071	0.0041	0.0032	−0.0064	0.0066	1.0000	0.0032
TCP_flag	0.0045	0.0300	0.0359	0.0484	0.0574	−0.0013	0.0032	1.0000

From the previous results, we found that the correlations between PROT, TOS or TCP_flag, and traffic features are not obvious, whereas PZ, SIP, DIP, DP, and SP are close related with each other. This means PZ, SIP, DIP, DP, and SP play more important roles in characterizing abnormal traffic behaviors. We refer them as behavior-related variables. In our system, distributed agents are employed to collect packet header traces from different POPs, respectively. Each agent first aggregates the packet header traces and then transforms them into the behavior-related variables. In this way, the traffic behaviors are characterized.

2.2. Feature clustering

In this subsection, we demonstrate two feature clustering techniques to group together similar transformations in each traffic feature. Both of them aim to represent a traffic feature in symbol and are operated on the changed value sequence from each entropy time series. The purpose of conducting feature clustering is to reduce the computational complexity in the following data mining process.

The first feature clustering technique emphasizes on presenting the significant changes in each traffic feature. First, we define a cluster that group together the changed

Table I. Ability for single attribute to suggest abnormal events.

	PZ	SIP	DIP	SP	DP	PROT	TOS	TCP_flag
PASC	0.32	0. 33	0. 39	0. 65	0. 55	0. 40	0. 49	0. 52
PSCA	0.0306	0.0340	0.0388	0.0635	0.0561	0.0403	0.0475	0.0524

PASC, percent of anomalies that significantly change; PSCA, percent of significant changes that are anomalies; PZ, packet size; SIP, source IP address; DIP, destination IP address; SP, source port; DP, destination port; PROT, protocol; TOS, type of service; TCP_flag, TCP flag.

values, which do not represent significant changes of corresponding traffic feature at the time intervals. The instances of this cluster are divided into two bounds. The upper bound is the mean for the changed values, which are larger than 0. The lower bound is the mean for the changed values, which are less than 0. The changed values, which are between the upper bound and the lower bound at the time intervals, are put into the defined cluster; we mark the changed values in this cluster as "C0".

The rest of the changed values are then divided by K-means clustering [7]. K-means clustering method offers an optimized approach to obtain symbolic representation for the entropy sequences of traffic features. We intend to have four value clusters: "C1" is for the changed values, which represent evident increases of corresponding traffic feature at the time intervals, "C2" is for the changed values, which represent surprising increases of corresponding traffic feature at the time intervals, "C3" is for the changed values, which represent evident decreases of corresponding traffic feature at the time intervals, and "C4" is for the changed values, which represent surprising decreases of corresponding traffic feature at the time intervals. However, the clusters "C2" and "C4" may be void when there are no surprising changes. To decide the optimized number of clusters, we use the Davies–Bouldin index (DBI) [8] as a measurement.

In our problem, DBI denotes a set of values, which are calculated by the sum of the standard deviations for arbitrary two clusters over the Euclidean distance between the centroid of the two clusters. There are C(n, 2) values for DBI in total. Denote the max value for DBI as max (*DBI*). The smaller the max (*DBI*), the better the separation of the clusters and the tightness inside the clusters.

By investigating the DBI for the separation of the clusters from 2 to 4, we choose the best number for the separation of the clusters. After conducting the K-means clustering on the changed values, we mark the cluster with symbol. When there are only two clusters, the one whose instants have values that is larger than 0 is marked with "C1"; the other is marked with "C3". When there are three or four clusters, we compute the mean for the values in each cluster. The cluster that has a mean above 0 is marked with "C1" or "C2", whereas the cluster that has a mean less than 0 is marked with "C3" or "C4". In particular, we mark the cluster with a mean that is closer to 0 as "C1" or "C3".

The second feature clustering technique emphasizes on the difference of variation trend in each traffic feature. To the end, we first divide the changed values into two parts: the first part is for the changed values, which represent increases of corresponding traffic feature at the time intervals, and the second part is for the changed values, which represent decreases of corresponding traffic feature at the time intervals. We intend to have four value clusters: "C1" is for the changed values, which represent slight increases of corresponding traffic feature at the time intervals, "C2" is for the changed values, which represent evident increases of corresponding traffic feature at the time intervals, "C3" is for the changed values, which represent slight decreases of corresponding traffic feature at the time intervals, and "C4" is for the changed values, which represent evident decreases of corresponding traffic feature at the time intervals. Similar to the aforementioned situation, the clusters "C2" and "C4" may be void sometimes. We use the same method described in the first feature clustering method to optimize the number for the separation of the clusters, and mark them with symbol.

2.3. Behavioral data mining

Behavioral data mining is employed to discover the most significant patterns in network interactions with respect to typical behavior. In this subsection, we present two learning techniques, one supervised method and one unsupervised method, for behavioral data mining. The supervised method builds time series graphs [9] and unveil the most significant patterns with respect to typical behavior by reference [30], which reduces the dimension of the involved data set. The unsupervised method mines the abnormal communication patterns at local by distributed agents. Both behavioral data mining methods operate on the behavior-related variables, that is, PZ, SIP, DIP, DP, and SP.

2.3.1. Supervised learning

We use the second feature clustering technique to obtain a symbolic representation of the entropy-based behavior-related data. On the basis of the symbolic representation, we could unveil the most significant patterns with respect to typical behavior by date mining. In our previous paper [10], we create time series graphs and mine the closed frequent graph patterns, which represent the most significant patterns with respect to typical behavior, from both the graph set including abnormal behaviors and the graph set without abnormal behaviors. For convenience, we summarize the techniques proposed in our previous paper [10] later in the text.

We first create a graph to describe the behavior-related variables and their relationships at each time interval. In this graph, the behavior-related variables are represented as nodes, which are connected to corresponding routers. The edges connecting nodes to routers denote their subordinate relationships, whereas edges between routers denote their connection relationships in logical topology. In each time interval, there is a graph, and graphs over time constitute time series graphs.

In the time series graphs, typical behaviors could be indicated by a set of frequent graph patterns. These graph patterns may refer to several kinds of behavior-related variables and their corresponding routers. However, when the number of routers increases, the number of the frequent graph patterns grows explosively. To reduce the dimension of the involved data set, we extract the behavior characteristics by closed frequent graph mining. It could make the solution scalable, while most of the meaningful information for frequent graph patterns remains. On the basis of closed frequent graph mining, we could obtain a set of frequent graph patterns from both the graph set including

abnormal behaviors and the graph set without abnormal behaviors, respectively. We use them as the candidates for emerging rules, which we present in Section 2.4.1.

2.3.2. Unsupervised learning

We use the first feature clustering technique to obtain a symbolic representation of the entropy-based behavior-related data. Because most of the anomalies in the communication network [3] emerge in two or three traffic features, we first mine the unusual subpatterns that contain two or three symbols at local, and then decide the abnormal communication patterns by calculating the abnormity coefficient of the merged patterns.

To discover the unusual subpatterns that may indicate abnormal behaviors, we introduce the support count of patterns [11]. Support reflects the frequent degree of a pattern and shows the frequency of the pattern. The smaller the support is, the greater the likelihood that network traffic anomaly may occur. We focus on computing the support count for two kinds of subpatterns. The first one is called 2-itemset subpattern, which reflects the transformations in two selected traffic features and the relationships between them. The other one is called 3-itemset subpattern, which reflects the transformations in three selected traffic features and the relationships among them.

The support of the subpattern is defined as the ratio that the number of the pattern divided by the total number of instances. When the support of a subpattern is less than a given value, we regard it as an unusual subpattern. It is important to realize that the unusual 2-itemset subpattern is more valuable than the unusual 3-itemset subpattern in network traffic anomaly detecting, as it reflects the most direct and basic relationship of multi-time series.

On the basis of the mining results of 2-itemset subpatterns and 3-itemset subpatterns, we could decide the abnormal communication patterns by calculating the abnormity coefficient of the merged patterns. The merged patterns are the combination of all the 2-itemset subpatterns and 3-itemset subpatterns, which are corresponding to the same POP. The abnormity coefficient of a merged pattern is defined as follows.

$$W_t(i) = -\log\left(\Pi\left(\mathrm{sup2}_1 * \mathrm{sup2}_2 * \ldots \mathrm{sup2}_k\right) * \Pi\left(\mathrm{sup3}_1 * \mathrm{sup3}_2 * \ldots \mathrm{sup3}_m\right)\right) \quad (2)$$

where t is the sampling time interval, i is the serial number of POP, $\mathrm{sup2}_i (1 \le i \le k)$ is the unusual 2-itemset subpattern that belongs to the inspected POP, $\mathrm{sup3}_i (1 \le i \le m)$ is the unusual 3-itemset subpattern that belongs to the inspected POP. The greater the abnormity coefficient is, the more possibility that the pattern is an abnormal communication pattern.

2.4. Behavior classification

We present two behavior classification methods for the supervised method and the unsupervised method separately. In particular, the classification method based on emerging patterns is for the supervised method, whereas the other classification method based on voting is for the unsupervised method.

2.4.1. Behavior classification based on emerging patterns

The emerging rule for detecting abnormal behaviors is first proposed in our previous paper [10]. For convenience, we summarize this technique later.

An emerging pattern [12] could be regarded as a rule that distinguishes a certain behavior from others. In particular, the emerging patterns for abnormal behaviors are the patterns that could distinguish the abnormal behaviors from normal ones. They could be selected from the closed frequent graph patterns that we introduce in Section 2.3.1. If a closed frequent graph pattern from the graph set including abnormal behaviors is not frequently observed in the graph set without abnormal behaviors, we call it an emerging pattern or an emerging rule for detecting abnormal behaviors. The support of an emerging rule in a graph set is the frequency that it appears in the graph set. The emerging ratio of an emerging rule for a graph set measures the ability that the emerging rule distinguishes two graph sets. The larger the emerging ratio, the greater the ability.

Whether there is an abnormal event in a certain time interval is determined by an abnormity coefficient. It is calculated from the observation of the emerging rules in the graph set including abnormal behaviors. Different emerging rules have different supports and emerging ratios, thus, contribute to the abnormity coefficient differently. We use the abnormity coefficient to classify the graphs in time intervals to two sets: one is the graph set that includes distributed correlated abnormal events, and the other is the graph set that does not include distributed correlated abnormal events. The larger the abnormity coefficient is, the more possibility that the graph includes distributed correlated abnormal events.

2.4.2. Behavior classification based on voting

On the basis of the abnormal communication patterns described in Section 2.3.2, whether the abnormal communication pattern has a network-wide impact is voted on the basis of the similarity comparison between the inspected POP and its neighbors.

We first specify the traffic features, which are reflected in the abnormal communication pattern. Then, we calculate the correlation coefficients to measure the relationship between the inspected POP and its neighbors on the changed value sequence for the specified traffic features. We could have the same number of correlation coefficients as the number of specified traffic features from the comparison between the inspected POP and each of its neighbors. Note that when a distributed correlated abnormal event occurs, it will cause the relevant behavior-related variables to have similar transformations at different POPs. However, because of the huge background, some of the transformations

may be not obvious. Thus, we select the max value of the correlation coefficients from the comparison between the inspected POP and each of its neighbors.

A vote is operated on the basis of the max values of the correlation coefficients. If the max value of the correlation coefficients is larger than Q, it means that this neighbor agrees that the abnormal communication pattern has a network-wide impact. Otherwise, it means that this neighbor disagrees. Each neighbor has a vote, and the result is measured by the fraction of positive votes that agrees that the abnormal communication pattern has a network-wide impact over the total votes, which we regard as R. From the empirical test, Q is chosen as 0.6, whereas R is specified as 0.5.

3. PERFORMANCE EVALUATION

3.1. Data collection

We implement a prototype of the proposed system and evaluate it using real traffic data collected from Abilene [6], the same data sets as in [2]. Abilene is the Internet2 backbone network, which connects with 200 universities in the U.S.A. It contains nine nodes across the U.S. mainland. We use distributed agents to collect Netflow traces. NetFlow technology is a prevalent IP/MPLS standard in the Internet, which can analyze and measure the IP data flows that pass through network equipment.

We have 2-week flow data, which were from Abilene's IP-level sampling flow data (packets sampling 1/100, cycle sampling at intervals of 5 min) for the period 1 January 2008–14 January 2008. The size of the flow data for analysis in our experiment is 485 GB. Every 5 min, flow statistics constituted the data in a time interval, and there are 288 sampling points for a day.

To obtain the ground truth for abnormal events detected on the basis of the measurements from POPs, we manually inspected the collected data for marking the distributed correlated abnormal events, which is according to the manual inspection introduced in [2]. Because of the enormous work of manual inspection, we have only identified all the DDoS events at present. The data in the first week contains 27 DDoS events, and the data in the next week contains 21 DDoS events. To have a clear background, we remove all other possible anomalies, which are indicated in the following methods. Both the supervised method and the unsupervised method use the data in the second week for detecting, whereas the supervised method uses the data in the first week for training.

3.2. Performance evaluation

The original entropy sequences and the time series of changed values at Internet2 POP ATLT are illustrated in Figure 2. It is shown that the time series of changed values effectively express unusual traffic behavior from the original ones.

The supervised method is performed the same as we did in our previous paper [10]. For convenience, we also shortly summarize it in this paper. The supervised method constructed time series graphs from the changed value sequences. Using the marked data, the time series graphs were divided into two graph sets: the graph set including abnormal behaviors and the graph set without abnormal behaviors. 1051 candidates of rules with the support threshold of 40% are obtained by mining closed frequent graph patterns in the abnormal graph sets. And, the emerging rules are selected from these candidates. The threshold of emerging ratio for valuable rules is specified as 1.0 in this experiment.

The unsupervised method mines the unusual 2-itemset subpatterns and 3-itemset subpatterns at each POP, respectively. The subpattern, whose support count is less than 0.02, is regarded as unusual subpattern. The abnormal communication patterns are identified by the outliers in the abnormity coefficients of the merged patterns. The threshold for voting is the same as specified in Section 2.4.2.

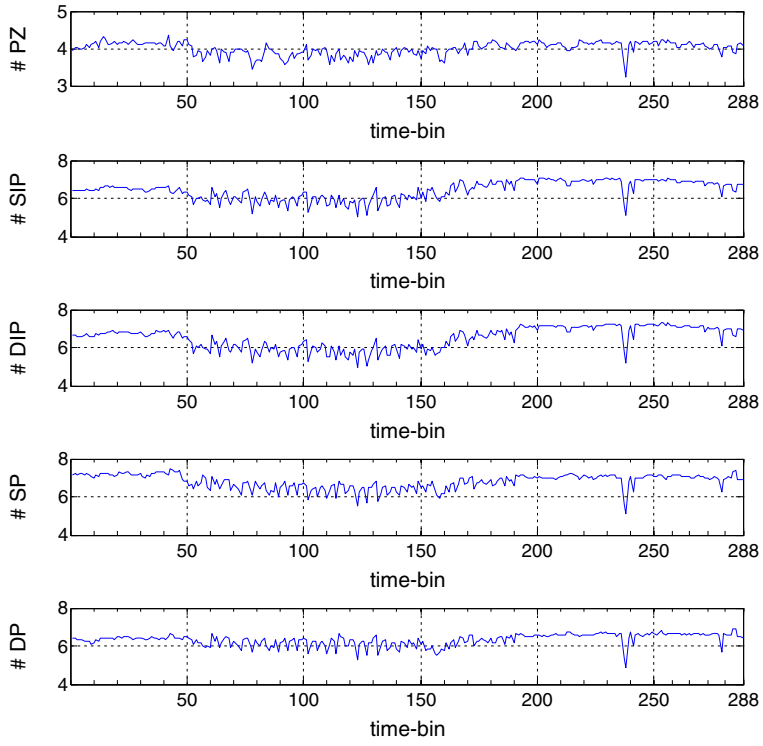
We evaluate the performance of both methods in terms of two metrics: the detection ratio (DR) and the false positive ratio (FPR). DR denotes the fraction of anomalies that are detected among all the anomalies existed, and FPR denotes the fraction of detections that are mistaken as real anomalies among all the detections.

Using the supervised method, 20 DDoS events are detected, and 19 of them are true DDoS events. This result shows that our supervised method has a DR of 90.5%. The FPR of the supervised method is 5%. However, using the unsupervised method, we detected 18 DDoS events, and 16 of them are true DDoS events. This result shows that our unsupervised method has a DR of 76.2% and a FPR of 11.1%.

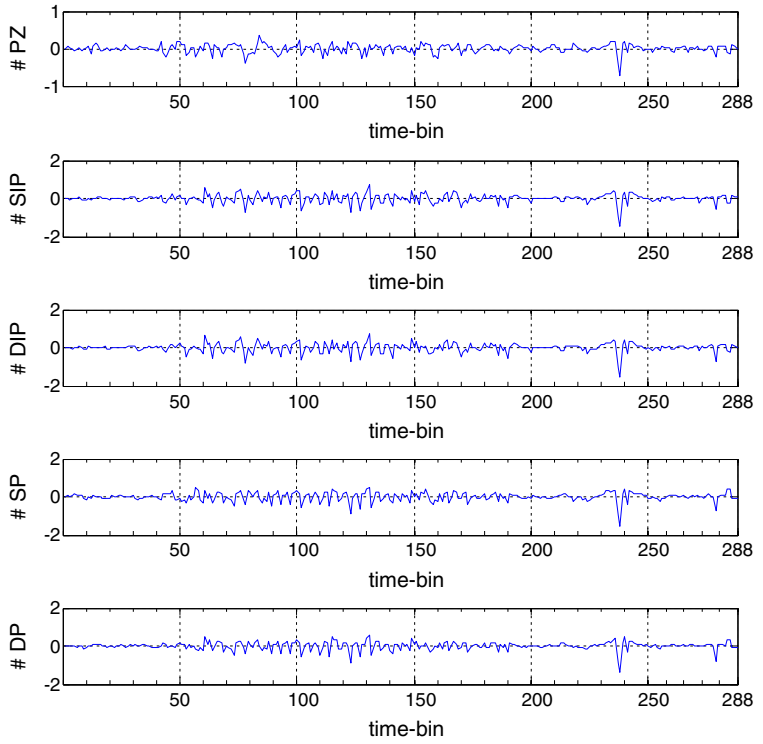
For the purpose of comparison and explanation of our proposed distribution algorithm, we also evaluate a PCA-based detection method [2] as we did in our previous paper [10], which is well-known to detect network-wide anomalies. We list the results for the three methods in Table II as follows. The results illustrated in the table show that both the supervised method and the unsupervised method are superior to the conventional detection method with regard to the FPR. In particular, the supervised method and the unsupervised method decreased the FPR by 15.8% and 9.7% compared with the PCA-based detection method, respectively. This is because the PCA-based detection method could not distinguish correlated abnormal events from independent abnormal events. For the detection ratios, the results of the supervised method and PCA-based detection are comparable. However, the detection ratio of the unsupervised method is lower.

4. RELATED WORK

Most of the related work can be categorized into network disruption diagnosis, network anomaly detection, and network traffic behavior monitoring.



(a) original entropy sequences



(b) time-series of changed values

Figure 2. The original entropy sequences and the time series of changed values at Internet2 core POP ATLT. (a) original entropy sequences and (b) time series of changed values.

Security Comm. Networks 2014; **7**:904–913 © 2013 John Wiley & Sons, Ltd.
DOI: 10.1002/sec

Table II. Detection results compared with PCA-based method.

Supervised		Unsupervised		PCA-based method	
Detection	20	Detection	18	Detection	24
DDoS	19	DDoS	16	DDoS	19
DR	90.5%	DR	76.2%	DR	90.5%
FPR	5%	FPR	11.1%	FPR	20.8%

DR, detection ratio; FPR, false positive ratio.

Network disruption diagnosis

Wu *et al*. [13] proposes an approach for analyzing routing dynamics from routing and traffic data in an IP network to provide alerts for significant disruptions. It designs an online system that converts millions of BGP update messages a day into a few dozen actionable reports about significant routing disruptions. There are several other works [14–16], which also analyze routing dynamics to detect network events that affect a large number of Internet destinations or a large amount of traffic. In [17], Xu *et al*. propose a method based on PCA to obtain a set of clusters from a BGP update stream and identify the underlying events to debug and troubleshoot BGP routing problems. Huang *et al*. [18] propose a method using network-wide analysis of routing information to diagnose network disruptions. It applies a multivariate analysis technique on dynamic routing information and detects every reported disruption with a low rate of false alarms.

Network anomaly detection

Most of the early anomaly detection methods [1,3,19,20] have treated anomalies as deviations in the overall traffic volume (number of bytes or packets). These volume-based anomaly detection methods have been successful in isolating large traffic changes (such as bandwidth DDOS), but fail to identify a large class of anomalies that do not cause detectable disruptions in traffic volume. More sophisticated techniques using the distribution of traffic features are then proposed. Lakhina *et al*. proposed a PCA-based detection method, which separates the high-dimensional space occupied by a set of network traffic measurements into two disjoint subspaces corresponding to normal and anomalous network conditions, and then detects anomalies by analyzing in the anomalous subspace [2]. The main advantage of this approach lies in its use of correlations in time series data from multiple links. Li *et al*. used random aggregations of IP flows rather than origin-destination (OD) flows to detect network anomalies [4]. This improvement enables more precise identification of the underlying causes for network anomalies. Huang *et al*. suggested an approach that avoids excessive use of the network-wide communication in PCA-based detection methods and enables real-time detection [5]. Zhang *et al*. presented a framework that infers network level anomalies from widely available data aggregates [21]. This work also introduces the first anomaly detection algorithm that can handle routing changes and missing data. Duffield *et al*. suggested rule-based anomaly detection on IP flows [22]. It exploits correlations between packet and flow level information, and associates packet level alarms with various features of the flows from the same traffic. Several other methods are introduced to detect traffic anomalies using various features in the network traffic [23–25].

Network traffic behavior monitoring

Most of the research has analyzed specific aspects of traffic or applied metrics to identify significant network events of interest, such as techniques for identifying heavy hitters [26], port scans [27], DDoS [28], and worms [29]. Most related to our work is [31], and it aims to profile the traffic patterns.

Our work is different from the network disruption diagnosis work because it focuses on the network traffic instead of routing dynamics. It is different from the network anomaly detection work because it could only deal with the outliers, and our work could also be applied to the benign events. In other words, there lacks a system to detect a comprehensive set of suspicious behaviors of interest in the backbone networks. Different from the prior work that focuses on specific aspects of network behavior, in this paper, we propose to develop a general behavior identification system to detect distributed correlated abnormal events in backbone networks.

5. CONCLUSION

In this paper, we present a novel system for detecting distributed correlated abnormal events in backbone networks. A set of data mining techniques are used to facilitate the detection of distributed correlated abnormal events. In our proposed system, traffic behavior representation is constructed to define and select traffic features for describing the traffic behaviors of interest, feature clustering is performed to group together similar transformations in each feature, behavioral data mining is employed to discover the most significant patterns in network interactions with respect to typical behavior, and behavior classification is used to expose the behaviors of interest. Experiment results on an ISP backbone network present the effectiveness of our proposed methods for determining distributed correlated abnormal events.

ACKNOWLEDGEMENTS

We would like to thank all the members of this project: Huilan Chen, Yan Liu, Chong Zhang, Hangyu Hu, and Chunyan Xia. We also would like to thank the anonymous reviewers for their constructive comments and suggestions. This research was supported by the National Nature Science Foundation of China (60872033 and 61201127)

REFERENCES

1. Lakhina A, Crovella M, Diot C. Diagnosing network-wide traffic anomalies. In *Proceedings of ACM SIGCOMM*, 2004.
2. Lakhina A, Crovella M, Diot C. Mining anomalies using traffic feature distributions. In *Proceedings of ACM SIGCOMM*, 2005.
3. Lakhina A, Crovella M, Diot C. Characterization of network-wide anomalies in traffic flows(short paper). In *Proceedings of ACM SIGCOMM Conference on Internet Measurement*, 2004.
4. Li X, Bian F, Crovella M, Diot C. Detection and identification of network anomalies using sketch subspaces. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement*, 2006.
5. Huang L, Nguyen XL, Garofalakis M, Hellerstein J, Jordan M, Joseph A, Taft N. Communication-efficient online detection of network-wide anomalies. In *Proceedings of IEEE INFOCOM*, 2007.
6. [Online]. Available: http://www.internet2.edu/network/, 2008.
7. Selim S, Ismail M. K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; **6**(1): 81–87.
8. Davies DL, Bouldin DW. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1979; **1**(2):224–227.
9. Zhou Y, Hu G. GNAED: a data mining framework for network-wide abnormal event detection in backbone networks. In *Proceedings of IEEE International Performance Computing and Communications Conference*(*IPCCC*), 2011.
10. Zhou Y, Hu G. Network-wide anomaly detection based on routers' connecting relationships. *IEICE Transactions on Communications* 2011; **E94.B**(8): 2239–2242.
11. Zhou Y, Hu G, He W. Using graph to detect network traffic anomaly. In *Proceedings of International Conference on Communications, Circuits and Systems* (*ICCCAS*), 2009.
12. Dong G, Li J. Efficient mining of emerging patterns: discovering trends and differences. In *Proceedings of ACM SIGKDD*, 1999.
13. Wu J, Mao Z, Rexford J, Wang J. Finding a needle in a haystack: pinpointing significant BGP routing changes in an IP network. In *Proceedings of USENIX Symposium on Networked Systems Design and Implementation* (*NSDI*), 2005.
14. Teixeira R, Rexford J. A measurement framework for pin-pointing routing changes. In *Proceedings of ACM SIGCOMM Workshop on Network Troubleshooting*, 2004.
15. Feldmann A, Maennel O, Mao ZM, Berger A, Maggs B. Locating Internet routing instabilities. In *Proceedings of ACM SIGCOMM*, 2004.
16. Caesar M, Subramanian L, Katz R. Towards localizing root causes of BGP dynamics. *Technical Report UCB/CSD-04-1302*, 2003. [Online]. Available: http://www.cs.uiuc.edu/~caesar/papers/rootcause.pdf
17. Xu K, Chandrashekar J, Zhang Z-L. A first step to understand inter domain routing dynamics. In *Proceedings of ACM SIGCOMM Workshop on Mining Network Data (MineNet)*, 2005.
18. Huang Y, Feamster N, Lakhina A, Xu J. Diagnosing network disruptions with network-wide analysis. In *Proceedings of ACM SIGMETRICS*, 2007.
19. Roughan M, Griffin T, Mao ZM, Greenberg A, Freeman B. Combining routing and traffic data for detection of IP forwarding anomalies. In *Proceedings of Joint International Conference on Measurement and Modeling of Computer Systems*, 2004.
20. Barford P, Kline J, Plonka D, Ron A. A signal analysis of network traffic anomalies. In *Proceedings of ACM SIGCOMM Workshop on Internet Measurement*, 2002.
21. Zhang Y, Ge Z, Greenberg A, Roughan M. Network anomography. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement*, 2005.
22. Duffield N, Haffner P, Krishnamurthy B, Ringberg H. Rule-based anomaly detection on IP flows. In *Proceedings of IEEE INFOCOM*, 2009.
23. Ahmed T, Coates M, Lakhina A. Multivariate online anomaly detection using kernel recursive least squares. In *Proceedings of IEEE INFOCOM*, 2007.
24. Chhabra P, Scott C, Kolaczyk E, Crovella M. Distributed spatial anomaly detection. In *Proceedings of IEEE INFOCOM*, 2008.
25. Kline J, Nam S, Barford P, Plonka D, Ron A. Traffic anomaly detection at fine time scales with bayes nets. In *Proceedings of International Conference on Internet Monitoring and Protection* (*ICIMP*), 2008.
26. Krishnamurthy B, Sen S, Zhang Y, Chen Y. Sketch-based change detection: methods, evaluation, and

applications. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement*, 2003.
27. Jung J, Paxson V, Berger A, Balakrishna H. Fast portscan detection using sequential hypothesis testing. In *Proceedings of IEEE Symposium on Security and Privacy*(*S&P*), 2004.
28. Lu K, Wu D, Fan J, Todorovic S, Nucci A. Robust and efficient detection of DDoS attacks for large-scale Internet. *Computer Networks* 2007; **51**(18): 5036–5056.
29. Kim HA, Karp B. Autograph: toward automated, distributed worm signature detection. In *Proceedings of ACM Security Technical Program*, 2004.
30. Yan X, Han J. Closegraph: mining closed frequent graph patterns. In *Proceedings of ACM SIGKDD*, 2003.
31. Xu K, Zhang Z, Bhattacharyya S. Internet traffic behavior profiling for network security monitoring. *IEEE/ACM Transactions on Networking* 2008; **16**(6): 1241–1252.