IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 14, 2017, accepted March 15, 2017, date of publication April 4, 2017, date of current version August 8, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2689001

Anomaly Detection Based on LRD Behavior Analysis of Decomposed Control and Data Planes Network Traffic Using SOSS and FARIMA Models

BASIL AsSADHAN[1], KHAN ZEB[1], JALAL AL-MUHTADI[2], AND SALEH ALSHEBEILI[3]

[1]Department of Electrical Engineering, College of Engineering, and Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh 11421, Saudi Arabia
[2]Center of Excellence in Information Assurance (CoEIA), and Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11653, Saudi Arabia
[3]Department of Electrical Engineering, College of Engineering, and KACST-TIC in RF and Photonics for the e-Society (RFTONICS), King Saud University, Riyadh 11421, Saudi Arabia

Corresponding Authors: Basil AsSadhan (bsadhan@ksu.edu.sa); Khan Zeb (ksailkhan@ksu.edu.sa)

This work was supported by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Saudi Arabia under Award 10-INF1279-02.

ABSTRACT The detection of anomalies in network traffic, such as low volume attacks and abnormalities, has become a pressing problem in today's large volume of Internet traffic. To this end, various anomaly detection techniques have been developed, including techniques based on long-range dependence (LRD) behavior estimation of network traffic. However, the existing LRD-based techniques analyze the aggregated WHOLE (control plus data) traffic, which might not be sufficient to detect short-duration and low-volume attacks and abnormalities in the traffic. This is because such anomalies might pass unnoticed in large volume of the normal background traffic. To address this issue, we propose a method that examines the LRD behavior of control and data planes traffic separately, which improves the detection efficacy. For LRD behavior analysis, the proposed method integrates the correlation structures of second-order self-similar and fractional autoregressive integrated moving average models. The performance of the proposed method is empirically evaluated and validated over a relatively recent real Internet traffic captured at King Saud University's network. The analysis and results demonstrate that the proposed method efficiently detects such low volume and short duration attacks and abnormalities in the traffic, which would not be detected by merely analyzing the aggregated WHOLE traffic without decomposing it into control and data planes traffic.

INDEX TERMS Anomaly detection, intrusion detection, Internet traffic, LRD, self-similarity, network traffic analysis, network security, control plane traffic, data plane traffic, SOSS model, FARIMA model.

I. INTRODUCTION

The volume of Internet traffic is persistently growing due to continuous emergence of new technologies and high bandwidth applications, and the pervasiveness of its users. According to Cisco global IP traffic forecast [1], the annual global IP traffic has exceeded the Zettabyte (ZB)[1] figure by the end of 2016. With such humungous growth, this figure will reach 2.3 ZB per year by 2020. This persistent and rapid growth of Internet traffic brings many challenging issues, including the detection of anomalies.

Anomalies can be defined as some patterns in data that do not imitate network traffic normal behavior. Network operators frequently face a wide range of such patterns in network traffic. Anomalous patterns could be benign abnormalities due to technical or physical problems, such as network outage, high-rate flows and sudden changes due to flash crowds [5]. On the other hand, they could be due to malicious illegitimate activities, for example, cyber intrusions, distributed denial of service (DDoS) attacks, worm propagation, port scanning, credit card frauds etc. [2], [3], which could lead to catastrophic consequences and threaten the proper operation of networks.

These malicious activities are also growing with time, which is evident from various surveys and reports. For example, according to [4], attacks on network infrastructures, data centers and customers, particularly DDoS attacks, are increasing in size, complexity, and frequency. DDoS attacks have increased round about sixty times in size in the past

[1]$1\text{ZB} = 10^{21}$ Bytes

eleven years, i.e., from 8 Gbps eleven years ago to 500 Gbps in 2015. Moreover, the frequency of DDoS attacks has escalated in 2015 with larger volumetric attacks as compared to 2014 [4]. DDoS attacks cause anomalies in network traffic by consuming and exhausting victims' resources. In order to overcome such abnormalities and attacks in today's humongous network traffic, the development of robust, accurate, real-time and efficient detection techniques is inevitable, which is a challenging and open research problem yet to be solved by the network community.

To this end, numerous methods have been developed for the detection of abnormalities and attacks in network traffic. Overall, these methods lie in two domains: signature based detection and anomaly based detection. In the former, the detection methods are based on predefined signatures, which are built on the basis of the characteristics and features of previously known attacks. Thus, such signatures are used for the detection of the corresponding attacks. While in the latter case, the detection methods are based on certain underlying models, which are labeled based on traffic normal behavior. The methods in this group detect and characterize anomalous patterns in the traffic based on deviation from the underlying labeled models. Once the traffic is characterized anomalous, then the corresponding data is investigated more closely to identify the root cause of the anomaly. Moreover, anomaly based detection techniques can detect zero-day attacks, i.e., previously unknown attacks, whereas signature based methods cannot. For this reason, anomaly based detection methods have gathered much attention of researchers over the last few decades.

Various anomaly based detection methods have been proposed in the literature based on a wide range of models, including the self-similar and long-range dependence (LRD) nature of Internet traffic, which has been introduced more than two decades ago by Leland *et al.* [6]. It is shown in [6] that self-similarity and LRD behavior are ubiquitous in aggregated network traffic. This nature of the network traffic is used as an underlying model for the detection of any possible anomalies, which is an efficient way for volume based anomaly detection. In such anomaly detection methods, the aggregated network traffic is monitored for its self-similar and LRD behavior, where deviation from this behavior indicates an anomaly in the traffic. However, the existing techniques in this domain analyze LRD behavior in the aggregated WHOLE (control plus data) traffic. In this way, short duration and low volume attacks and abnormalities might not be detected. This is because they might be buried under the large volume of normal traffic, which might not affect the overall statistical characteristics of the aggregated WHOLE traffic considering the LRD behavior. Besides, many of the attacks and abnormalities are established and carried out in the control plane traffic [7], [8]. Consequently, due to the small volume of control plane traffic compare to the WHOLE traffic, such attacks and abnormalities might be skipped by merely looking at the aggregated WHOLE traffic.

To address this issue, we extend our preliminary work in [9], where we only presented basic results on packet count analysis, which are extended and further explicated in this paper. We propose a robust LRD based anomaly detection method that analyzes the LRD behavior of aggregated control and data planes traffic separately using an integrated LRD model, which is based on the integration of the correlation structures of second order self-similar (SOSS) and fractional autoregressive integrated moving average (FARIMA) LRD models. In addition, the proposed method further narrows down the LRD analysis of traffic for anomaly detection by further splitting the control and data planes traffic in different directions with respect to the enterprise network. Hence, the proposed method is comparatively more robust and efficient considering the large volume of Internet traffic. Moreover, for comparative analysis, we also analyze the aggregated WHOLE traffic along with the aggregated decomposed traffic. The main contributions of this paper include:

1) LRD behavior analysis of decomposed network traffic through byte count feature along with packet count feature and their cross comparative analysis.
2) Experimental evaluation and validation of the proposed anomaly detection method using recent real normal and anomalous Internet traffic traces.
3) Comparison between the results of the proposed method and the results of analyzing the aggregated WHOLE traffic.
4) Theoretical and experimental analysis of LRD behavior variation with the variations in traffic volume, aggregation interval and window size.
5) Empirical identification and evaluation of an appropriate aggregation interval, which is required to aggregate the network traffic, and a time window size, which is required for the analysis of LRD behavior of the corresponding aggregated traffic.

The rest of this paper is organized as follows; Section II briefly describes the phenomenon of self-similarity and LRD. It also provides the details of LRD estimation in Internet traffic and the role of aggregation interval and window size in LRD behavior analysis. Section III provides a thorough literature review of the related work. In Section IV, we present and explain the proposed methodology. Section V presents the experimental results and discussion. This section includes the description of KSU's dataset, the results on the selection of appropriate aggregation interval and window size, and the performance evaluation and validation of the proposed method using KSU's dataset. Lastly, Section VI presents the concluding remarks and future directions.

II. BACKGROUND

A. SELF-SIMILARITY AND LONG-RANGE DEPENDENCE (LRD)

Generally, a process that statistically looks the same irrespective of scaling in time or space is known as self-similar process. Kolmogorov first identified self-similar processes in 1941 [10]. The well-known example of self-similar

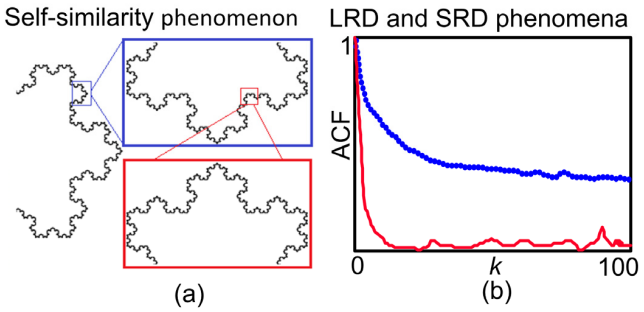

FIGURE 1. The phenomena of self-similarity and LRD: (a) Koch curve (b) LRD and SRD.

processes is fractal. The notion of fractal arises from the fractal geometry. When the fractal is considered for different resolutions, it shows the characteristic of similarity. Figure 1 (a) depicts the phenomenon of self-similarity in a well-known Koch curve [11]. If we zoom the Koch curve with different degrees of magnifications, it looks the same. The most well-known statistical example of self-similar processes is second-order self-similar (SOSS) process.

On the other hand, LRD is a property of a process that shows strong statistical dependence over large time lags. In other words, a process whose autocorrelation function decays very slowly (i.e., hyperbolically) and the process depends on long past values, such process exhibits LRD behavior. This implies that all the values at any time are correlated in a non-negligible way with values at future instant. In contrast, if the autocorrelation function of a process decays very fast (i.e., exponentially), then such process is short-range dependence (SRD). In such process, the dependence does not live for a very long time. Figure 1 (b) depicts the phenomena of LRD and SRD. In the plot, the blue dotted curve with hyperbolic decay represents a typical LRD process, while the solid red curve with abrupt decay represents a typical SRD process. The variable k represents the corresponding lag values of the autocorrelation function.

The phenomena of LRD and self-similarity are closely related with subtle differences. LRD describes the statistical significant correlations of the behavior of a time-dependent process across large time scales, while self-similarity defines the phenomenon in which the behavior of a process is preserved irrespective of scaling in space or time [12]. These phenomena in a process can be studied through a well-known Hurst (H) parameter, which is emerged after the work of the British hydrologist H. E. Hurst on the Nile river minima in 1951 [13].

Mathematically, a continuous process $Y(t)$ is considered self-similar if it satisfies the condition,

$$Y(t) =_d a^{-H} Y(at), \forall a > 0, t \geq 0 \; and \; 0 < H < 1,$$

where H is the corresponding Hurst parameter, a is a scaling parameter, and the notation $=_d$ represents the equality in distribution [14], [15]. This means that the normalized time scaled version $Y(at)$ of $Y(t)$ with a normalization factor of a^{-H} is equal in distribution to $Y(t)$. Where H can take any real value from the interval $0 < H < 1$. However, negative values of H are not acceptable because this will cause the process $Y(t)$ to be no longer a measurable process. Similarly, H with a value of zero is of no interest because it implies that for all values of t greater than zero, $Y(t) = Y(at)$, with probability of one. Moreover, when $Y(t)$ has a finite variance and stationary increments, then the incremental process of $Y(t)$ can be evaluated as follows: $X_i = Y_i - Y_{i-1}$ $(i = 1, 2, 3 \ldots)$. Where X_i (in our study represents the time series of packet and byte counts) has an autocorrelation function given by [16];

$$\rho_H(k) = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right), k \geq 1. \quad (1)$$

Any process with an autocorrelation function of (1) is referred to as an SOSS process. The aggregated process $X^{(m)}(k)$ of X_i has the following form [18], [19]:

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i), k = 1, 2, 3, \ldots,$$

where m is the level of aggregation. This implies that the original series X_i is divided into non-overlapping blocks of size m and averaged over each block. The index, k, identify the block. It can be shown that for all integers k,

$$X(i) =_d m^{1-H} X^{(m)}(i) \quad (2)$$

A stationary process $X(i)$ is exactly SOSS if it satisfies (2) for all m. This means that $m^{1-H}X^{(m)}$ has the same second order statistical properties, i.e., variance and autocorrelation, as X, for all m. Conversely, a stationary process is called asymptotically SOSS if (2) holds as $m \to \infty$. This implies $m^{1-H}X^{(m)}$ has the same second order statistics, i.e., variance and autocorrelation, as that of X, when $m \to \infty$ [17], [18].

Correspondingly, if X_i exhibits LRD behavior, then its hyperbolically decaying autocorrelation function can be represented as follows;

$$\rho_\beta(k) \sim k^{-\beta}, \; ask \to \infty, \; \beta \in (0, 1) \quad (3)$$

where $\beta = 2 - 2H$, is the corresponding LRD parameter, which measures the degree of LRD in X_i. Equation (3) implies that the autocorrelation function of LRD process decays slowly, therefore, the absolute sum of its values approaches infinity, i.e., $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$. On the contrary, the autocorrelation function of an SRD process decays fast such that $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$ [16], [19]. In case of LRD, H takes values between 0.5 and 1, whereas for SRD, its values are less than 0.5. It is worth noticing that in the LRD behavior analysis, H cannot take values of 1, greater than 1, and 0.5. Because H with a value of 1 implies that $\rho_H(k) = 1$ for all k. Likewise, the value of H greater than 1 is of no use because it contradicts the fact that $|\rho_H(k)| \leq 1$ for all k. Similarly, H with a value of 0.5 shows an uncorrelated process.

Furthermore, the most well-known models of LRD processes are fractional Gaussian noise, i.e., which represent SOSS process [18], [20], and fractional autoregressive

integrated moving average (FARIMA) process [21]. Any FARIMA (*p, d, q*) process can be represented using the standard FARIMA (*0, d, 0*) [19]. The parameter d determines the LRD behavior, parameters p and q allow for more flexible modelling of the short-range properties. The parameter d can take real values from the interval $(-1/2, 1/2)$. A FARIMA (*0, d, 0*) process exhibits LRD behavior if and only if $d \in (0, 1/2)$. The autocorrelation function of SOSS process is given in (1) and the autocorrelation function of the standard FARIMA (*0, d, 0*) process has the following form,

$$\rho_d(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(1+k-d)} = \prod_{i=1}^{k} \frac{(k-i+d)}{(k-i+1-d)} \tag{4}$$

where the parameter $d = 1/2 - \beta/2$ determines the LRD behavior and $\Gamma(.)$ is Euler's gamma function.

Additionally, the LRD behavior has been observed and analyzed in various fields, for instance, econometrics, hydrology, biophysics, linguistics, earth sciences, and Internet traffic modeling [22]–[24]. About two decades ago, Leland *et al.* [6] introduced the notion of self-similarity and LRD in aggregated network traffic. Later on, it was found that self-similarity and LRD behavior also occur in WAN traffic [25], WWW traffic [26], and protocol level traffic [24], [27]. This self-similar and LRD nature of network traffic is mainly due to the multiplexing of a large number of ON/OFF sources that transfer files whose sizes are heavy-tailed [26], [28]. Moreover, the LRD behavior in network traffic is analyzed through various features including packet count, byte count, packet inter-arrival time, and flow count [6], [29], [30].

B. LRD ESTIMATION IN NETWORK TRAFFIC

For accurate capturing and forecasting of LRD behavior in Internet traffic, the selection of a proper LRD estimation tool and other related parameters such as aggregation interval and window size are crucial. Here we explain and discuss these parameters and LRD estimation tool that we use in this work.

An aggregation interval is the time duration over which the number of packets/bytes are counted, which results in a time series of packets/bytes counts' bins. A window size is the time-window over which the analysis of LRD behavior of the corresponding aggregated traffic in the form of time series of packets/bytes' bins is carried out. The selection of values for these parameters depend on the traffic rate, the higher the traffic rate, the smaller the aggregation interval and window size should be and vice versa. Since the traffic rate changes over time, therefore, different traffic behaviors occur at different time scales [12], [31], [32]. The traffic rate during daytime is higher than the traffic rate during nighttime. Similarly, the traffic rate during working hours of weekdays is higher than the traffic rate during weekends. This is because there will be more active sources (users) in the network during daytime, particularly working hours as compared to weekend and nighttime. In addition, the LRD phenomenon in network traffic depends on the number of ON/OFF sources to be multiplexed. Therefore, the observations of the degree of LRD behavior of network traffic varies with the variation of scale (aggregation interval). Consequently, we need to adjust the scale according to the traffic rate.

Moreover, it is presented in [24] that the LRD behavior in network traffic depends on the network utilization, i.e., when utilization is low, e.g., overnight, then the degree of LRD decreases, and vice versa. Thus in order to accurately capture the LRD behavior in traffic overnight, during weekend and during non-busy hours of the day, the identification of distinct aggregation intervals is required for every individual case based on the number of active sources in the network as opposed to the case of traffic during working hours. Since during these times, the networks have less number of active sources to be multiplexed, and hence less volume of traffic as compared to busy hours on working days where the traffic is more bursty in nature. Such cases induce false alarms in the detection of anomalies in Internet traffic based on LRD behavior analysis (i.e., declaring that the traffic under analysis deviates from LRD behavior, which indicates an anomaly in the traffic, when in fact such event occurs due to the inadequate traffic and not an anomaly). This is one of the limitations of the LRD based anomaly detection schemes. Therefore, such cases either need to be avoided while analyzing the LRD behavior or an appropriate aggregation interval should be identified for each individual case based on the number of active sources in the network, which could be achieved by developing an adaptive method for the aggregation of traffic.

Similarly, the proper selection of window size depends on the intensity of underlying traffic. According to [33], it is necessary to identify the minimum required window size to obtain reliable LRD measurements. The LRD measurements become less accurate when using small window sizes. It causes more false alarms, as the failure to exhibit LRD behavior would be due insufficient data and not due to an anomaly. On the other hand, if the window size gets larger, then it could lead to miss detection of short duration and low volume anomalies. Because they might be buried under the normal traffic, hence, might not be detected. Likewise, larger window size could slow down the analysis. Besides, there is the possibility that we might come across different types of non-stationarities, particularly the trends i.e., upward or downward trends in the traffic, which could mislead the LRD behavior's results. Since according to [12] and [23], the definition of LRD is based on stationarity, and all LRD estimation techniques assume a stationary time-series; therefore, their estimates are quite sensitive to the existence of non-stationarities in the analysis. Experiments in [33] and [34] show that window sizes of 15–30 minutes are practical and sufficient for the LRD analysis in modern LANs Ethernet traffic.

Furthermore, the accurate detection and estimation of LRD behavior in network traffic is also highly dependent on the LRD estimation tools/methods. Among several LRD estimation methods, the simplest and more accurate method to test

whether the underlying traffic exhibits LRD behavior or not is to test whether it follows SOSS/FARIMA LRD model or not. For this reason, in the proposed method, we incorporate the Kettani and Gubner's LRD estimator known as the ''Optimization Method'' (OM) [16], [35], which is based on the aforementon models. OM is shown in the literature, for example in [36] and [37], to be among the well-known LRD behavior estimation methods, which is considered as a more simple, accurate, and faster.

The OM uses either SOSS model or FARIMA (*0, d, 0*) model for H parameter estimation in any given process. Let X_i represent the process of number of packets or bytes in the i^{th} interval of a given trace of data. The OM first tests whether X_i fits any of the aforementioned models, and if so, then it gives an estimation of the LRD parameter $\hat{\beta}$ i.e., the H parameter in case of SOSS model, and the d parameter in case of FARIMA model. OM evaluates this fitting of data to the underlying model through a Curve-Fitting Error (CFE) function, $E_K\left(\hat{\beta}\right)$, given by:

$$E_K\left(\hat{\beta}\right) = \frac{1}{4K}\sum_{k=1}^{K}\left(\rho_{\hat{\beta}}(k) - \hat{\rho}(k)\right)^2 \tag{5}$$

$\rho_{\hat{\beta}}(k)$ represents the autocorrelation function of the underlying model (i.e., SOSS or FARIMA (*0, d, 0*)), which is given in (1) and (4), respectively. Whereas $\hat{\rho}(k)$ is the sample autocorrelation function of the data to be analyzed, and K is the largest possible value of lags k that minimizes the edge effect in calculation of $\hat{\rho}(k)$. The sample autocorrelation function, $\hat{\rho}(k)$, can be estimated by normalizing the estimated autocovariance function, $\hat{C}_X(k)$ by the estimated variance, $\hat{\sigma}_X^2$ of the process X_i as follows:

$$\hat{\rho}(k) = \frac{\hat{C}_X(k)}{\hat{\sigma}_X^2} \tag{6}$$

The autocovariance function, $\hat{C}_X(k)$ is estimated by:

$$\hat{C}_X(k) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{\mu}_X\right)\left(X_{i+k} - \hat{\mu}_X\right)$$

where $\hat{\mu}_X$ is the sample mean of the process and can be estimated as follows: $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i$. The estimated variance of the process is evaluated as follows: $\hat{\sigma}_X^2 = E\left[\left(X_i - \hat{\mu}_X\right)^2\right]$, which is equal to $\hat{C}_X(0)$. Therefore, (6) can be written as:

$$\hat{\rho}(k) = \frac{\hat{C}_X(k)}{\hat{C}_X(0)}$$

In the proposed method, both SOSS and FARIMA models are integrated concurrently, which are explained in Section IV.

III. RELATED WORK

Since the discovery of self-similar and long-range dependence (LRD) nature of aggregated network traffic [6], the concept of self-similarity and LRD has initiated studies in several areas, including the detection of anomalies in network traffic [37]–[57]. Network traffic has been shown to be self-similar and exhibit LRD behavior under normal conditions. However, anomalous conditions, such as congestion, devices failure, and cyber intrusions can cause loss of LRD behavior. Consequently, this loss of self-similar and LRD behavior can be used to detect such anomalous events. In [37] and [38], it is demonstrated that normal Internet traffic in the absence of anomalies preserves second order self-similar (SOSS) property, while anomalous traffic distract from this structure of normal traffic. Thus, the anomalous traffic can be detected by monitoring the SOSS property of the traffic. Mirosław and Dymora in [39] present that the Hurst (H) parameter estimation could be used to detect anomalies in certain types of traffic, such as HTTP, E-mail, and SSL. In this method, the corresponding H values are compared with the confidence interval of normal values for the detection of any possible anomaly in the traffic. On the other hand, Inacio *et al.* [43] show that the degree of self-similarity increases during an attack, which may be used to suspect malicious activities and trigger further monitoring mechanisms. Similarly, according to Lee *et al.* [42], the statistical nature of spam traffic shows high degree of self-similarity compared to normal traffic. Thus, this distinctive behavior can help in detecting spam traffic. Moreover, Yan and Wang [44] propose an anomaly detection method for the security evaluation of LAN traffic by using H parameter variation analysis of four different metrics i.e., all packets, TCP packets, UDP packets and ARP packets. Similarly, in [45]–[49], considering the self-similar and LRD nature of the traffic, the variation in H parameter is used for the detection of possible anomaly in the traffic.

Furthermore, one of the major causes of anomalies in network traffic is the continuous rapid growth of DDoS attacks. DDoS attacks perturb the normal flow of network traffic by capturing network recourses. Since it affects the volume of traffic, it could contribute to the deviation of network traffic from LRD behavior. To date, various approaches based on self-similar and LRD nature of network traffic have been proposed for the detection of DDoS attacks. Liu *et al.* [50] use the autocorrelation and H parameter measurements for the early detection of network traffic anomalies caused by DDoS attacks. Similarly, Nurohman and Purwanto [51] develop a method for the detection of anomalies, particularly DDoS attacks based on kolmogorov-smirnov test and H estimator. This method can differentiate normal traffic from abnormal traffic by exploiting the self-similar nature of traffic. Moreover, Kaur *et al.* [40] use wavelet-based estimation of LRD to differentiate between flash crowds and pulsating DDoS attacks. Likewise, Zhang *et al.* [41] develop a method in the wavelet domain in light of LRD behavior of network traffic for the detection of outliers such as DDoS attacks. Conversely, Jian-Qi *et al.* [52] propose a DoS attack detection approach based on the distributional features of packet compositions (i.e., source and destination address and ports), which shows composition self-similarity (CSS) in local-world network (number of hosts). Where the traffic

anomalies (DoS attacks) cause changes in the distribution of addresses or ports observed in traffic, hence, affect the H parameter of CSS. Similarly, in [53]–[56] and references therein, the self-similar and LRD behavior of network traffic is used in one way or another for the detection of DDoS attacks in network traffic. Overall, the literature shows that in the presence of anomaly, the self-similar and LRD behavior of network traffic varies. Thus, this variation in the behavior of network traffic can be modeled to detect the corresponding possible anomalies in the traffic.

On the other hand, besides the LRD model based anomaly detection methods, numerous anomaly detection methods based on various other models/approaches have been proposed in the literature, for instance the methods in [57]–[63]. In [57], a scheme called stream projected outlier detector (SPOT) is developed for anomaly detection in high-dimensional data streams from wireless network. SPOT is comprised of learning and detection stages. In the learning stage, it incorporates an outlier-based mechanism where with the help of multiobjective genetic algorithm (MOGA), the projected anomalies subspaces are searched in multi-dimensional data streams, which are then used in the detection stage for the detection of possible anomalies. Likewise, in [58] an algorithm is developed for the detection of anomalies in sensors' data with uncertainties. The algorithm incorporates belief-rule-based association rule mining for overcoming uncertainties in the data, such as vagueness, ignorance, ambiguity, incompleteness, and imprecision. The algorithm is evaluated on two different datasets against other methods in the same domain with comparatively better results. Moreover, in [59], a semi-supervised discriminative restricted Boltzmann machine based anomaly detection technique is proposed. The notion of this method is to find any inherent similarity between normal and anomalous traffic for characterization. Based on experiments, it is observed that the performance of the proposed classifier suffers when it is tested in a network other than the network from where the training data was taken. Therefore, further investigation of the anomalous behavior and its differences to the inherent nature of the normal traffic is suggested. On the other hand, in [60], a users' behavior analysis based anomaly detection scheme is proposed, which characterizes normal and abnormal users' behavior using principal component analysis (PCA). The method analyzes the users' database access behavior and web browsing behavior for such characterization. In addition, a method for anomaly detection based on maximum and relative entropy estimations is introduced in [61]. The maximum entropy is used to estimate the packet distribution of normal traffic that is used as a baseline profile, whereas the relative entropy estimation of the traffic under observation is used to identify anomalous behaviors in the traffic. Conversely, in [62], an anomaly detection method based on non-linear characteristics analysis of aggregated IP traffic flows is proposed. The method constructs a baseline normal traffic profile and then against this profile, it analyzes the aggregated IP flow traffic for non-stationary events and hidden recurrence patterns using recurrence quantification analysis (RQA) and support vector machines (SVM). Based on such analysis the anomalous traffic is detected and classified. Similarly, in order to understand and analyze the dynamic nature of the traffic for anomalies, in [63], a distributed sensors network based anomaly detection system is proposed. In this system, sensor agents are installed at different nodes at the network, which collect the traffic at their network interfaces and preprocess it to extract features of interest that could best describe the traffic dynamics by using independent component analysis. The collected features' data is then aggregated at a central collector for further analysis and classification of normal and anomalous events using a decision tree classifier.

Apart from the related work presented, considerable research has been devoted to establishing anomaly detection methods based on LRD model and various other state-of-the-art models. However, to the best of our knowledge none of them, except in our previous preliminary work [9], [64], consider analyzing the LRD behavior of control and data traffic separately in different directions with respect to the enterprise network. In our proposed method, we introduce a paradigm shift in the LRD behavior analysis based anomaly detection methods from the analysis of aggregated WHOLE (control plus data) traffic to the analysis of aggregated decomposed traffic as discussed in the next section.

IV. METHODOLOGY: NETWORK ANOMALY DETECTION USING LRD ANALYSIS OF PACKET AND BYTE COUNTS IN DECOMPOSED NETWORK TRAFFIC

In the proposed method, the Internet traffic is preprocessed and analyzed using its intrinsic LRD behavior. The basic framework of the proposed method and a flow chart of the complete process of network traffic analysis are depicted in Figure 2 and Figure 3, respectively, where the traffic analysis is mainly comprised of two modules: Internet traffic preprocessing and LRD estimation. The traffic preprocessing module is further comprised of three processes: TCP traffic filtering, traffic decomposition, and packet and byte count extraction. After capturing the bidirectional Internet traffic i.e., incoming and outgoing, with respect to LAN at the router of the network as shown in Figure 2, the traffic is preprocessed by filtering the TCP traffic and decomposing it into different subgroups of control and data planes in different directions with respect to the enterprise network as depicted in Figure 3. Control plane traffic contains packets that set, maintain, or tear down a connection, whereas the data plane traffic contains packets that are responsible for the actual transfer of data [7], [8]. We decompose the traffic into control and data planes based on the TCP flags and the length fields in the TCP and IP header of each packet. For further decomposition in different directions with respect to the enterprise network, we use the corresponding source and destination IP addresses fields in each packet. We treat TCP packets having any of these flags: SYN, FIN or RST as control packets. Since these packets are used to establish, maintain, or tear down a TCP connection. Moreover, bare acknowledgements (ACKs)

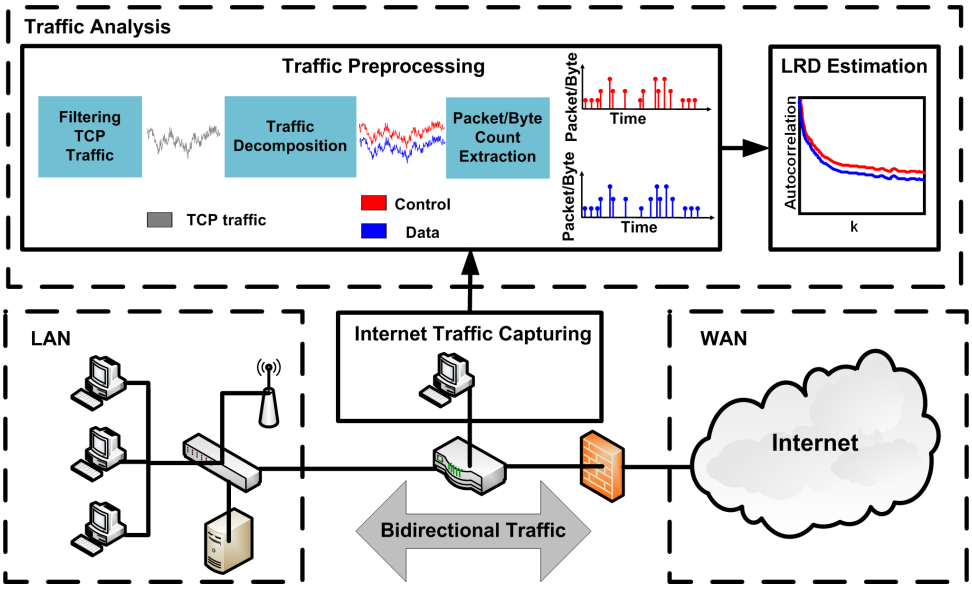

FIGURE 2. A framework of the proposed anomaly detection method.

are also treated as control packets, and the rest of packets, including piggybacked acknowledgements, are treated as data packets. In contrast to piggybacked ACKs, bare ACKs do not carry payloads. After this decomposition, the packet and byte counts in the corresponding subgroups are extracted by aggregating them in bins using suitable aggregation interval, i.e., binswidth. Such bins give us the time series of packets and bytes counts. In the LRD estimation module, SOSS and FARIMA tests are applied to such time series of the corresponding traffic subgroups to estimate the LRD parameter i.e., H and d. Since SOSS and FARIMA models, which are given in the OM method in [16] and [35], are integrated in the proposed method as the normal model for anomaly detection; therefore, based on such estimates, the traffic is classified as either normal or anomalous.

In the preprocessing phase, the TCP traffic is decomposed into the following traffic subgroups:

- Incoming control (IC) traffic.
- Outgoing control (OC) traffic.
- Bidirectional control (BC) traffic (combination of IC and OC).
- Incoming data (ID) traffic.
- Outgoing data (OD) traffic.
- Bidirectional data (BD) traffic (combination of ID and OD).
- Incoming whole (IW) traffic (combination of ID and IC).
- Outgoing whole (OW) traffic (combination of OD and OC).
- Bidirectional WHOLE (BW) traffic (combination of BC and BD).

The bidirectional WHOLE traffic is the combination of control and data in both directions with respect to the network, which in fact represents the undecomposed network traffic. Note that for the comparison purpose, we also analyze this bidirectional WHOLE traffic. Because most of the previous works in the area of LRD behavior analysis are based on analyzing this traffic. Hence, in this way we can demonstrate the comparison and performance efficacy of the proposed method. Furthermore, we note here that the proposed method focuses on TCP traffic, since most of Internet traffic is associated with TCP [7], and TCP traffic constitutes 80 – 90% of the Internet traffic [65]. Besides, TCP traffic can be easily decomposed into control and data planes traffic using the header information in TCP packets as discussed in [7] and [8]. For further decomposition of traffic into subgroups based on direction, the corresponding source and destination IP addresses fields in each TCP packet are used. After the traffic is decomposed, then the corresponding packet count and byte count series i.e., packet/byte bins, in the traffic subgroups are extracted using appropriate aggregation interval i.e., binwidth. The time series of such bins is analyzed for LRD behavior using appropriate time windows.

In the LRD estimation phase, the extracted time series of packet and byte counts in the form of bins in all of the traffic subgroups are analyzed against the underlying LRD model. For LRD model, SOSS and FARIMA models are integrated in the proposed method, which are provided in the OM method in [16] and [35]. According to OM, if the minimum error in (5) for any of these models is less than a threshold value of 1×10^{-3}, then the data under analysis fits the underlying model and the minimizer $\hat{\beta}$ is the value of LRD parameter ($\hat{H}$ *or* $\hat{d}$). This implies that the traffic under analysis exhibits LRD behavior. In contrast, if the minimum error is greater than the threshold, 1×10^{-3}, for any of the models, then either the data under analysis does not follow the assumed model or the volume of the available data is not sufficiently high for the right decision of LRD behavior. Therefore, in such case the value of $\hat{\beta}$ is of no importance i.e., then it does not

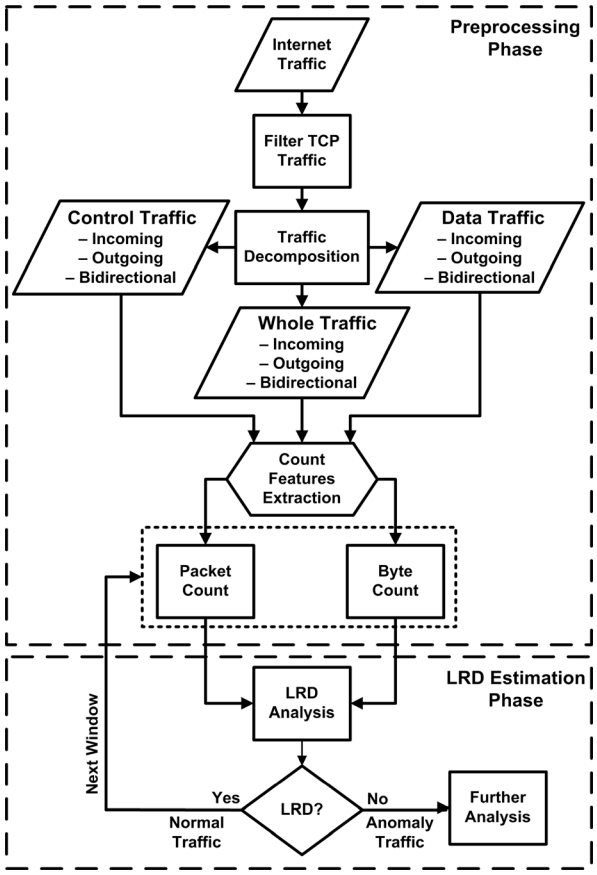

FIGURE 3. Flow diagram describing the methodology.

represent the corresponding LRD. The selection of threshold value of 10^{-3} in [16] and [35] is made to set the probability of false alarm (i.e., declaring that the process does not exhibit LRD behavior where in fact it does), to less than 0.05.

Moreover, since we integrate both SOSS and FARIMA models in our method, therefore, the traffic under analysis is tested against both models concurrently in an appropriate iterative overlapping sliding time window with 50% shift. In each window, the LRD in packet and byte count bins of the traffic subgroups is estimated by fitting the correlation structure of the underlying data in (5) to the correlation structures of SOSS and FARIMA models, given in (1) and (4) respectively, concurrently. In this way, if any of the traffic subgroups deviates from the correlation structures of both the underlying models, i.e., the minimum error in (5) is greater than 1×10^{-3} for both models, then the traffic under analysis is classified as anomalous. Such traffic is considered for further investigation in order to find the root cause of the corresponding anomaly. This allows the verification of whether the detected anomalous behavior is due to benign normal activities, or due to attacks and cyber intrusions. Note that in such cases, the corresponding estimated LRD parameters i.e., H and d, are ignored, as they do not represent the LRD process. Conversely, if all of the traffic subgroups follow the correlation structure of one or both of the underlying models i.e., the minimum error in (5) is less than 1×10^{-3} for one or both models, then the corresponding traffic is classified as normal. As the process is iterative, so after a window slides with 50% shift, the next window start estimating LRD in the packet and byte counts of the traffic sequences. We use sliding window with a 50% shift in order to get accurate and faster results rather than waiting for the next entire window to pass. Moreover, using a 50% shift instead of 100% shift increases the chances that a given abnormity is well covered by a single window. However, using a shift larger than 50% will less likely achieve this. Besides, larger shifts slow down the analysis process. On the other hand, using a shift smaller than 50% will speed up the process and provide better coverage of a given abnormality, but it comes at the expense of more processing without significant improvements.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. DATASET

We present experimental results on analyzing a relatively recent real Internet traffic dataset, which was captured at King Saud University (KSU) network on Dec. 22, 2012 to Feb. 9, 2013 [65]. We use KSU dataset because it contains real (not simulated) and relatively recent Internet traffic. In addition, this dataset reflects newer traffic patterns, with focus on social media, online streaming and other newer applications.

The dataset contains traffic of intra-LAN communication and the communication between KSU's LAN and the Internet. The dataset is comprised of more than 11 Tera (10^{12}) bytes of Internet traffic, of which 90% is worth TCP traffic. At the time of capturing this dataset, the network had approximately more than 10,000 active hosts. In addition, the IP addresses in this dataset are all anonymized in order to protect the identity and preserve the privacy of external and internal hosts. Since we are interested in TCP traffic, therefore, the TCP traffic in the dataset is filtered out and the information encompasses in each TCP packet is preprocessed and well organized in comma separated value (CSV) files. The following information is provided for each TCP packet:

- Time stamp
- Source IP address
- Source port
- Destination IP address
- Destination port
- TCP Flag (e.g., SYN, FIN, RST, or no flag is set)
- Data sequence number of the packet
- Data sequence number of the expected return data
- Acknowledgment sequence number of the expected next data
- Receiver window size i.e., the number of bytes that receiver can receive
- Total length of the frame

We use this information to decompose the traffic and to extract the corresponding packet and byte counts sequences. Moreover, this information is also used to closely investigate any possible anomalous behavior after the detection, in order to reach to the root cause of the corresponding anomalous behavior.

B. EMPIRICAL SELECTION OF APPROPRIATE AGGREGATION INTERVAL AND WINDOW SIZE

In order to produce accurate results, we start our analysis with the empirical evaluation of aggregation interval and window size for the underlying dataset because they are crucial factors in the LRD behavior analysis of aggregated traffic. We empirically identify appropriate values for these parameters that are capable of capturing the LRD behavior in the WHOLE traffic, i.e., bidirectional control plus data traffic, as well as in the decomposed traffic.

We demonstrate the empirical selection of these parameters by analyzing the LRD behavior in traffic of different times of days and nights from the underlying dataset using SOSS and FARIMA models. We conduct multiple experiments by aggregating the traffic with several aggregation intervals i.e., 0.1 – 3 seconds, and then analyzing with a number of different iterative overlapping sliding windows, e.g., 300, 600, 1200, and 1800 packets/bytes' bins (each bin is resulted from the number of packets or bytes aggregated in the concern interval), with 50% shift. In accordance to the aggregation intervals/binswidth used, these windows of analysis in terms of time ranges from 30 seconds to 90 minutes, which are calculated as follows; $binswidth \times \text{numberofbins} = \text{windowsize}$. For example, an aggregation interval of 0.1 sec and time window of analysis of 300 bins cover 30 seconds of the traffic trace (i.e., $0.1\ sec \times 300 = 30\ sec$). Similarly, an aggregation interval of 3 seconds and time window of 1800 bins represents 90 minutes of the traffic trace.

As a conclusive observation, here we present the results on three different traffic scenarios from the KSU' dataset: working hours' traffic, late night's traffic and weekend's traffic, which cover the traffic of busy and non-busy hours i.e., higher loads and lower loads. We analyze these traces by using window sizes of 1200 and 1800 bins. These traffic scenarios include Saturday (8 to 10 am) traffic, Saturday (12 to 2 am) traffic and Friday (8 to 10 am) traffic,[2] respectively. Table 1 shows the average traffic rates in these traffic traces in kilo packets per second (KPPS) and megabits per second (Mbps).

From our initial exhaustive iterative analysis, we observed that the window sizes of 300 and 600 bins are not sufficient to capture the LRD behavior in all of the traffic subgroups; as a result, we do not consider them. We also found that a window size of 1200 bins is not sufficient to capture the LRD behavior in most of the traffic subgroups. In contrast, a window size of 1800 bins is large enough to capture LRD behavior in the all of the traffic subgroups. However, for the sake of comparison, here we also present results on window size of 1200 bins.

The results of analyzing LRD behavior in the aggregated packet and byte counts of decomposed traffic in Trace 1 given in Table 1 using aforementioned window sizes and various aggregation intervals are shown in Figure 4. The x-axis in the Figure represents the aggregation intervals/binswidths over which the traffic is aggregated, and the y-axis represents the percentage failures of capturing LRD behavior in the two hours traffic of trace 1 using time windows sizes of 1200 and 1800 bins respectively. Since we use sliding window with 50% shift, therefore, the percentage shows the failures to capture the LRD behavior in each window as we move along the two hours traffic. In other words, these percentage

[2]During capturing of the traffic, Thursday and Friday were weekend days in the Kingdom of Saudi Arabia

TABLE 1. Average traffic rates in saturday and friday's traffic traces (2-hours each).

Decomposed Traffic		Saturday Trace 1 8 am – 10 am		Saturday Trace 2 12 am – 2 am		Friday Trace 3 8 am – 10 am	
Direction	Traffic	KPPS	Mbps	KPPS	Mbps	KPPS	Mbps
Incoming	Control	1.8	0.9	0.1	0.1	0.2	0.1
	Data	10.7	110.4	0.5	4.6	0.7	5.7
	whole	12.5	111.3	0.6	4.7	0.9	5.8
Outgoing	Control	7.6	4.0	0.3	0.2	0.5	0.3
	Data	1.87	11.4	0.2	1.5	0.3	1.2
	whole	9.5	15.4	0.5	1.7	0.8	1.5
Bidirectional	Control	9.4	4.9	0.4	0.2	0.7	0.4
	Data	12.6	121.8	0.7	6.2	1.0	6.9
	WHOLE	22.0	126.7	1.1	6.4	1.7	7.3

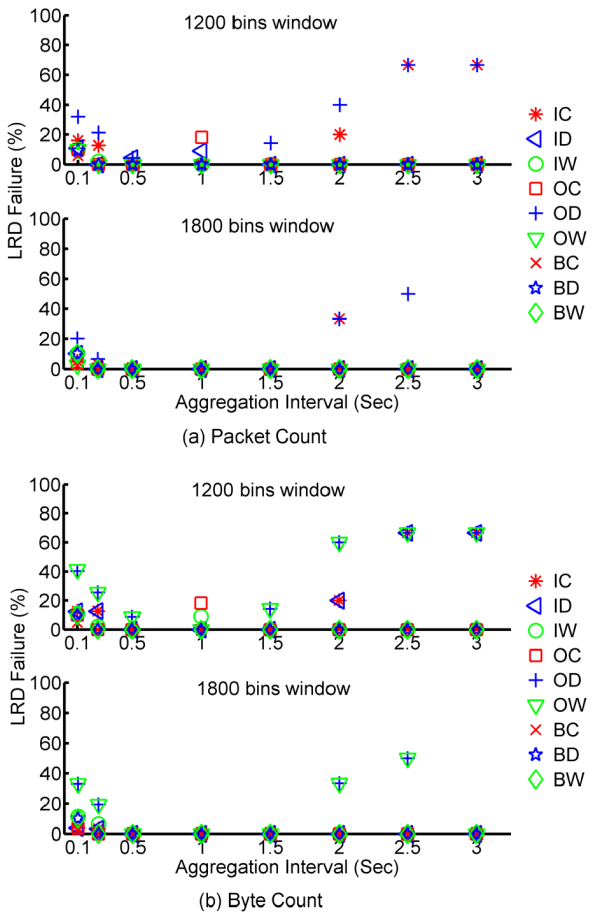

FIGURE 4. Percentage failures in capturing of LRD behavior in traffic of Saturday traffic, Trace 1 (8 am – 10 am), against various aggregation intervals and window sizes (a) packet count analysis (b) byte count analysis.

failures in capturing LRD behavior are based on the number of observations for each aggregation interval over a length of two hours traffic. Hence, these failures do not imply the failure of capturing the LRD behavior in the entire traffic trace, but rather the failure in a particular window of analysis. The percentage is obtained by dividing the number of failures by the total number of observations or windows of analysis. Whereas the number of observations is evaluated as follows;

$$\text{number of observations} = \frac{\textit{traffic trace length}}{\textit{window size}} + \frac{\textit{traffic trace length} - \textit{window size}}{\textit{window size}}$$

For example, if the traffic is aggregated using an aggregation interval of 0.1 second, then a window size of 1200 bins covers 120 seconds (2 minutes) of the traffic trace. Thus, using this time window for analysis with a sliding period half of its size gives us 119 observations over a period of 2-hours (120 minutes) traffic, i.e.,

$$\frac{120 \text{ minutes}}{2 \text{ minutes}} + \frac{120 \text{ minutes} - 2 \text{ minutes}}{2 \text{ minutes}} = 119 \text{ observations}$$

Similarly, in case of an aggregation interval of 3 seconds, and analysis window of 1800 bins, we have only 1 observation over a 2-hours period because it covers 5400 seconds (90 minutes) of the traffic. The top plot in Figure 4 (a) shows the results of analyzing the LRD behavior in packet count of all of the traffic subgroups using a window size of 1200 bins, while the bottom plot shows the results for a window size of 1800 bins. Likewise, Figure 4 (b) shows the results of analyzing the byte count feature. It can be seen from the Figure that the aggregation interval of 0.1 second with both window sizes fails to capture the LRD behavior in certain slots of all of the traffic subgroups. However, when we increase the aggregation interval and window size, the failures decrease. At aggregation intervals of 0.5 second and 1 second and with a window size of 1800 bins, the LRD behavior is captured in both packet and byte count sequences of all of the traffic subgroups as can be seen from Figure 4. This window size of 1800 bins corresponds to 15 and 30 minutes when the traffic is aggregated with an aggregation interval of 0.5 and 1 second, respectively. Nonetheless, if we further increase the aggregation interval and window size, then it could induce non-stationarity in the analysis, which could mislead the results. On the other hand, a window size of 1200 bins fails to capture the LRD behavior in certain traffic subgroups, particularly the *incoming control* and *outgoing data* at most of the aggregation intervals including 0.5 second and 1 second as shown in Figure 4. This might be due to the low volume of traffic in these two traffic subgroups as can be seen from Table 1. However, the failures in capturing the LRD behavior are comparatively only in few traffic subgroups at aggregation interval of 0.5 second and higher. This is because an increment in the aggregation interval might aggregate more sources, which leads to an increase in volume of underlying traffic in the corresponding windows.

On the other hand, Figure 5 shows the results of analyzing late night traffic, i.e., Trace 2 in Table 1. It can be seen from Figure that most of the aggregation intervals and window sizes fail to capture the LRD behavior in both packet and byte count of almost all the traffic subgroups including the WHOLE traffic. The reason behind this is the very low volume of traffic during this time in the concern traffic trace as shown in Table 1. However, the LRD behavior is captured in all of the traffic subgroups at larger aggregation intervals other than the aggregation interval of 0.5 and 1 second as shown in Figure 5, which are not appropriate for anomaly detection analysis. Comparatively, similar results are observed for Friday's traffic trace as shown in Figure 6. Although, in both traffic traces i.e., Trace 2 and 3, the WHOLE traffic exhibits LRD behavior at larger aggregation intervals and window sizes. Nevertheless, the incoming control and the outgoing data traffic fail to exhibit the LRD behavior in both packet and byte counts sequences at most of the aggregation intervals and window sizes used. This is because the volume of incoming control and outgoing data is lower as compared to the WHOLE traffic as can be seen from Table 1. Moreover, since Friday is a weekend day and late night is an idle time; therefore, there might be less number of active users, which result in low volume of traffic.

In conclusion, after conducting extensive experiments on the KSU's dataset, we observe that using aggregation intervals of 0.5 and 1 second to aggregate the traffic enable us to capture the LRD behavior successfully by analyzing the corresponding aggregated traffic with a window size of 1800 bins, i.e., 15 and 30 minutes, respectively. Therefore, these aggregation intervals and window sizes are appropriate for LRD behavior analysis of traffic, particularly for anomaly detection. Note that these aggregation intervals are selected based on the empirical analysis of KSU's dataset, where the average traffic rate is around 1 KPPS. However, for any given dataset, if the traffic rate decreases, then the aggregation interval should be increased and vice versa.

C. NETWORK ANOMALY DETECTION: EXPERIMENTAL EVALUATION AND VALIDATION

The KSU's dataset contains few traffic traces that are identified with anomalous patterns in the traffic. We analyze four of such traffic traces in this work to test the detection efficacy of our proposed method. These traffic traces are two hours in length each and were captured on Sunday, Monday, Tuesday, and Wednesday of the first week in the KSU's dataset during the hours of 08:00:00 – 10:00:00, 14:00:00 – 16:00:00, 10:00:00 – 12:00:00, and 12:00:00 – 14:00:00, respectively. The average traffic rates in these traffic traces are 147.5 Mbps, 88.4 Mbps, 145.3 Mbps, and 143.7 Mbps, respectively. We select these traffic traces since they are during busy hours on working days and contain anomalies in them. We find that Monday and Wednesday's traffic traces contain malicious traffic, whereas the other two encompass abnormalities in them. We highlight the corresponding anomalous patterns in the course of the traffic through the proposed method and

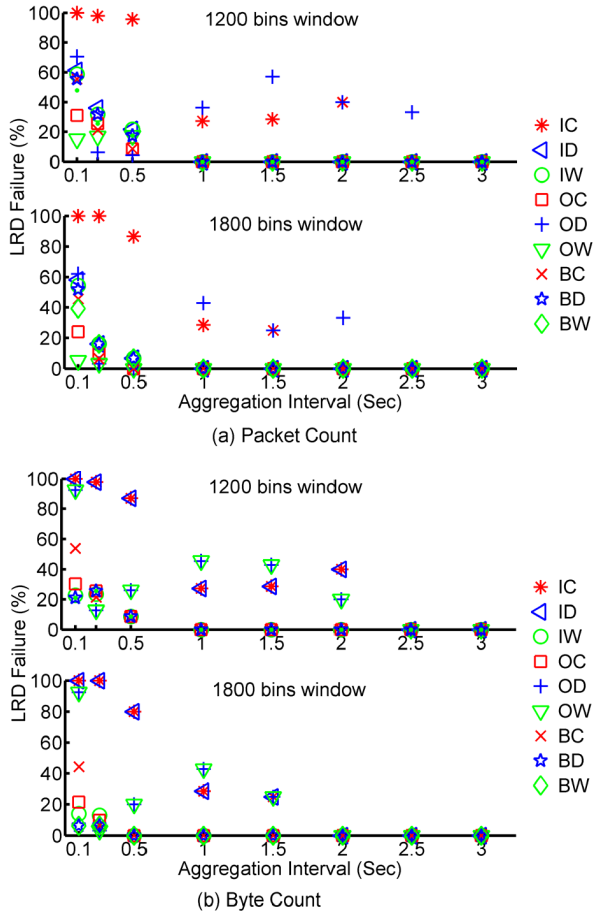

FIGURE 5. Percentage failures in capturing of LRD behavior in traffic of Saturday traffic, Trace 2 (12 am – 2 am), against various aggregation intervals and window sizes (a) packet count analysis, (b) byte count analysis.

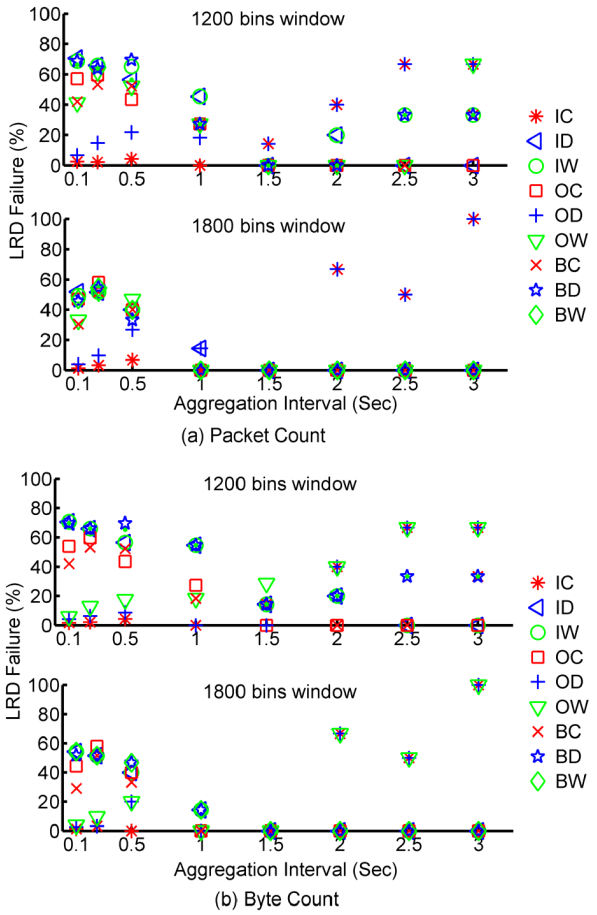

FIGURE 6. Percentage failures in capturing of LRD behavior in of Friday traffic, Trace 3 (8 am – 10 am), against various aggregation intervals and window sizes (a) packet count analysis, (b) byte count analysis.

demonstrate how we further investigate them for the identification of the actual root cause in Wednesday and Monday's traffic.

Firstly, we demonstrate the presence of anomalous patterns in the concern traffic traces based on the concept presented in [7] and [8]. According to [7] and [8], the data traffic generation is based on the control traffic generation; therefore, the two traffic planes should have similar time variation during benign normal conditions, whereas they might have dissimilar time variation during abnormal conditions. Consequently, we can depict the anomalous condition by plotting the bidirectional control and data planes traffic simultaneously. For instance, the bidirectional control and data planes normal traffic from KSU's dataset that was captured on Sunday during, 06:00:00 – 08:00:00, shows similar time variations as shown in Figure 7. The left plot represents the volume of packet count, whereas the right plot shows the volume of byte count. The byte count is plotted using log-scale. As we can see from the two plots in the Figure, there is a similar time variations between the control and data planes traffic i.e., the data follows the control closely.

On the other hand, if we look at the bidirectional control and data planes traffic of the aforementioned four anomalous traffic traces, we can see that there is dissimilarity between the time variations of the two traffic groups as shown in Figures 8 and 9. In these plots, there are abrupt spikes in the bidirectional control traffic, which disrupt the similarity between the two traffic planes. Such dissimilarity indicates anomalous event in the traffic. The volume of spikes is huge in Monday and Wednesday's traffic traces compared to Sunday and Tuesday's traffic traces. We can see that the dissimilarity in the behavior of control and data p lanes traffic starts at the time of abrupt spikes in the control plane traffic as shown in Figures 8 and 9. However, before and after this time, the data plane traffic closely follows the track of the control plane traffic. These spikes highlight the presence of anomalous behavior in the corresponding slots of the traffic. Since certain anomalous behaviors manifest themselves mainly in the control plane traffic, so by comparison between the control and data planes, we can observe them.

Secondly, we present the results of applying our proposed method to the aforementioned four anomalous traffic traces using window sizes of 15 and 30 minutes. Table 2 shows the result of analyzing the packet count in the abnormal duration of Sunday's traffic trace using a window size of 30 minutes.

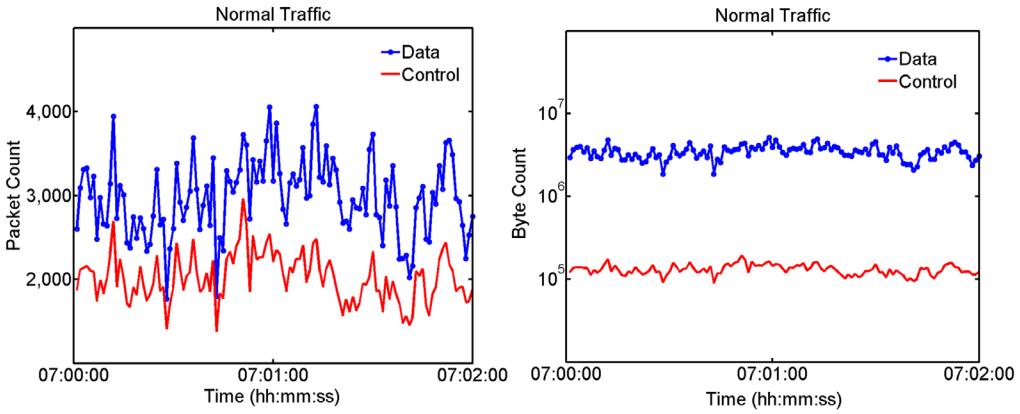

FIGURE 7. Similarity in the behavior of bidirectional control and data planes of normal traffic.

The corresponding traffic is aggregated with an aggregation interval of 1 second. We can see from Table 2 that the error value is less than the threshold value of 1×10^{-3} for all of the traffic subgroups using both SOSS and FARIMA models. This implies that all traffic subgroups including aggregated WHOLE traffic exhibit LRD behavior. Similar, results are observed for the corresponding byte count in the abnormal duration of Sunday's traffic trace as can be seen from Table 3. Although there is an abnormality in this particular duration of traffic, the resultant misdetection is due to the use of large window size for analysis. Nevertheless, when we decrease the window size to 15 minutes, which results from the aggregation interval of 0.5 seconds, the abnormality is detected in both packet and byte count of the *incoming control* traffic as can be seen from Tables 4 and 5 respectively. At the same time, the abnormality is not detected in the rest of the

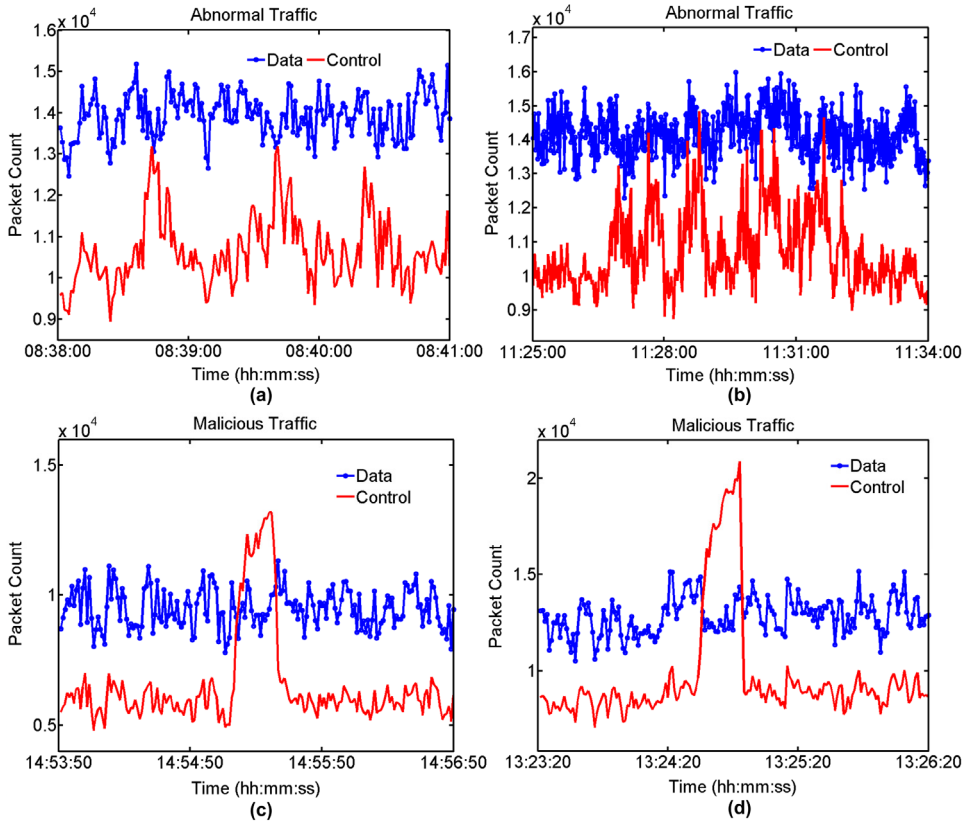

FIGURE 8. Packet count in the bidirectional control and data traffic showing abnormal traffic in (a) Sunday's traffic trace (b) Tuesday's traffic trace, and malicious traffic in (c) Monday's traffic trace (d) Wednesday's traffic trace.

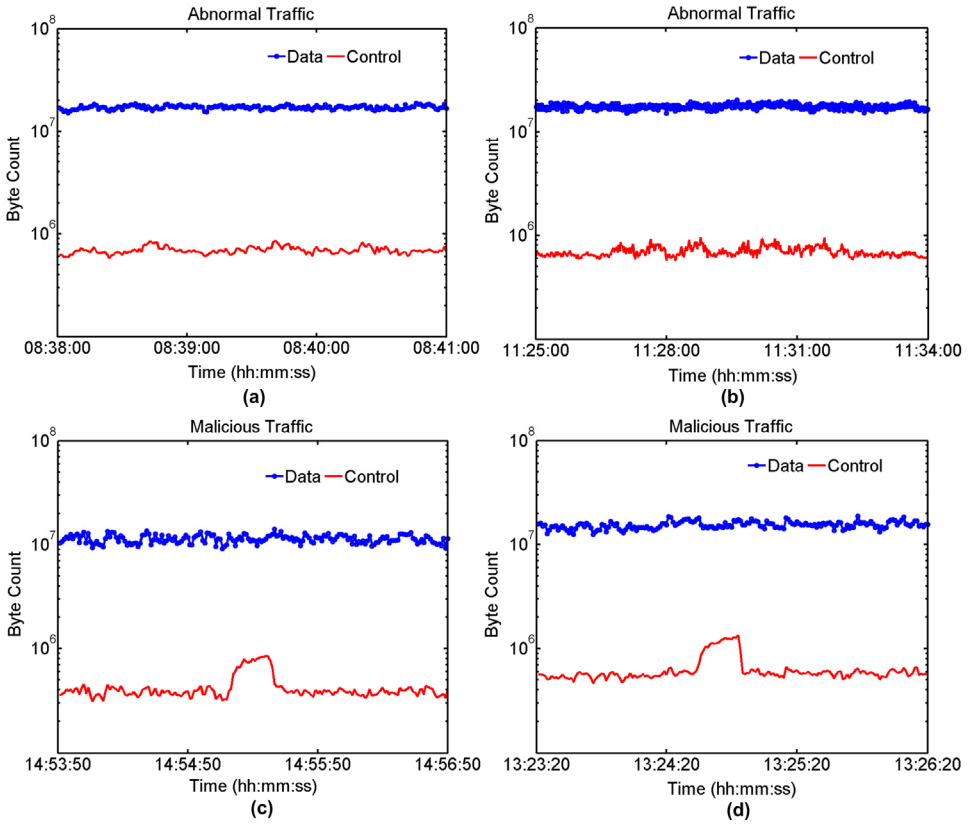

FIGURE 9. Byte count in bidirectional control and data planes traffic showing abnormal traffic in (a) Sunday's traffic trace (b) Tuesday's traffic trace, and malicious traffic in (c) Monday's traffic trace (d) Wednesday's traffic trace.

traffic subgroups, particularly the WHOLE traffic. This is because the percentage volume of packet and byte counts in the aggregated incoming control traffic is 8% and 0.7% of the aggregated WHOLE traffic as shown in Tables 4 and 6, respectively. Therefore, the overall effect of the abnormality in the control is buried under the WHOLE traffic; hence, the abnormality is not detected in the aggregated WHOLE traffic. Note that when the value of error for any of the tests is greater than 1×10^{-3}, then the corresponding estimated values of the LRD parameters i.e., $\hat{H}$ and $\hat{d}$, are ignored and denoted by "--" in the Tables. Since in such cases, the traffic under analysis do not exhibit LRD behavior.

Moreover, Tables 6 and 7 show the results of analyzing the abnormal traffic in Tuesday's traffic trace through packet count and byte count. In this case, the concern abnormality is detected in the aggregated *bidirectional control* traffic as can be seen from Tables 6 and 7. However, it is not detected in the other subgroups, including the aggregated WHOLE traffic.

TABLE 2. LRD behavior analysis of packet count sequences in abnormal duration of sunday's traffic trace (08:00:00 – 10:00:00) using window size of 30 minutes and aggregation interval of 1 second.

Decomposed Traffic		Vol.	SOSS Test		FARIMA Test		LRD?
Direction	Traffic	(%)	H	Error	d	Error	
Incoming	Control	8	0.89	4.92E-04	0.39	6.69E-04	yes
	Data	49	0.89	1.97E-04	0.4	1.15E-04	yes
	whole	57	0.89	9.39E-05	0.4	1.07E-04	yes
Outgoing	Control	34	0.87	1.20E-04	0.38	1.57E-04	yes
	Data	9	0.89	5.77E-04	0.4	7.19E-04	yes
	whole	43	0.87	1.98E-04	0.38	2.47E-04	yes
Bidirectional	Control	42	0.87	2.50E-04	0.38	3.56E-04	yes
	Data	58	0.89	1.16E-04	0.4	1.53E-04	yes
	WHOLE	100	0.88	1.27E-04	0.39	1.29E-04	yes

TABLE 3. LRD behavior analysis of byte count sequences in abnormal duration of sunday's traffic trace (08:00:00 – 10:00:00) using window size of 30 minutes and aggregation interval of 1 second.

Decomposed Traffic		Vol.	SOSS Test		FARIMA Test		LRD?
Direction	Traffic	(%)	H	Error	d	Error	
Incoming	Control	0.7	0.88	3.8E-04	0.39	5.2E-04	yes
	Data	86.3	0.89	2.0E-04	0.4	1.2E-04	yes
	whole	87.0	0.89	1.9E-04	0.4	1.1E-04	yes
Outgoing	Control	2.9	0.87	1.1E-04	0.38	1.6E-04	yes
	Data	10.1	0.9	6.7E-04	0.41	9.2E-04	yes
	whole	13.0	0.9	6.7E-04	0.41	9.1E-04	yes
Bidirectional	Control	3.6	0.87	2.5E-04	0.37	3.3E-04	yes
	Data	96.4	0.89	9.8E-05	0.4	1.3E-04	yes
	WHOLE	100.0	0.89	9.3E-05	0.4	1.2E-04	yes

TABLE 4. LRD behavior analysis of packet count sequences in abnormal duration of sunday's traffic trace (08:00:00 – 10:00:00) using window size of 15 minutes and aggregation interval of 0.5 second.

Decomposed Traffic		Vol.	SOSS Test		FARIMA Test		LRD?
Direction	Traffic	(%)	*H*	Error	*d*	Error	
Incoming	Control	8	--	**1.8E-03**	--	**2.0E-03**	**No**
	Data	49	0.87	2.2E-04	0.38	1.3E-04	yes
	whole	57	0.87	1.8E-04	0.38	1.4E-04	yes
Outgoing	Control	34	0.85	3.4E-04	0.36	3.8E-04	yes
	Data	9	0.88	4.9E-04	0.39	6.5E-04	yes
	whole	43	0.86	2.7E-04	0.37	3.5E-04	yes
Bidirectional	Control	42	0.86	8.4E-04	--	1.1E-03	yes
	Data	58	0.87	1.6E-04	0.38	1.2E-04	yes
	WHOLE	100	0.87	2.1E-04	0.38	2.5E-04	yes

TABLE 6. LRD behavior analysis of packet count sequences in abnormal duration of tuesday's traffic trace (10:00:00 – 12:00:00) using window size of 15 minutes and aggregation interval of 0.5 second.

Decomposed Traffic		Vol.	SOSS Test		FARIMA Test		LRD?
Direction	Traffic	(%)	*H*	Error	*d*	Error	
Incoming	Control	7	0.9	9.57E-04	--	1.08E-03	yes
	Data	52	0.82	4.90E-04	0.34	3.96E-04	yes
	whole	59	0.85	4.96E-04	0.36	4.70E-04	yes
Outgoing	Control	34	0.87	9.42E-04	--	1.05E-03	yes
	Data	7	0.75	5.97E-04	0.26	6.23E-04	yes
	whole	41	0.85	6.30E-04	0.36	6.77E-04	yes
Bidirectional	Control	41	--	**1.15E-03**	--	**1.34E-03**	**No**
	Data	59	0.81	4.42E-04	0.32	3.62E-04	yes
	WHOLE	100	0.85	5.72E-04	0.36	5.86E-04	yes

This is because the abnormality is contained in the control traffic and the control traffic has lower volume compared to the rest of the aggregated traffic. Moreover, similar to Sunday's abnormality, the volume of Tuesday abnormality is small; therefore, in the analysis we opt for a window size of 15 minutes rather than 30 minutes.

Conversely, Tables 8 – 11 show the results of LRD behavior analysis of packet and byte counts in the malicious traffic of Monday and Wednesday's traffic traces, respectively. Since the volume of malicious traffic in these two traffic traces is high compare to that of Sunday and Tuesday's abnormalities as shown in Figures 8 – 9. Therefore, such malicious traffic is detected with even using a larger aggregation interval of 1 second and window size of 30 minutes. From Tables 8 – 9, we can see that similar to Sunday's traffic, the *incoming control* traffic fails to exhibit LRD behavior. In this case, since the volume of malicious packets present in the control traffic is very high, therefore, it even causes the *bidirectional control* traffic to fail to exhibit LRD behavior. However, in the case of packet count, the rest of the traffic subgroups still exhibit LRD behavior, particularly the aggregated WHOLE traffic. This is because the percentage volume of packet count in the incoming and bidirectional control traffic is 6% and 39% of the WHOLE traffic, respectively, as shown in Table 8, thus, the effect of malicious traffic is overshadowed under the WHOLE traffic. Yet, this anomaly is also detected in the *outgoing data* and *outgoing whole* traffic in the corresponding byte count as given in Table 9. The reason for this detection is the overall impact of the anomaly on these two traffic subgroups. Furthermore, it can be seen from Tables 10 – 11 that the control plane traffic in all directions fails to exhibit LRD, while the other subgroups still exhibit LRD behavior. Since it is the incoming and outgoing control traffic, which carry the malicious packets; consequently, the bidirectional traffic is also affected. Nevertheless, in the case of WHOLE traffic, the overall effect is overshadowed. From these results, we can infer that the anomalous behavior is carried in the incoming and/or outgoing control plane traffic. Therefore, it could not be detected by merely analyzing the aggregated WHOLE traffic i.e., bidirectional whole traffic.

In addition, the overall comparison of packet and byte count results is given in Table 12. The tick mark (✓) in the Table represents the detection of the anomaly in the concern traffic subgroups, while the cross mark (✗) signifies the misdetection. As can be seen from the Table 12, the control traffic is the most effective subgroup to detect anomalies in both packet and byte counts. This is because the anomalous traffic is mainly present in the control plane traffic. Moreover, the packet and byte count show similar results for all the traffic traces except for Monday's abnormal traffic, where

TABLE 5. LRD behavior analysis of byte count sequences in abnormal duration of sunday's traffic trace (08:00:00 – 10:00:00) using window size of 15 minutes and aggregation interval of 0.5 second.

Decomposed Traffic		Vol.	SOSS Test		FARIMA Test		LRD?
Direction	Traffic	(%)	*H*	Error	*d*	Error	
Incoming	Control	0.7	--	**1.6E-03**	--	**1.9E-03**	**No**
	Data	86.3	0.86	2.1E-04	0.37	1.2E-04	yes
	whole	87.0	0.86	2.1E-04	0.37	1.2E-04	yes
Outgoing	Control	2.9	0.85	2.9E-04	0.36	3.3E-04	yes
	Data	10.1	0.9	4.8E-04	0.4	6.5E-04	yes
	whole	13.0	0.9	4.7E-04	0.4	6.2E-04	yes
Bidirectional	Control	3.6	0.86	8.0E-04	0.36	9.5E-04	yes
	Data	96.4	0.86	1.8E-04	0.37	1.4E-04	yes
	WHOLE	100.0	0.86	1.8E-04	0.37	1.3E-04	yes

TABLE 7. LRD behavior analysis of byte count sequences in abnormal duration of tuesday's traffic trace (10:00:00 – 12:00:00) using window size of 15 minutes and aggregation interval of 0.5 second.

Decomposed Traffic		Vol.	SOSS Test		FARIMA Test		LRD?
Direction	Traffic	(%)	*H*	Error	*d*	Error	
Incoming	Control	0.6	0.9	8.7E-04	0.41	9.9E-04	yes
	Data	89.6	0.82	5.4E-04	0.33	4.3E-04	yes
	whole	90.2	0.82	5.4E-04	0.33	4.3E-04	yes
Outgoing	Control	2.9	0.87	9.6E-04	--	1.1E-03	yes
	Data	6.9	0.76	7.5E-04	0.27	7.7E-04	yes
	whole	9.8	0.75	7.6E-04	0.26	7.8E-04	yes
Bidirectional	Control	3.5	--	**1.1E-03**	--	**1.3E-03**	**No**
	Data	96.5	0.81	5.7E-04	0.32	4.5E-04	yes
	WHOLE	100.0	0.81	5.6E-04	0.32	4.5E-04	Yes

TABLE 8. LRD behavior analysis of packet count sequences in malicious duration of monday's traffic trace (14:00:00 – 16:00:00) using window size of 30 minutes and aggregation interval of 1 second.

Direction	Traffic	Vol. (%)	SOSS Test		FARIMA Test		LRD?
			H	Error	*d*	Error	
Incoming	Control	6	--	**8.8E-03**	--	**9.4E-03**	**No**
	Data	54	0.92	1.1E-04	0.43	1.3E-04	yes
	whole	60	0.93	1.6E-04	0.43	2.2E-04	yes
Outgoing	Control	33	0.92	7.5E-04	0.42	8.7E-04	yes
	Data	7	0.93	2.9E-04	0.43	2.6E-04	yes
	whole	40	0.93	4.1E-04	0.43	6.7E-04	yes
Bidirectional	Control	39	--	**1.9E-03**	--	**2.2E-03**	**No**
	Data	61	0.93	7.6E-05	0.44	1.6E-04	yes
	WHOLE	100	0.93	2.0E-04	0.43	4.1E-04	yes

TABLE 10. LRD behavior analysis of packet count sequences in malicious duration of wednesday's traffic trace (12:00:00 – 14:00:00) using window size of 30 minutes and aggregation interval of 1 second.

Direction	Traffic	Vol. (%)	SOSS Test		FARIMA Test		LRD?
			H	Error	*d*	Error	
Incoming	Control	7	--	**6.9E-03**	--	**7.5E-03**	**No**
	Data	52	0.90	2.3E-04	0.41	2.3E-04	yes
	whole	59	0.90	7.5E-04	0.41	8.6E-04	yes
Outgoing	Control	33	--	**1.9E-03**	--	**2.2E-03**	**No**
	Data	8	0.94	8.3E-05	0.45	8.8E-05	yes
	whole	41	0.91	9.5E-04	0.41	1.2E-03	yes
Bidirectional	Control	40	--	**4.4E-03**	--	**4.9E-03**	**No**
	Data	60	0.91	1.9E-04	0.42	2.0E-04	yes
	WHOLE	100	0.91	8.8E-04	--	1.0E-03	yes

the outgoing data and whole traffic also fail to exhibit LRD behavior. Our extensive LRD behavior analysis of decomposed network traffic reveals that despite the similarities in the results of packet and byte count features, the packet count feature is a better choice for LRD based anomaly detection using SOSS and FARIMA models. This is because the byte count exhibits more variations in the results due to its high dependency on the underlying application used.

Finally, the validation of the proposed method is demonstrated through investigating the detected anomalies in the traffic. Since according to the proposed method, if an anomalous event is detected, then it is further investigated to identify the root cause of the event as shown in Figure 3. Therefore, we dug in the aforementioned anomalous traffic and did packet-by-packet header inspection. Here we only present the results for Wednesday and Monday's anomalous traffic traces because they contain malicious packets. Since the anomalies are mainly detected in the control plane traffic, therefore, we filtered out the control traffic in the concern anomalous durations of the aforementioned traffic traces. After looking at the packets' header information in each duration, we found that most of the packets are exchanged between two hosts, i.e., an internal host and an external host (inside and outside KSU's network). Consequently, we further narrow down the analysis by filtering the traffic between these two hosts. Then by investigating the TCP port and sequence numbers in the header of each packet in the concern traffic, we observed that the anomalies in Wednesday and Monday's traffic traces are due to Acknowledgements-storm (Ack-storm) denial of service (DoS attacks). It is a type of low-rate DoS attack that is induced by the exploitation of a design flaw in the TCP protoc ol specifications. According to [66], an Ack-storm DoS attack occurs when a host receives a packet with an acknowledgement number field (the receiver's sequence number) larger than the one expected (acknowledgement to the packet that is not yet sent), then the host discards this packet and resends the last ACK packet. In this way, the two hosts are trapped in an infinite loop of sending and receiving bare ACKs back and forth. This loop continues when both the hosts keep receiving the packets. It only stops, either when the packets are dropped, or when a reset (RST) packet aborts the connection.

Furthermore, we observed that around 358,664 bare ACKs packets were exchanged between the two hosts in the case Wednesday' malicious event, which mounted the Ack-storm DoS attach that lasted for around 20 seconds. This huge volume of packets generated anomaly in the corresponding control plane traffic as can be seen in the form of a spike in Figures 8 (d) and 9 (d). Such malicious traffic causes the LRD behavior failure in the corresponding control traffic as

TABLE 9. LRD behavior analysis of byte count sequences in malicious duration of monday's traffic trace (14:00:00 – 16:00:00) using window size of 30 minutes and aggregation interval of 1 second.

Direction	Traffic	Vol. (%)	SOSS Test		FARIMA Test		LRD?
			H	Error	*d*	Error	
Incoming	Control	0.5	--	**8.8E-03**	--	**9.4E-03**	**No**
	Data	91.7	0.92	6.0E-05	0.43	1.2E-04	yes
	whole	92.2	0.92	6.7E-05	0.43	1.1E-04	yes
Outgoing	Control	2.8	0.92	7.4E-04	0.42	8.6E-04	yes
	Data	5.0	--	**2.0E-03**	--	**2.3E-03**	**No**
	whole	7.8	--	**1.1E-03**	--	**1.3E-03**	**No**
Bidirectional	Control	3.3	--	**1.9E-03**	--	**2.1E-03**	**No**
	Data	96.7	0.92	1.3E-04	0.43	7.3E-05	yes
	WHOLE	100.0	0.92	1.6E-04	0.43	7.5E-05	yes

TABLE 11. LRD behavior analysis of byte count sequences in malicious duration of wednesday's traffic trace (12:00:00 – 14:00:00) using window size of 30 minutes and aggregation interval of 1 second.

Direction	Traffic	Vol. (%)	SOSS Test		FARIMA Test		LRD?
			H	Error	*d*	Error	
Incoming	Control	0.6	--	**6.8E-03**	--	**7.4E-03**	**No**
	Data	89.5	0.91	2.3E-04	0.41	2.9E-04	yes
	whole	90.1	0.91	2.3E-04	0.41	3.2E-04	yes
Outgoing	Control	2.8	--	**1.9E-03**	--	**2.2E-03**	**No**
	Data	7.1	0.93	1.8E-04	0.44	9.2E-05	yes
	whole	9.9	0.93	1.9E-04	0.44	9.9E-05	yes
Bidirectional	Control	3.4	--	**4.4E-03**	--	**4.8E-03**	**No**
	Data	96.6	0.91	2.4E-04	0.42	2.5E-04	yes
	WHOLE	100.0	0.91	2.6E-04	0.42	2.7E-04	yes

TABLE 12. Summary and comparison of the results of packet and byte count features in decomposed network traffic and WHOLE traffic.

Decomposed Traffic		Sunday's Abnormal Traffic		Tuesday's Abnormal Traffic		Monday's Malicious Traffic		Wednesday's Malicious Traffic	
Direction	Traffic	Packet Count	Byte Count	Packet Count	Byte Count	Packet Count	Byte Count	Packet Count	Byte Count
Incoming	Control	✓	✓	×	×	✓	✓	✓	✓
	Data	×	×	×	×	×	×	×	×
	Whole	×	×	×	×	×	×	×	×
Outgoing	Control	×	×	×	×	×	×	✓	✓
	Data	×	×	×	×	×	✓	×	×
	Whole	×	×	×	×	×	✓	×	×
Bidirectional	Control	×	×	✓	✓	✓	✓	✓	✓
	Data	×	×	×	×	×	×	×	×
	WHOLE	×	×	×	×	×	×	×	×

can be seen from Tables 10 – 12. Likewise, we observed that the attack in Monday's traffic trace lasted for about 23 seconds as shown in Figures 8 (c) and 9 (c) in the form of a sudden surge in control plane traffic, which causes failure of LRD behavior in the corresponding traffic as can be seen from Tables 8 – 9 and Table 12. In order to validate these findings, we further analyze the behavior of traffic when the attack packets are removed. Thus, we filter out the corresponding malicious packets of Ack-storm DoS attack from the packets of normal traffic in Wednesday's traffic trace as shown in Figure 10. The left plot in Figure 10 shows bidirectional control and data plane traffic without filtering the malicious traffic, whereas the right plot shows the corresponding subgroups after filtering the malicious packets. It can be seen from the Figure that after filtering out the malicious packets, the spike in the control plane vanishes and the control and data planes start to exhibit similar time variations. Similar results are observed for Monday's malicious traffic.

The experimental observations show that the presence of short duration and low volume abnormalities and attacks in the control traffic affect the LRD behavior of the aggregated control plane traffic. However, the same is not found in the case of overall aggregated WHOLE traffic. This shows the detection efficacy of the proposed method as compared to the methods that only analyze aggregated WHOLE traffic without decomposing it into control and data planes. Thus, the proposed method can be an efficient platform for online threats detection, since it narrows down the analysis to control and data planes of the traffic in different directions with respect to the enterprise network.

VI. CONCLUSION AND FUTURE WORK

In this paper, we present and empirically evaluate an effective volume based anomaly detection method that analyzes the LRD behavior in Internet traffic through both packet and byte counts of control and data planes traffic separately. Through experiments on real dataset, we observe that the proposed method efficiently detects anomalies such as low volume attacks and abnormalities. However, the same is not observed when the traffic is not decomposed into control and data planes traffic. This means that such anomalies might not be

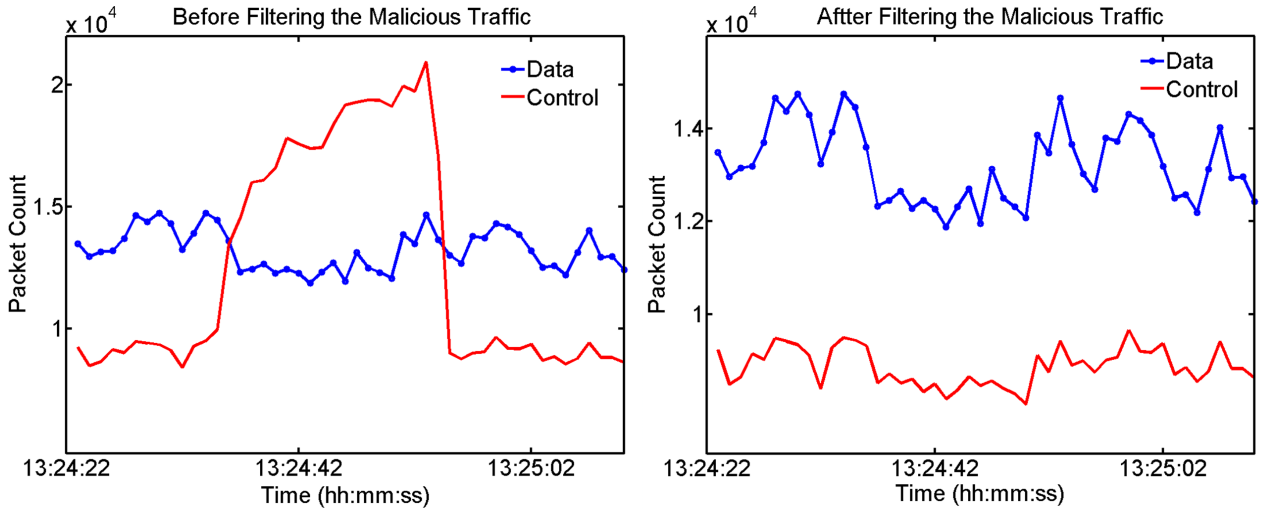

FIGURE 10. Packet count in bidirectional control and data planes traffic of Wednesday's malicious traffic trace before and after removing the malicious traffic.

detected by those LRD model based methods, which only consider the aggregated WHOLE traffic without decomposing it into control and data planes. Since we observe that in the presence of an anomalous behavior in the control plane of the traffic, the aggregated control plane traffic fails to exhibit LRD behavior, whereas the aggregated WHOLE traffic still exhibits LRD behavior. This is because the control plane traffic constitutes a small percentage of the WHOLE traffic; therefore, the effect of anomaly in the control traffic is not necessarily high enough to cause the LRD failure in the aggregated WHOLE traffic. Hence, the effect of anomaly is overshadowed when looking at the aggregated WHOLE traffic. In addition, the decomposition of network traffic results in reducing the volume of some of the individual traffic planes, namely the incoming control and outgoing data compared to the aggregated WHOLE traffic, i.e., bidirectional whole traffic. This provides a fine granular analysis platform in terms of parallel observations of different traffic subgroups of control and data for online analysis at the cost of few prior extra computations during decomposition process. We experimentally demonstrate and validate such scenario by analyzing malicious traffic that contain the Acknowledgement storm (Ack-storm) DoS attacks.

Additionally, since we analyze the statistical behavior of aggregated network traffic; therefore, we show the importance of aggregation interval and window size in the analysis of LRD behavior of network traffic, particularly for anomalies detection. Through extensive experiments on real Internet traffic dataset using various aggregation intervals and different window sizes, we show that the aggregation intervals of 0.5 second and 1 second and window sizes of 15 minutes and 30 minutes, respectively are appropriate for capturing LRD behavior in the underlying traffic. These aggregation intervals are suitable for the traffic rate of at least 1 kilo packets per second (KPPS). We notice that these aggregation intervals and window sizes are large enough to capture the LRD behavior in traffic and small enough to detect low volume and short duration anomalies as well as to avoid the non-stationarity effect.

In our future work, we aim for the analysis of inter-arrival time of the decomposed traffic in comparison to packet and byte counts. We will also insert different types of generated attacks and abnormalities into the control traffic of real Internet traffic for further investigation. This will enable us to calculate the accuracy, false positives and false negatives of the proposed system. In addition, we will perform a comparative analysis of the proposed scheme with other LRD based schemes.

REFERENCES

[1] Cisco VNI, "The zettabyte era—Trends and analysis," Cisco, San Jose, CA, USA, White Paper 1465272001812119, Jun. 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, p. 15, 2009.

[3] J. Mazel, P. Casas, R. Fontugne, K. Fukuda, and P. Owezarski, "Hunting attacks in the dark: Clustering and correlation analysis for unsupervised anomaly detection," *Int. J. Netw. Manage.*, vol. 25, no. 5, pp. 283–305, 2015.

[4] Arbor Network the Security Division of NETSCOUT. (2016) *11th Annual Worldwide Infrastructure Security Report (WISR)*. [Online]. Available: https://www.arbornetworks.com/images/documents/WISR2016_EN_Web.pdf

[5] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in *Proc. 2nd ACM SIGCOMM Workshop Internet Meas.*, Nov. 2002, pp. 71–82.

[6] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[7] B. AsSadhan, H. Kim, J. M. F. Moura, and X. Wang, "Network traffic behavior analysis by decomposition into control and data planes," in *Proc. 4th Int. Workshop Secur. Syst. Netw. (SSN)*, Miami, FL, USA, Apr. 2008, pp. 1–8.

[8] B. AsSadhan and J. M. F. Moura, "An efficient method to detect periodic behavior in botnet traffic by analyzing control plane traffic," *J. Adv. Res.*, vol. 5, no. 4, pp. 435–448, 2014.

[9] K. Zeb, B. AsSadhan, J. Al-Muhtadi, S. Alshebeili, and A. Bashaiwth, "Volume based anomaly detection using LRD analysis of decomposed network traffic," in *Proc. 4th IEEE Int. Conf. Innov. Comput. Technol. (INTECH)*, Luton, U.K., Aug. 2014, pp. 52–57.

[10] A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," *Dokl. Akad. Nauk SSSR*, vol. 30, no. 4, pp. 299–303, 1941.

[11] H. Von Koch, "Sur une courbe continue sans tangente, obtenue par une construction géométrique élémentaire," *Arkiv Matematik*, vol. 1, pp. 681–704, Oct. 1904.

[12] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-range dependence ten years of Internet traffic modeling," *IEEE Internet Comput.*, vol. 8, no. 5, pp. 57–64, Sep. 2004.

[13] H. E. Hurst, "Long-term storage capacity of reservoirs," *Trans. Amer. Soc. Civil Eng.*, vol. 116, pp. 770–808, 1951.

[14] W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, vol. 23. Cambridge, MA, USA: Birkhauser Boston Inc., 1998, pp. 27–53. [Online]. Available: http://dl.acm.org/citation.cfm?id=292595.292597

[15] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*. Hoboken, NJ, USA: Wiley, 2000.

[16] H. Kettani and J. A. Gubner, "A novel approach to the estimation of the long-range dependence parameter," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 6, pp. 463–467, Jun. 2006.

[17] D. Rincón, G. Ferrer, and X. Hernández, "Self-similar traffic in a commercial video on-demand system," in *Proc. 6th Open Eur. Summer School (EUNICE)*, Enschede, The Netherlands, 2000, pp. 181–188. [Online]. Available: https://www.simpleweb.org/ifip/Conferences/EUNICE/2000/paper8-2.pdf

[18] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[19] O. I. Sheluhin, S. M. Smolskiy, and A. V. Osin, *Self-Similar Processes in Telecommunications*. Hoboken, NJ, USA: Wiley, 2007.

[20] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York, NY, USA: Freeman, 1983, p. 173.

[21] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *J. Time Ser. Anal.*, vol. 1, no. 1, pp. 15–29, 1980.

[22] G. Samorodnitsky, "Long range dependence," *Found. Trends Stochastic Syst.*, vol. 1, no. 3, pp. 163–257, 2006.

[23] C. Park *et al.*, "Long-range dependence analysis of Internet traffic," *J. Appl. Statist.*, vol. 38, no. 7, pp. 1407–1433, 2011.

[24] P. Dymora, M. Mazurek, and D. Strzałka, "Computer network traffic analysis with the use of statistical self-similarity factor," *Ann. UMCS, Inf.*, vol. 13, no. 2, pp. 69–81, 2013.

[25] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, Jun. 1995.

[26] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[27] J. Domańska, A. Domańska, and T. Czachórski, "A few investigations of long-range dependence in network traffic," in *Information Sciences and Systems*. 2014, pp. 137–144.
[28] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Trans. Netw.*, vol. 5, no. 1, pp. 71–86, Feb. 1997.
[29] L. Zhang, Z. Zhu, and J. S. Marron, "Multiresolution anomaly detection method for long range dependent time series," UNC Dept. Statist. Oper. Res., Chapel Hill, NC, USA, Tech. Rep. UNC/STOR/07/12, 2008. [Online]. Available: http://stat-or-old.oasis.unc.edu/research/Current%20Reports/techpdf/ZhangZhuMarron07.pdf
[30] L. Yao, M. Agapie, J. Ganbar, and M. Doroslovacki, "Long range dependence in Internet backbone traffic," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 3. May 2003, pp. 1611–1615.
[31] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz, "The changing nature of network traffic: Scaling phenomena," *SIGCOMM Comput. Commun. Rev.*, vol. 28, no. 2, pp. 5–29, 1988.
[32] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido, "A nonstationary Poisson view of Internet traffic," in *Proc. 23rd Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 3. Mar. 2004, pp. 1558–1569.
[33] M. Y. Idris, H. Abdullah, and M. A. Maarof, "Iterative window size estimation on self-similarity measurement for network traffic anomaly detection," *Int. J. Comput. Inf. Sci.*, vol. 2, no. 2, pp. 83–91, 2004.
[34] M. F. Rohani, M. A. Maarof, A. Selamat, and H. Kettani, "Uncovering anomaly traffic based on loss of self-Similarity behavior using second order statistical model," *Int. J. Comput. Sci. Netw. Secur.*, vol. 7, no. 9, pp. 116–122, 2007.
[35] H. Kettani and J. A. Gubner, "On the detection of LRD phenomena," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, 2012, pp. 320–326.
[36] H. Sheng, Y. Chen, and T. Qiu, *Fractional Processes and Fractional-Order Signal Processing: Techniques and Applications*. London, U.K.: Springer-Verlag, 2012.
[37] M. F. Rohani, M. A. Maarof, A. Selamat, and H. Kettani, "LoSS detection approach based on ESOSS and ASOSS models," in *Proc. 4th Int. Conf. Inf. Assurance Secur. (ISIAS)*, 2008, pp. 192–197.
[38] M. F. Rohani, M. A. Maarof, A. Selamat, and H. Kettani, "Multi-level sampling approach for continous loss detection using iterative window and statistical model," *IIUM Eng. J.*, vol. 11, no. 2, pp. 137–149, 2010.
[39] M. Mazurek and P. Dymora, "Network anomaly detection based on the statistical self-similarity factor for HTTP protocol," *Przeglad Elektrotechniczny*, vol. R.90, no. 1, pp. 127–130, Jan. 2014. [Online]. Available: https://www.infona.pl/resource/bwmeta1.element.baztech-6e4c65a5-e2e5-4b10-bafd-211e50c7e9ef
[40] G. Kaur, V. Saxena, and J. P. Gupta, "A novel multi scale approach for detecting high bandwidth aggregates in network traffic," *Int. J. Secur. Appl.*, vol. 7, no. 5, pp. 81–100, 2013.
[41] L. Zhang, Z. Zhu, and J. S. Marron, "Multiresolution anomaly detection method for fractional Gaussian noise," *J. Appl. Statist.*, vol. 41, no. 4, pp. 769–784, 2014.
[42] J. R. Lee, H.-D. Jeong, D. McNickle, and K. Pawlikowski, "Self-similar properties of spam," in *Proc. 5th Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput. (IMIS)*, 2011, pp. 347–352.
[43] P. R. Iácio, M. M. Freire, M. Pereira, and P. P. Monteiro, "Analysis of the impact of intensive attacks on the self-similarity degree of the network traffic," in *Proc. 2nd Int. Conf. Emerg. Secur. Inf., Syst. Technol. (SECURWARE)*, 2008, pp. 107–113.
[44] R. Yan and Y. Wang, "Hurst parameter for security evaluation of LAN traffic," *Inf. Technol. J.*, vol. 11, no. 2, pp. 269–275, 2012.
[45] X. Ye, J. Lan, and W. Huang, "Network traffic anomaly detection based on self-similarity using FRFT," in *Proc. 4th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, 2013, pp. 837–840.
[46] X. Wang, L. Pang, Q. Pei, and X. Li, "A scheme for fast network traffic anomaly detection," in *Proc. Int. Conf. Comput. Appl. Syst. Modeling (ICCASM)*, 2010, pp. V1-592–V1-596.
[47] H. Jiang and L. Pang, "Fast network traffic anomaly detection based on iteration," in *Proc. 7th Int. Conf. Comput. Intell. Secur. (CIS)*, 2011, pp. 1006–1010.
[48] J.-S. R. Lee, S.-K. Ye, and H.-D. J. Jeong, "Detecting anomaly teletraffic using stochastic self-similarity based on Hadoop," in *Proc. 16th Int. Conf. Netw.-Based Inf. Syst. (NBiS)*, 2013, pp. 282–287.
[49] J.-S. R. Lee, S.-K. Ye, and H.-D. J. Jeong, "*ATMSim*: An anomaly teletraffic detection measurement analysis simulator," *Simul. Model. Pract. Theory*, vol. 49, pp. 98–109, Dec. 2014.
[50] L. Liu, X. Jin, G. Min, and L. Xu, "Anomaly diagnosis based on regression and classification analysis of statistical traffic features," *Secur. Commun. Netw.*, vol. 7, no. 9, pp. 1372–1383, 2013.
[51] H. Nurohman and Y. Purwanto, "Traffic anomaly based detection: Anomaly detection by self-similar analysis," in *Proc. Int. Conf. Control, Electron., Renew. Energy Commun. (ICCEREC)*, 2015, pp. 1–6.
[52] Z. Jian-Qi, F. Feng, C.-K. Kim, Y. Ke-Xin, and L. Yan-Heng, "A DoS detection method based on composition self-similarity," *KSII Trans. Internet Inf. Syst.*, vol. 6, no. 5, pp. 1463–1478, 2012.
[53] Z. Xia, S. Lu, and J. Li, "DDoS flood attack detection based on fractal parameters," in *Proc. 8th Int. Conf. Wireless Commun., Netw. Mobile Comput. (WiCOM)*, 2012, pp. 1–5.
[54] Z. Xia, S. Lu, J. Li, and J. Tang, "Enhancing DDoS flood attack detection via intelligent fuzzy logic," *Informatica*, vol. 34, no. 4, pp. 497–507, 2010.
[55] A. Takahashi, R. Igarashi, H. Ueda, Y. Iwaya, and T. Kinoshita, "Network anomaly detection based on R/S pox diagram," *Int. J. Soc. Mater. Eng. Resour.*, vol. 17, no. 2, pp. 186–192, 2010.
[56] L. Liang Fu, H. Mao Lin, M. A. Orgun, and Z. Jia-Wan, "An improved wavelet analysis method for detecting DDoS attacks," in *Proc. 4th Int. Conf. Netw. Syst. Secur. (NSS)*, 2010, pp. 318–322.
[57] J. Zhang, Q. Gao, H. Wang, and H. Wang, "Detecting anomalies from high-dimensional wireless network data streams: A case study," *Soft Comput.*, vol. 15, no. 6, pp. 1195–1215, 2011.
[58] R. Ul Islam, M. S. Hossain, and K. Andersson, "A novel anomaly detection algorithm for sensor data under uncertainty," *Soft Comput.*, pp. 1–17, Nov. 2016, doi: 10.1007/s00500-016-2425-2. [Online]. Available: https://link.springer.com/article/10.1007/s00500-016-2425-2
[59] U. Fiore, F. Palmieri, A. Castiglione, and A. D. Santis, "Network anomaly detection with the restricted Boltzmann machine," *Neurocomputing*, vol. 122, pp. 13–23, Jul. 2013.
[60] M. Bi, J. Xu, M. Wang, and F. Zhou, "Anomaly detection model of user behavior based on principal component analysis," *J. Ambient Intell. Humanized Comput.*, vol. 7, no. 4, pp. 547–554, Aug. 2016.
[61] Y. Gu, A. McCallum, and D. Towsley, "Detecting anomalies in network traffic using maximum entropy estimation," in *Proc. 5th ACM SIGCOMM Conf. Internet Meas.*, Oct. 2005, p. 32.
[62] F. Palmieri and U. Fiore, "Network anomaly detection through nonlinear analysis," *Comput. Secur.*, vol. 29, no. 7, pp. 737–755, Oct. 2010.
[63] F. Palmieri, U. Fiore, and A. Castiglione, "A distributed approach to network anomaly detection based on independent component analysis," *Concurrency Comput., Pract. Exper.*, vol. 26, no. 5, pp. 1113–1129, Apr. 2014.
[64] B. AsSadhan, H. Kim, and J. Moura, "Long-range dependence analysis of control and data planes network traffic," presented at the Saudi Int. Innov. Conf. (SIIC), Leeds, U.K., 2008.
[65] A. Bashaiwth, "Network traffic analysis for botnet detection," M.S. thesis, Dept. Elect. Eng., College of Engineering King Saud Univ., Riyadh, Saudi Arabia, May 2015.
[66] R. Abramov and A. Herzberg, "TCP Ack storm DoS attacks," *Comput. Secur.*, vol. 33, pp. 12–27, Mar. 2013.

BASIL AsSADHAN received the M.S. degree in electrical and computer engineering from the University of Wisconsin and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University. He is currently an Assistant Professor with the Electrical Engineering Department, King Saud University. His research interests are in the areas of cybersecurity, network security and network traffic analysis, and anomaly detection.

KHAN ZEB received the B.Sc. degree in telecommunication engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2010, and the M.S. degree in electrical engineering from King Saud University, Riyadh, Saudi Arabia, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada. He was a Researcher with the Center of Excellence in Information Assurance, King Saud University, from Jan. 2015 to Dec. 2016. His research interests include cyber security, anomaly/intrusion detection, network traffic analysis, e-healthcare/m-healthcare systems, smart grids, IoTs, wireless networks, and optical communication.

JALAL AL-MUHTADI received the M.S. and Ph.D. degrees from the University of Illinois at Urbana–Champaign, USA. He is currently the Director of the Center of Excellence in Information Assurance, King Saud University, and also an Assistant Professor with the Computer Science Department, King Saud University. He has over 40 scientific publications in the areas of cybersecurity, information assurance, and Internet of Things.

SALEH ALSHEBEILI was the Chairman of the Electrical Engineering Department, King Saud University, from 2001 to 2005. He has over 25 years of teaching and research experience in the area of communications and signal processing. He was a member of the Board of Directors of the King Abdullah Institute for Research and Consulting Studies from 2007 to 2009, a member of the Board of Directors of the Prince Sultan Advanced Technologies Research Institute (PSATRI) from 2008 to 2017, the Managing Director of PSATRI from 2008 to 2011, and the Director of the Saudi-Telecom Research Chair from 2008 to 2012. He has been a Co-Founder and the Director of the Technology Innovation Center, RF and Photonics for the e-Society, funded by King Abdulaziz City for Science and Technology (KACST), since 2011. He is currently a Professor with the Electrical Engineering Department, King Saud University. He was on the Editorial Board of the *Journal of Engineering Sciences of King Saud University* from 2009 to 2012. He has also actively involved in the review process of a number of research journals, KACST general directorate grants programs, and national and international symposiums and conferences.