Characterising Anomalous Events using Change - Point Correlation on Unsolicited Network Traffic

Ejaz Ahmed, Andrew Clark and George Mohay

Information Security Institute,
Queensland University of Technology, Brisbane, Australia
{e.ahmed, a.clark, g.mohay}@qut.edu.au

Abstract. Monitoring unused or dark IP addresses offers opportunities to extract useful information about both on-going and new attack patterns. In recent years, different techniques have been used to analyze such traffic including sequential analysis where a change in traffic behavior, for example change in mean, is used as an indication of malicious activity. Change points themselves say little about detected change; further data processing is necessary for the extraction of useful information and to identify the exact cause of the detected change which is limited due to the size and nature of observed traffic. In this paper, we address the problem of analyzing a large volume of such traffic by correlating change points identified in different traffic parameters. The significance of the proposed technique is two-fold. Firstly, automatic extraction of information related to change points by correlating change points detected across multiple traffic parameters. Secondly, validation of the detected change point by the simultaneous presence of another change point in a different parameter. Using a real network trace collected from unused IP addresses, we demonstrate that the proposed technique enables us to not only validate the change point but also extract useful information about the causes of change points.

1 Introduction

Different techniques [1,2,3] have been proposed to monitor network traffic for malicious content. One compelling technique is to monitor unused IP addresses spaces [3,4,5,6] as this traffic has no reason to exist. Due to the absence of any legitimate activity associated with such addresses, the traffic observed is a result of different abnormal activities like traffic from hosts infected by worms, traffic generated by network probing tools or viruses, traffic from misconfigured nodes and backscatter traffic from distributed denial of service attacks. This provides an added advantage to network security researchers as the analysis is not complicated or distracted by complex, hard to analyze, legitimate traffic. While monitoring such addresses has been used for forensics [7,8,9,10,11,12] and attack detection [13,14,15], the analysis of traffic is largely a manual process. In this paper, we address this limitation of manual root cause identification and validation by correlating change points among different network traffic parameters. We do this by monitoring multiple network traffic parameters in parallel and applying change detection techniques on each of the monitored parameter separately.

Several characteristics can be used to define normal network behavior including traffic dynamics such as the type of traffic, volume of traffic and delay experienced by the network. One of the key observations in differentiating between normal and anomalous activity is the change in traffic dynamics. During normal operations traffic dynamics usually remain constant or vary slowly over time whereas during malicious activity they no longer remain relatively constant [16,17]. Detection of malicious activities can thus be considered as a change detection problem, detecting change in traffic parameter as quickly as possible [16].

In change detection, there is a sequence of observations whose statistical properties changes at some unknown point in time and the goal is to detect these changes as soon as possible. Usually extraction of information related to the change point is done by manual analysis which is not only time consuming but also limited due to the nature and size of the collected data.

In this paper, correlation of change points among different traffic parameters is used to automatically extract the information related to the anomaly. The significance of the proposed technique is two-fold. Firstly, automatic extraction of information related to change points by correlating change points detected across multiple traffic parameters; and secondly, validation of detected change point by the simultaneous presence of another change point in a different parameter.

The paper is organized as follows. Section 2 provides an overview of the related work and details our contributions. A brief overview of the change detection technique is provided in Section 3. Section 4 provides detail of the proposed change point correlation technique. Section 5 details the data being used in this paper and explains the traffic parameter being considered. Experimental results are provided in Section 6. Finally a conclusion and future directions are given in Section 7.

2 Related Work

The idea to use change detection techniques for detecting different anomalous behavior has existed for some time. Change detection has been used in analyzing network traffic to identify different traffic anomalies including worm detection [18,19], denial of service and distributed denial of service attack detection [20,21,22,23] , scan detection [17] and anomaly/fault detection [24,25,26]. Due to its simplicity and effectiveness, change detection has also been used in the analysis of unwanted traffic collected from unused IP addresses. In this regard Bu et al. [27] proposed detecting a worm outbreak based on a change in the inter-arrival time of packets sent by scanners. The authors have used a CUSUM change detection algorithm on packet inter arrival times as the first step in detecting a worm outbreak. The alarm from CUSUM algorithm is then analyzed in a second stage where a maximum likelihood estimation (MLE) is used to confirm the epidemic.

Limthong et al. [28] used the discrete wavelet transformation (DWT) technique to identify anomalies in unused IP addresses. The authors have analyzed three different types of packets namely TCP SYN, TCP SYN/ACK and UDP packets. The focus of the paper was to identify a suitable measurement interval and to analyze different DWT levels. On the other hand no information about detected anomalies is provided.

Ahmed et al. [14] used a nonparametric CUSUM change detection technique to identify unusual behavior in darknet traffic. The authors proposed a sliding window based memory management technique to identify changes in traffic behavior. The validity of the proposed technique had been tested using both a synthetic data set and real network traces collected from an unused IP address block. In this paper the authors have only validated their proposed technique using UDP traces collected from monitored data. In another study [15], the authors have also proposed detection of "nested" anomalous activities i.e. the commencement of a new session of anomalous activity whilst another anomalous activity is occurring .

Although change detection has received considerable attention in recent years, correlation between different change points has not been addressed. This might be because most of the research in this area makes use of a single parameter. However, multivariate and multichannel [29,30] change detection techniques have been proposed in the literature, where multiple traffic parameters are considered for detecting a change, the basic objective is to detect a change in any traffic parameter to raise an alarm.

Ide et al. [31] have used change point correlation to compare multiple time series from dynamic systems. The authors have used a singular spectrum transformation (SST) technique to identify change points in the time series. The similarity between different time series is obtained by visual inspection of change point scores. The focus of our method proposed in this paper is different as it correlates the change points identified in different traffic parameter time series to validate and automatically extract useful information related to change points identified in primary traffic parameter time series.

Use of change point correlation to identify problems in enterprise middle ware systems was proposed by Agarwal et al. [32]. The authors have used change point correlations (simultaneous occurrence of events in different time series in a given window) to identify problems associated with middle ware architectures. The authors have used the difference of means to identify changes in system time series. In order to reduce false alarms the authors have used a pre-selected threshold to identify a change along with pre-defined signatures for possible problems associated with the system. For the analysis of unsolicited traffic such as that observed while monitoring large unused address spaces, the correct formulation of signatures for various attacks is not possible. The lack of attack signatures, such as for the zero day attacks, limits the use of the proposed technique in analyzing such traffic. In addition, as mentioned by the authors, the effectiveness of the proposed technique largely depends upon the window size over which the difference of mean is calculated. We argue that the use of a fixed size window has two serious drawbacks, first it introduces a delay in detection and secondly the effect of a huge change in parameter behavior will effect the detection of subsequent behavioral changes at least for the duration of the window size. This makes it unsuitable for analysis of malicious network traffic and was addressed by Ahmed et al. [14].

The work most closely related to our work is of Kaplan et al. [33]. The authors have used change point correlation to identify simultaneous activities from multiple brain areas. The basic idea is to treat two change points in different parameters as correlated if thy appear close in time. Although our work is inspired by this work, there is a significant difference between the two approaches both in the change detection technique and data used. Firstly we used the change detection technique presented in [15]. Secondly our objective is to automatically extract information related to change points identified in traffic collected from unused IP addresses, whereas Kaplan et al. [33] have applied this technique on EEG data.

The focus of the technique proposed in this paper is to automatically extract information related to a change point identified in traffic targeting unused IP addresses. In our framework, information extraction is realized by a two stage process: change point detection; and change point correlation. In the first stage the change detection technique described in [15] is applied to the different time series collected from the unused IP address block. In the second stage the correlation among different change points is carried out to automatically extract the information related to the detected change point.

The novelty of the proposed technique lies in its ability to automatically extract information related to change points by correlating change points across different traffic parameters. In addition the proposed correlation based technique can also be used to validate change points. The effectiveness of the proposed technique is validated using a 17 month real network trace collected from a dedicated block of unused IP addresses.

3 Change Detection Technique

For change detection, a nonparametric CUSUM technique provided in [15] have been used. In this section a brief overview of the change detection technique will be provided, for a detailed explanation the interested reader is referred to [15].

Let $N_1, N_2, ..., N_n$ be the sequence of observations from the monitoring system at fixed time instances $t_{1,}t_{2,}....., t_n$. A malicious activity at time t_k will result in the change in statistical properties of the observed traffic parameter. Let μ_k be the mean traffic rate before and $\bar{X_k}$ be the mean traffic rate after a new anomalous activity, then the CUSUM score, S_k , can be calculated by Equation 1,

$$S_k = \max\left\{0, S_{k-1} + N_k - \mu_k - \alpha.\bar{X_k}\right\} \tag{1}$$

where α is a tuning parameter belonging to the interval (0,1) and is considered to be an upper bound on the estimated post-change traffic rate $\bar{X_k}$.

The choice of the tuning parameter, α, effects the performance of the algorithm as selecting too small or too large an α value will increase the false alarm rate. The α value can either be replaced with a constant value based on experimentation or can be calculated dynamically using the Equation 2:

$$0 \leq \alpha \leq 1 - \frac{\mu_k + h_k/T}{\bar{X_k}} \tag{2}$$

where h_k is the detection threshold at time t_k and T is the maximum time in which the change should be detected.

The average number of packers at time t_k can be estimated iteratively using an exponential weighted moving average (EWMA) given in Equation 3:

$$\bar{X_k} = (1-\beta).\bar{X}_{k-1} + \beta.N_k \tag{3}$$

where $0 < \beta < 1$ is a smoothing factor which gives more weight to the current observation.

The CUSUM score, S_k, given in Equation is then subject to a threshold test h in order to detect a change, S_k value greater or equal than the threshold will indicate a change in parameter properties. The threshold value h at time t_k is given in Equation 4:

$$h_k^{start} = \sigma_{k-1},\ h_k^{end} = 0.25.h_k^{start} \tag{4}$$

where h_k^{start} and h_k^{end} are the threshold values for the start and the end of the change point at time t_k and σ_{k-1} is the standard deviation of the data elements in current window at time t_k . In addition, to reduce the false alarm rate, an additional counter τ is used along with h_k^{end} to mark the end of a change point, the alarm is not canceled until timer τ reaches a specified value, see Equation 5.

$$Alarm = if(S_k \leq h_k^{end} AND \tau \geq 2, terminate, resume) \tag{5}$$

4 Change Point Correlation

In analyzing traffic, including that collected by monitoring unused IP addresses, the volume and nature of the observed traffic makes manual analysis of the traffic related to a change point a difficult and time consuming task. This is mainly because such traffic is usually unlabeled and largely malicious which makes every single change point worthy of further investigation.

To overcome this limitation, information extraction related to a change point can be specified as a problem of identifying the simultaneous occurrence of two or more change points in different traffic parameters. The change points in multiple parameters are considered as correlated if the time difference between them does not exceed a time threshold. The simultaneous occurrence of these change points in a given time window can then be used to automate the extraction of information related to a change point.

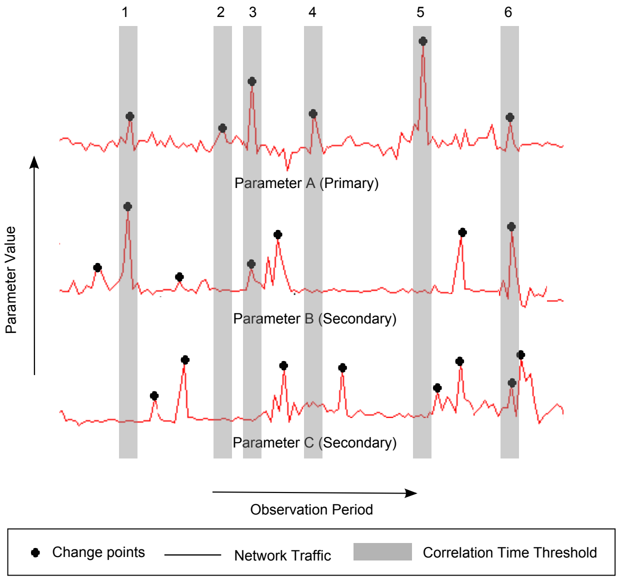


Fig. 1. Correlating change points in different traffic parameters.

Change points in different traffic parameters are considered to be correlated only if they appear simultaneously within a certain time threshold. This is illustrated in Figure 1 which shows three parameters namely A, B and C. The horizontal axis is the observation period and vertical axis is the measured parameter value. For information extraction, the total number of traffic parameters are divided in two categories namely primary and secondary parameters. For analysis only one primary parameter is selected whereas there can be any number of secondary parameters. The choice of the primary and the secondary parameter is arbitrary and largely depends upon type of network traffic and required analysis. For example the number of packets can be a good candidate for a primary parameter and number of unique sources and number of unique destination ports can both be secondary parameters. In Figure 1, parameter A is a primary while parameters B and C are the secondary parameters.

For change point correlation, each detected change point in the primary parameter is surrounded by a correlation time threshold window, represented by the vertical strips in Figure 1. The change points in the secondary parameters are considered to be correlated with the primary parameter if they fall within the time threshold window. In the above figure the first change point in the primary parameter A is correlated with the change point identified in secondary parameter B, whereas no change point within the time threshold window was identified in secondary parameter C. The close proximity of these change points can be used to explain and characterize the change identified in primary parameter A.

As a single change point can cover more than one time interval, the time threshold window can not be fixed and is calculated as a function of change point duration. Let t_s and t_e be the start and end of a change point then the correlation time threshold window for this change point can be calculated as

$$t_{tr}^{start} = t_s - n \quad , \quad t_{tr}^{end} = t_e + n \tag{6}$$

where t_{tr}^{start} and t_{tr}^{end} is the start and end of correlation time threshold window respectively and n is the correlation threshold allowance having value between 0 and m. The above equation will take n measurement points to the immediate left and n measurement points to the immediate right of the change point identified in primary parameter for correlation analysis.

4.1 Correlation Logic

One of the first questions that arises when correlating change points identified in different network traffic parameters is how to deal with change points identified in different secondary network traffic parameters, with a different change start and end time when compared with t_{tr}^{start} and t_{tr}^{end}. Let $\bar{t}_s$ and $\bar{t}_e$ be the time instances at which the change starts and ends in the secondary parameter. Then for the cases given in Equation 7, the change in the secondary parameter will not cover the correlation time threshold window.

$$(\bar{t_e} \geq t_{tr}^{end} > \bar{t_s} > t_{tr}^{start}) \parallel (t_{tr}^{end} > \bar{t_e} > \bar{t_s} \geq t_{tr}^{start}) \parallel (t_{tr}^{end} > \bar{t_e} > t_{tr}^{start} \geq \bar{t_s}) \tag{7}$$

Keeping in mind the above situations, the correlation logic needs to be enhanced in several ways when considering the duration and the significance of a change point in a secondary parameter with respect to a change in the primary parameter. For information extraction and validation of a change point in the primary parameter it is expected that a change in a secondary parameter will be consistent with the change in the primary parameter. This is given in Equation 8.

$$(\bar{t_e} \geq t_{tr}^{end} > t_{tr}^{start} \geq \bar{t_s}) \tag{8}$$

A simultaneous occurrence of change in both primary and secondary parameters, according to Equation 8, can be used as the key to evaluate these changes. In addition, the significance of the change identified in a secondary parameter can be considered as the magnitude of the change, in the primary parameter, explained by the secondary parameter. This requires that the corresponding change in the primary parameter exceed a certain threshold to be considered significant and can be specified as:

$$(Secondary \quad Parameter)/(Primary \quad Parameter) > X \tag{9}$$

where X is the threshold of the proportion of the change specifying how much of the change in the primary parameter can be explained by the change in a specific secondary parameter. Since we aim to use correlation for information extraction and validation of a change in the primary parameter, the choice of X will have a significant effect on the correlation. Selecting a large X value means only the secondary parameter which have a significant influence on the primary parameter will be considered in correlation. On the other hand, selecting a small X value will consider secondary parameters with little or no effect on the primary parameter as being correlated, which might not be useful in identifying the cause of a change in the primary parameter.

Although the duration and significance of the change in a secondary parameter gives higher confidence in the extracted information related to the change in primary parameter, the changes in secondary parameters with cases in Equation 7 cannot be ignored. In such cases the change in the secondary parameter is considered correlated with the change in the primary parameter only if the change in the secondary parameter is significant.

Different values for X can be used to correlate changes in secondary and primary parameters. A higher X value for changes in the secondary parameter with a duration according to Equation 7, and lower value if the duration is consistent with a change in the primary parameter (see Equation 8) can be used for correlating changes identified in the primary and secondary parameters. We use $X = 0.3$ where the change in the secondary parameter continues at least for the duration of the change in the primary parameter and $X = 0.5$ where the change in the secondary parameter is within the correlation time threshold window but does not continue for the duration of the change in the primary parameter.

4.2 Change Point Validation

The use of correlation among simultaneous changes in primary and secondary network traffic parameters includes a sort of additional validation of a non-nested (single) and nested change point identified in the primary parameter. These are taken into account only if at least one of the secondary parameters changes simultaneously within the given time threshold window.

In Figure 1 for change points 4 and 5, no change in the secondary parameter was identified within the time threshold window. Therefore the changes 4 and 5 in the primary parameter A cannot be validated and hence can be treated as false alarms. Validation of nested change points identified in the primary parameter is confirmed either by the presence of a nested change, or identification of a new change in at least one of the secondary parameters. This changed behavior in secondary parameters is used to confirm the identification of a nested change point in the primary parameter.

The occurrence of false change points, in the present context, does not effect the analysis as the traffic is by definition unsolicited and is either opportunistic or malicious. Every single packet appearing on the unused IP address blocks is unsolicited and we aim to understand what caused it. . The simplified algorithm of change point correlation is given in Figure 2

Algorithm: Change Point Correlation

```
for each parameter do
    compute change points
end for
for each change point in primary parameter do
    calculate $t_{tr}^{start}$ and $t_{tr}^{end}$ of the change point
    $correlation\_flag = 0$
        for each secondary parameter ($i$) do
            if $t_e^i \geq t_{tr}^{end}$ and $t_s^i \leq t_{tr}^{start}$ then
                if (secondary parameter)/(primary parameter) $> 0.3$
                    change point correlated, $correlation\_flag = 1$
                end if
            else if $(t_{tr}^{end} \geq t_s^i > t_{tr}^{start})$ or $(t_{tr}^{end} > t_e^i \geq t_{tr}^{start})$
                if (secondary parameter)/(primary parameter) $> 0.5$
                    change point correlated, $correlation\_flag = 1$
                end if
            end if
        end for
end for
if $correlation\_flag == 1$
    change point validated
end if
```

Fig. 2. Change point correlation.

5 Experimentation

To evaluate the effectiveness of the proposed method, we performed experiments on a real data set collected from a dedicated block of unused IP addresses. The data was collected for a period of about 20 months between 27 November 2006 and 10th August 2008. The monitoring system consists of a single class C address block. For experimentation, the size of the sliding window is set to 100 [14,15]. For correlation, the correlation time threshold allowance (n) is set to 1. Other parameters such as change detection threshold (h_k^{start}), change end threshold (h_k^{end}) and upper bound on mean (α) will be calculated dynamically.

In this experimentation we first categorized the packets based on their type into either TCP, UDP or ICMP. Then for each category different traffic parameters, that have the potential to explain the detected change in the primary parameter, have been extracted. Table 1 provides a summary of these parameters. Even though there are dependencies among these parameters, this does not effect our analysis as we aim to use these parameters in order to understand what causes a change in traffic behavior. Each of these parameters are measured at a regular time interval. First change points were identified in the individual traffic parameter time series using the technique described in Section 3. Change point correlation was then used to validate and automatically extract the causes of the change point in the primary parameter using the technique discussed in Section 4. We now provide a discussion of the results.

Parameter	Type	Description
N_{pt}^{k}	Primary	Total number of packets of type *pt* received during k^{th}time interval
S_{pt}^{k}	Secondary	Total number of unique source IP addresses sending packets of type *pt* during k^{th}time interval
P_{pt}^{k}	Secondary	Total number of unique destination ports receiving packets of type *pt* during k^{th}time interval
Y_{pt}^{k}	Secondary	Total number of unique payloads in packets of type *pt* during k^{th}time interval
s_{pt}^{k}	Secondary	Total number of packets of type *pt* from busiest source address during k^{th}time interval
d_{pt}^{k}	Secondary	Total number of packets of type *pt* received by busiest destination address during k^{th}time interval
p_{pt}^{k}	Secondary	Total number of packets of type *pt* received by busiest destination port during k^{th}time interval
$P_{pt}^{k,w}$	Secondary	Total number of packets of type *pt* received by well known destination ports, (0 to 1023), during k^{th}time interval
$P_{pt}^{k,r}$	Secondary	Total number of packets of type *pt* received by registered destination ports, (1024 to 49151), during k^{th}time interval
$P_{pt}^{k,d}$	Secondary	Total number of packets of type *pt* received by dynamic/private destination ports, (49152 to 65353), during k^{th}time interval

Table 1. Network traffic parameters.

6 Analysis Results

To evaluate the proposed approach, we have tested it on 20 months of real data collected from a dedicated unused IP address block, class C. The monitored architecture is a passive monitoring, contains no active component and no response is generated, of 256 unused IP addresses. Although the proposed technique has been tested on TCP, UDP and ICMP traffic, we will focus our discussion on UDP analysis. In the absence of any active component, due to the passive nature of the monitoring system, UDP being connectionless protocol is well suited for such an analysis.

Table 2 summarizes the change points identified in the measured traffic parameters for 20 months of the collected traffic. The total number of change points detected in each parameter is given in the total column, this includes both nested and non-nested change points. The nested column provides the number of nested change points identified in each parameter. The last column lists the number of correlated change points. As described in Section 4 for the primary parameter, the total number of UDP packets per unit of time, the correlated column lists the change points for which simultaneous change points were identified in at least one of the secondary parameters. For secondary parameters the correlated column gives the number of change points occurring simultaneously with the primary parameter.

Parameter	Number of Change Points		
	Total	Nested	Correlated
N^k_{udp}	45	12	45
S^k_{udp}	22	5	14
P^k_{udp}	31	10	15
Y^k_{udp}	23	4	10
s^k_{udp}	31	6	19
d^k_{udp}	47	13	13
p^k_{udp}	58	6	38
$P^{k,w}_{udp}$	62	5	19
$P^{k,r}_{udp}$	59	13	21
$P^{k,d}_{udp}$	48	7	6

Table 2. Summary of the change points identified in UDP traffic.

The change points identified in the primary parameter are shown in Figure 3. In this figure the top graph shows the number of UDP packets, primary parameter, observed on the Darknet and the bottom graph shows the change points identified. The vertical axis in the bottom graph represents the outcome of the change detection algorithm with values greater than 1 representing the number of nested changes detected by the change detection algorithm presented in Section 3.

Figure 4 shows the result of correlation of change points identified in primary and some of the secondary parameters for the period of about 9 weeks from 27 November 2006 to 31 February 2007. In this graph only three secondary parameters are shown due to the space limitations, but note that the analysis is performed on all 9 secondary parameters for the period of about 20 months. In this graph the horizontal axis represents time in days, left vertical axis represents number of UDP packets and right vertical axis represents the outcome of change detection algorithm (dotted lines). The correlation time window is shown by the vertical gray portions.

In figure 4 , the first change point in primary parameter is correlated with change point in p^k_{udp}, where k is 30/11/2006. It is also correlated with $P^{k,w}_{udp}$, not shown in the above figure. This helps in not only automatically extracting useful information related to change point but also validates the change point identified in the primary parameter. The observed change point was due to an increase in the traffic on destination port 137, more than 75 % of the total UDP traffic was targeted on this port. This port is used by NETBIOS name service and also by Trojan Msinit. The source ports used by these sources were either 137 or 1024+n which confirms the existence of worm looking for unprotected network shares.

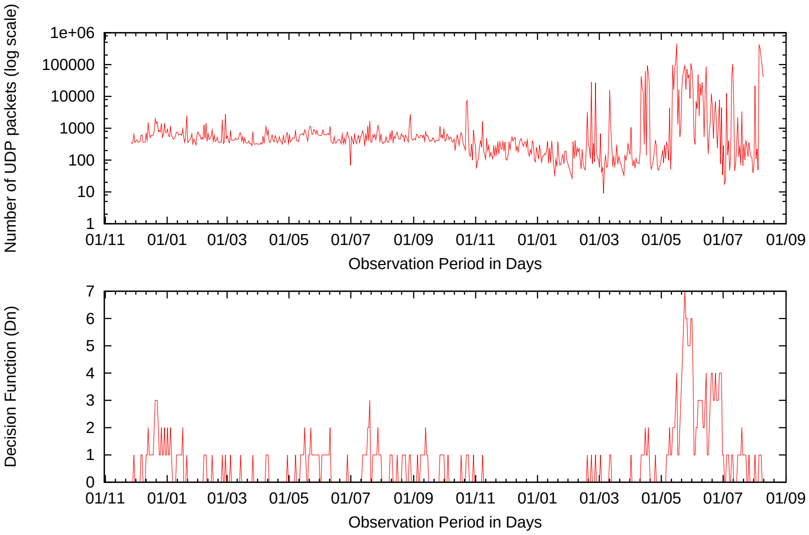
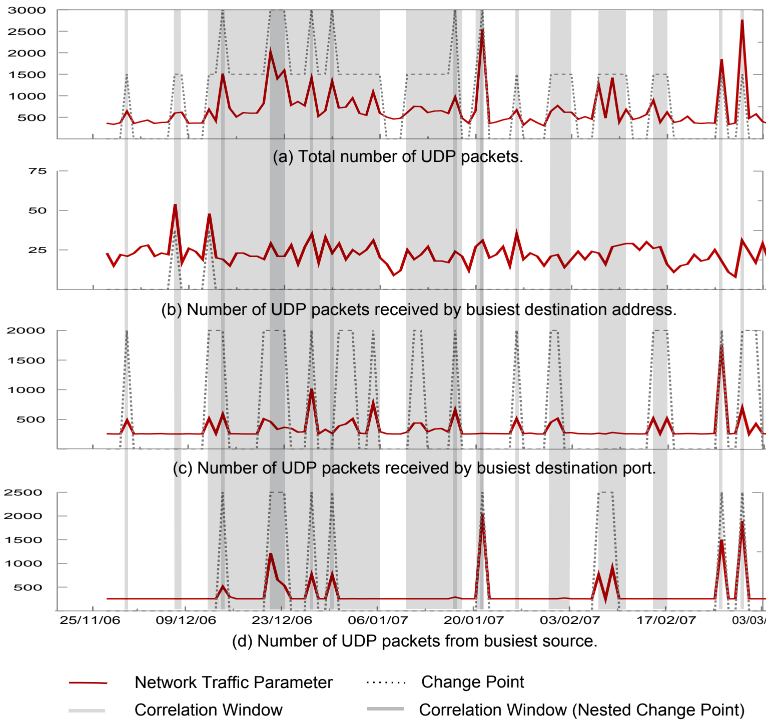


Fig. 3. Change points detected in primary parameter (UDP traffic).

The second change point was observed on 7/12/2006 which continued for two days. This change point was correlated with two secondary parameters d^k_{udp} and $P^{k,r}_{udp}$, not shown in the above figure. According to the correlation algorithm described in Section 4 the change point identified in d^k_{udp} is not significant (less than 30%) and thus can be discarded. On the other hand the change point identified in $P^{k,r}_{udp}$ is not only significant but also correlated with change in primary parameter. This change point was due to an increase in the UDP traffic on four destination ports 1030-1033 and 1434 collectively receiving more than 50% of the total UDP traffic. During this change point no significant activity either by specific source/destination address or on specific destination port was observed which was confirmed by the lack of change point in these secondary parameters.

The first nested change point in N^k_{udp} was observed from 12/12/2006 to 5/01/2007. Extracting useful information related to this change point using current correlation algorithm is a difficult task due to the complex correlation of change points observed in secondary parameters. During this period change points in 6 secondary parameters were identified with no change continued for the whole duration. Currently analysis of such complex correlation patterns is not considered and is part of our future work.



(a) Total number of UDP packets.

(b) Number of UDP packets received by busiest destination address.

(c) Number of UDP packets received by busiest destination port.

(d) Number of UDP packets from busiest source.

Fig. 4. Change point correlation (UDP traffic).

The second nested change point was observed from 11/01/2007 to 17/01/2007. During this period nested change was observed on the last day. This change point was correlated with change points in p^k_{udp} and $P^{k,r}_{udp}$. Observe that correlated nested change point was identified in both p^k_{udp} and $P^{k,r}_{udp}$, both observe nested change point on 17/01/2007. The simultaneous occurrence of these nested change points in secondary parameters, p^k_{udp} and $P^{k,r}_{udp}$, validates the nested change point identified in primary parameter N^k_{udp}. This change was due to the increased MS SQL slammer activity on port 1434 which received more than 60% of the total UDP traffic. The nested change was also due to an increase in the traffic on the same port. Even though the same destination port was targeted in both changes, the simultaneous occurrence of nested changes in primary and secondary parameters alerts to the significant change in UDP traffic on port 1434 during pre-nested and nested changes, traffic on port 1434 is almost doubled during nested change.

An interesting correlation was observed for change point identified in primary parameter on 12/03/2008 which continued for the next day. During this period 8 out of 9 secondary parameters were observed to have correlated change point with primary parameter. In addition nested change point was observed in both $P^{k,w}_{udp}$and p^k_{udp} whereas no nested change point was observed in the primary parameter. This was due to the fact that the change detection algorithm only identifies nested change if there is a significant increase in the parameter value. Although there was significant increase in both $P^{k,w}_{udp}$ and p^{k}_{udp}, an increase in the primary parameter was not observed. In fact the primary parameter was decreased from over 15000 UDP packets to just 1086 packets during that time. In addition the change point in the other 4 secondary parameters was only for one day and was not observed on the second day of the change in primary parameter. During the analysis it was observed that the change on 12/03/2008 was due to the increased activity on a single destination port 13276 on destination address x.x.x.221 which was targeted by the large number of sources. This behavior was not observed on the second day of the change which was confirmed by the lack of change point in S^{k}_{udp}, Y^{k}_{udp}, s^{k}_{udp} and $P^{k,r}_{udp}$.

Although the proposed technique is limited to only simple correlations, our analysis showed some encouraging results. Using the proposed technique we were able to automatically extract and validate more than 60% of the change points, including both non-nested and nested changes, identified in the primary parameter.

7 Conclusion and Future Directions

In this paper, we have proposed a technique for correlating change points among different traffic parameters in order to automatically extract the information related to the anomalous event. The motivation behind our work was to to automate the extraction of information related to change points and to validate the detected change points by correlating change points in primary and different secondary traffic parameters. The applicability and usability of the proposed algorithm is analyzed with the help of real network traces collected from dedicated block of unused IP addresses. It is observed that the proposed technique is not only helpful in automatically extracting information related to detected changes but also can be used to validate the identified changes at the first place.

Even though the proposed technique is limited to only simple correlations, the preliminary results are indeed encouraging. We believe that the algorithm can be improved to detect and validate complex correlations and is part of our future work. In addition we aim to use the proposed technique in automatically generating signatures related to detected change points which can then be used to detect similar behaviors in the future.

References

1. Gregory B. White, Eric A. Fisch, and Udo W. Pooch. *Computer System and Network Security.* CRC Press, 1995.
2. Xuxian Jiang and Dongyan Xu. Collapsar: a vm-based architecture for network attack detention center. In *Proceedings of the 13th conference on USENIX Security Symposium*, pages 2–2, Berkeley, CA, USA, 2004. USENIX Association.

3. Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson. The internet motion sensor: A distributed blackhole monitoring system. In *Proceedings of Network and Distributed System Security Symposium*, pages 167–179, 2005.
4. Uli Harder, Matthew Johnson, Jeremy T. Bradley, and William J. Knottenbelt. Observing internet worm and virus attacks with a small network telescope. In *Proceedings of the 2nd Workshop on Practical Applications of Stochastic Modelling*, pages 113–126, June 2005.
5. David Moore, Colleen Shannon, Geoffrey M. Voelker, and Stefan Savage. Network telescopes: Technical report. Technical Report tr-2004-04, July 2004.
6. Nicolas Vanderavero, Xavier Brouckaert, Olivier Bonaventure, and Baudouin L. Charlier. The honeytank: a scalable approach to collect malicious Internet traffic. *International Journal of Critical Infrastructures*, 4(1):185–205, 2008.
7. David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. Inside the slammer worm. *IEEE Security & Privacy Magazine*, 1(4):33–39, 2003.
8. Moheeb A. Rajab, Fabian Monrose, and Andreas Terzis. Worm evolution tracking via timing analysis. *Proceedings of the ACM workshop on Rapid malcode*, pages 52–59, 2005.
9. David Moore, Geoffrey M. Voelker, and Stefan Savage. Inferring Internet denial-of-service activity. *Proceedings of the USENIX Security Symposium*, pages 9–22, 2001.
10. Abhishek Kumar, Vern Paxon, and Nicholas Weaver. Exploiting underlying structure for detailed reconstruction of an internet-scale event. In *Proceedings of the ACM Internet Monitoring Conference*, 2005.
11. Paul Barford, Rob Nowak, Rebecca Willett, and Vinod Yegneswaran. Toward a model for source addresses of internet background radiation. In *Proceedings of the Passive and Active Measurement Conference of Stochastic Modelling*, March 2006.
12. Kensuke Fukuda, Toshio Hirotsu, Osamu Akashi, and Toshiharu Sugawara. Correlation among piecewise unwanted traffic time series. In *Proceeding of the IEEE Global Telecommunications Conference*, pages 1–5, December 2008.
13. Sohraab Soltani, Syed Ali Khayam, and Hayder Radha. Detecting malware outbreaks using a statistical model of blackhole traffic. In *Proceeding of the IEEE International Conference on Communications*, pages 1593–1597, May 2008.
14. Ejaz Ahmed, Andrew Clark, and George Mohay. A novel sliding window based change detection algorithm for asymmetric traffic. In *Proceedings of the IFIP International Conference on Network and Parallel Computing*, pages 168–175, Oct. 2008.
15. Ejaz Ahmed, Andrew Clark, and George Mohay. Effective change detection in large repositories of unsolicited traffic. In *Proceedings of the Fourth International Conference on Internet Monitoring and Protection*, May 2009.
16. Michele Basseville and Igor V. Nikiforov. *Detection of abrupt changes: theory and application.* Englewood Cliffs, NJ: Prentice Hall, 1993.
17. Jaeyeon Jung, V. Paxson, A.W. Berger, and H. Balakrishnan. Fast portscan detection using sequential hypothesis testing. *Proceedings of IEEE Symposium on Security and Privacy*, pages 211–225, May 2004.
18. Jeffrey Chan, Christopher Leckie, and Tao Peng. Hitlist worm detection using source IP address history. In *Proceedings of Australian Telecommunication Networks and Applications Conference*, 2006.

19. Chen Bo, Bin-Xing Fang, and Xiao-Chun Yun. A new approach for early detection of internet worms based on connection degree. In *Proceedings of International Conference on Machine Learning and Cybernetics*, volume 4, pages 2424–2430 Vol. 4, Aug. 2005.
20. Haining Wang, Danlu Zhang, and Kang G. Shin. Change-point monitoring for the detection of DoS attacks. *IEEE Transactions on Dependable and Secure Computing*, 1(4):193–208, 2004.
21. Wei Chen and Dit-Yan Yeung. Defending against tcp syn flooding attacks under different types of ip spoofing. In *Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies*, pages 38–38, April 2006.
22. Alexander G. Tartakovsky, Boris L. Rozovskii, Rudolf B. Blazek, and Hongjoong Kim. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing*, 54(9):3372–3382, Sept. 2006.
23. Vasilios A. Siris and Fotini Papagalou. Application of anomaly detection algorithms for detecting SYN flooding attacks. *Computer Communications*, 29(9):1433–1442, 2006.
24. M. Thottan and Chuanyi Ji. Anomaly detection in IP networks. *IEEE Transactions on Signal Processing*, 51(8):2191–2204, Aug. 2003.
25. H. Hajji. Statistical analysis of network traffic for adaptive faults detection. *IEEE Transactions on Neural Networks*, 16(5):1053–1063, Sept. 2005.
26. Hassan Hajji. Baselining network traffic and online faults detection. In *IEEE International Conference on Communications*, volume 1, pages 301–308, May 2003.
27. Tian Bu, Aiyou Chen, S. Vander Wiel, and Thomas Woo. Design and evaluation of a fast and robust worm detection algorithm. In *Proceedings of the 25th IEEE International Conference on Computer Communications*, pages 1–12, April 2006.
28. Kriangkrai Limthong, Fukuda Kensuke, and Pirawat Watanapongse. Wavelet-based unwanted traffic time series analysis. In *International Conference on Computer and Electrical Engineering*, pages 445–449, Dec. 2008.
29. A.G. Tartakovsky and V.V. Veeravalli. An efficient sequential procedure for detecting changes in multichannel and distributed systems. In *Proceedings of the Fifth International Conference on Information Fusion*, volume 1, pages 41–48 vol.1, 2002.
30. Osman Salem, Sandrine Vaton, and Annie Gravey. A novel approach for anomaly detection over high-speed networks. In *Proceedings of the European Conference on Computer Network Defense*. INFO - Dépt. Informatique (Institut TELECOM ; TELECOM Bretagne), October 2007.
31. Tsuyoshi Idé and Keisuke Inoue. Knowledge discovery from heterogeneous dynamic systems using change-point correlations. In *Proceedings of the SIAM international conference on data mining*, 2005.
32. Manoj K. Agarwal, Manish Gupta, Vijay Mann, Narendran Sachindran, Nikos Anerousis, and Lily Mummert. Problem determination in enterprise middleware systems using change point correlation of time series data. In *Proceedings of the 10th IEEE/IFIP Network Operations and Management Symposium*, pages 471–482, April 2006.
33. Alexander Ya. Kaplan and Sergei L. Shishkin. Application of the change-point analysis to the investigation of the brain electrical activity. *B.E.Brodsky, B.S.Darkhovsky. Nonparametric Statistical Diagnosis: Problems and Methods*, pages 333–388, 2000.