International Journal of Innovative Computing, Information and Control
Volume **15**, Number **4**, August **2019**

ICIC International ©2019 ISSN 1349-4198
pp. **1323–1336**

TIME AND FREQUENCY COMPONENTS ANALYSIS OF NETWORK TRAFFIC DATA USING CONTINUOUS WAVELET TRANSFORM TO DETECT ANOMALIES

Mohammed Alharbi[1] and Marwan Ali Albahar[2]

[1]School of Computer Science
University of the Incarnate Word
4301 Broadway St, San Antonio, TX 78209, USA
m7medfahd@hotmail.com

[2]School of Management Science
Ibn Rushd College
King Abdul Aziz Rd., Abha, P.O. Box 447, Saudi Arabia
marwanalialbahar@gmail.com

Received November 2018; revised March 2019

Abstract. *There are various host-based methods and network-based methods to monitor network intrusions in real time, but they are limited in the context of identifying anomalies activities in the network. In this research study, in order to boost security in network intrusion systems, one method is to apply signal processing strategies which include powerful continuous wavelet transform methods that consist of different mother wavelets to detect any anomalies in network site traffic data. The percentage deviation metric was used to assess the quality of performance of the wavelets in detecting anomalous network activities such as brute force, port scan and DoS attacks. Results obtained from the analysis showed that Morlet wavelet performed better than the other implemented wavelets for detecting anomalies in traffic signal data based on the lowest percentage deviation value.*

Keywords: Continuous wavelet transform (CWT), Network intrusion anomalies, Percentage deviation, Signal processing

1. **Introduction.** Soon after the seminal report published by Anderson, intrusion detection has been explicitly discussed in [1]. Conventionally, the techniques of the intrusion detection had been categorized into anomaly detection and misuse detection. It is assumed that a majority of attacks leave behind a set of signatures within the network stream packets or an audit trail. Misuse detection indicates that attacks are often detectable by identifying these signatures or assessing the behaviors of the network traffic or the audit trails. On the other hand, the approach of misuse detection is confined to recent known attacks. One of the daunting challenges of misuse detection is to identify the new attacks or variations within the known attacks. Hence, to overcome the limitations of the misuse detection, Denning formulated the concept of anomaly detection within his seminal report [2]. The assumptions of Denning were based on the fact that violations in the security have the potential to be detected if the abnormal system usage patterns accessed through the audit data are inspected. In this regard, the majority of the anomaly detection approaches are aimed to generate normal activity profiles by calculating different metrics, so intrusion is only detected once the behavior of the actual system deviates from the normal and original profiles. On the basis of the features of

DOI: 10.24507/ijicic.15.04.1323

the monitored sources, there are two types of anomaly detection: network-based and the host-based. Generally, the host-based anomaly detection system works on the local monitored host and makes use of the audit trail data or the log files as the primary source of obtaining information. However, the drawback of the host-based anomaly is the inability to identify coordinated and distributed attacks by reflecting on the patterns in the network traffic. In contrast, the network-based anomaly technique is designed to safeguard the entire network against the intrusion through consistent monitoring of traffic on the network in specific sensors or designed hosts. Thereby ensure the protection of a large number of computers running simultaneously on different operating systems (OS) against different kinds of remote attacks: distributed denial-of-service attacks (DoS), port scans, and propagation of computer worms, which are potential threats to the current infrastructure of the Internet. This research paper presents the development of an intrusion detection system which is based on the continuous wavelet transform technique. The intrusion detection technique utilized a percentage of deviation metric to assess the quality of performance of the wavelets in detecting the anomalous network activities such as port scan, DoS attacks, and brute force.

***Contribution*:** The main contributions of this paper are as follows.

1- The first contribution is to design a real-time wavelet analysis of the network traffic and assess multiple wavelets based on their performance to detect anomalies in the network such as port scans, floods, and DoS.
2- The second contribution of this research is to analyze the wavelet filters to utilize the percentage deviation metric for assessing the wavelet filters once applied on the datasets comprising a multitude of the anomalies.

2. **Wavelet Analysis – Continuous Wavelet Transform.** Wavelet analysis is a powerful mathematical method to decompose data such as signals as well as images hierarchically [3]. This method can be employed to extract important information from different data types at different resolution levels. The wavelet transforms allow a particular function to be depicted using a rough general trend of the original signal and this is followed by a family of detailed complementary shapes or information that represent add-ons to the rough general shape of the original signal. This complementary information is required to get the transformed original data by adding one level to another until the best resolution level is reached. Technically speaking, the wavelet transform is mathematical algorithms that are utilized to divide a continuous-time signal into various scale components or frequency components [3]. The wavelet coding techniques are suitable for many practical applications where acceptable degradation and scalability are crucial because of the multi-resolution facet, which is inherited in algorithm. So far, wavelet analysis has achieved worldwide acceptance, especially in signal processing. Hence, continuous wavelet transforms were employed to analyze network traffic data various and popularly used mother wavelet methods that are (i) Haar, (ii) Morlet, (iii) Coiflet, (iv) Daubechies, (v) Mexican Hat wavelets.

Continuous wavelet transform (CWT) is actually a type of transform that is considered redundant and the coefficients that are produced from the continuous wavelet transform depend on the wavelet type. Moreover, these coefficients that are generated are quite difficult to interpret or understand. In this section, one simple synthetic signal (See Figure 1) was produced using Matlab programming language to demonstrate the interpretation of the CWT coefficients and the common mother wavelet (Haar wavelet) is applied to the signal at different scales. Wavelets are able to detect a discontinuity or any singularity in the signal. In other words, any abrupt changes inside the signal indicate wavelet coefficients with large absolute values.

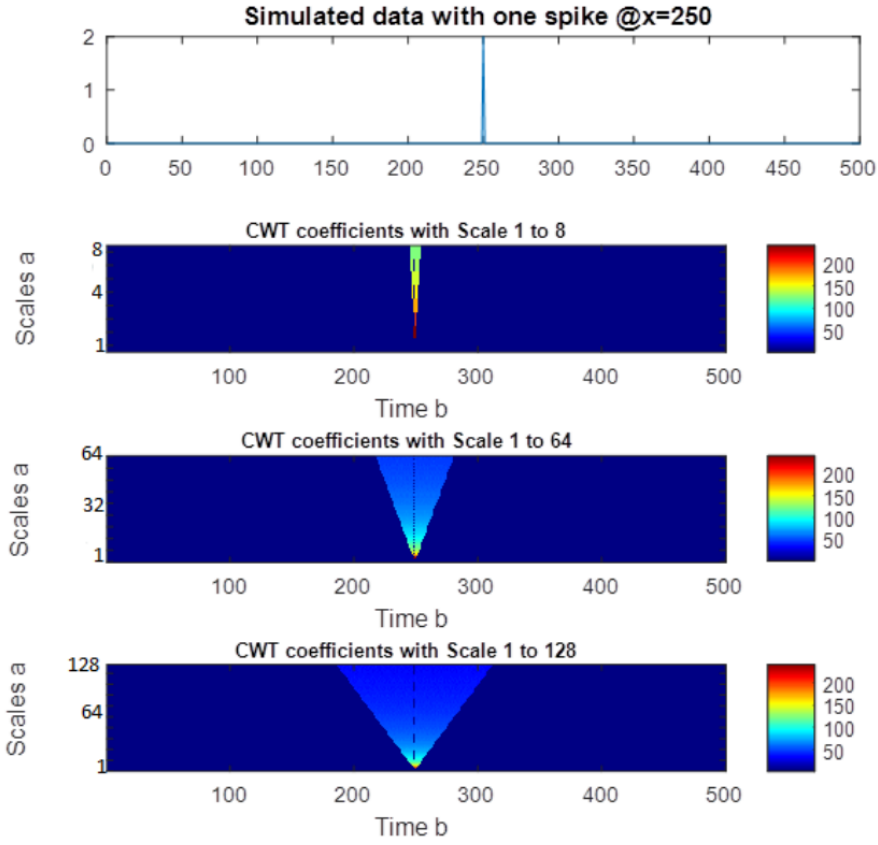


FIGURE 1. (color online) An example of the continuous wavelet transform applied on a synthetic data with one spike at $x = 250$ using different maximum number of scales (8, 64 and 128)

3. **Related Work.** For the detection of network intrusion a wavelet analysis approach is commonly used due to its characteristic time-frequency property which promotes signal splitting into variant components at multiple frequencies [4,5]. [6] suggested a technique for anomaly-based network intrusion detection which used the relations that are inherent among the wavelet coefficients of a self-similar function in a distinguished way and the technique showed the presence or absence of an anomaly as well as its location in the data. Also, it delivered the empirical outcomes on the KDD dataset to verify the proposed approach. A novel modeling technique was employed to study network signals by analyzing the network anomalies proposed in [7], where the authors combined the system identification theory with the wavelet approximation. In [8], authors proposed a wavelet based method, known as WIND to identify network problems and failures. Furthermore, the authors recommended a system by which the methods could be employed for plotting the values of RTT of a network path. While the authors in [9] highlighted that congestion on shared links in networks made use of the wavelets. It indicated a high correlation in those two paths that share the congested link between their one-way delays. Also, the authors pointed out that a Daubechies 6 wavelet has the highest correlation to congestion implementation and thus it was utilized as a mother wavelet for denoising of the wavelet. Additionally, in [10], the signal processing techniques were applied by authors in intrusion detection systems where a framework namely, Waveman, was also developed and implemented to get a wavelet-based real-time analysis of network traffic anomalies. In addition, the authors utilized two metrics (entropy and percentage deviation) to assess the performance of multiple wavelet functions while detecting multiple anomalies. The authors in

[11] proposed a wavelet-based technique that has the potential to identify network anomalies in real time. More specifically, the technique was based on wavelet analysis and integrated use of sketches to disclose the anomalies at the router level in the collected data. Likewise, the authors recommended a multi-timescale analysis to improve the detection rate. However, authors in [12] leveraged spectral analysis and statistical analysis techniques for network anomaly detection. Whereas in [13] the authors used multiple features in order to define the network traffic information, these sets of features were based on different metrics. They then proposed a novel detection mechanism of network traffic anomaly based on high-order statistical analysis and analytical discrete wavelet transform (ADWT). To validate this mechanism, the authors evaluated their method through real traffic dataset which was collected from a public server.

The motivation for this paper is to validate that the percentage deviation value can be employed to create a real-time network intrusion detection system, and also we intend to show that some of the wavelets have better performance when utilized in a real-time network intrusion detection system. Our objective is to detect the wavelet(s) that perform better than others through experiments conducted on the traffic datasets.

4. **Methodology.**

The repeated transformation in wavelet is the summation of signal multiplied by scaled, shifted version of wavelet over a complete time span of the signal. The wavelet coefficients are thus generated by this process that is a function of scale as well as position. The main steps that are involved in continuous wavelet transform are as follows.

(1) A particular wavelet is chosen, and it is then compared to a section at the beginning of the original signal.

(2) Then a number C is computed which represents how closely the wavelet is correlated to that section of the signal. The higher the value of C means that there is higher similarity between this wavelet and the signal.

(3) Then the wavelet is shifted or moved to the right of the original signal. The step (1) & step (2) are repeated until it covers the whole duration of the signal.

(4) Then the wavelet is scaled or stretched, and the steps (1) to (3) are repeated.

(5) The step (1) through step (4) for all scales are repeated.

After the application of this algorithm, wavelet coefficients are produced at different scales by different sections or portions of the signal. The coefficients represent the results of a regression of the original signal that is performed on the wavelets. The CWT coefficient plots are the time-scale view of the signal.

4.1. **Experimental setup – simulated network.**

The external server (in Figure 2), via the Internet, provides two services: (1) a file synchronization service (Seafile) and (2) a public webserver. In addition, numerous types of realistic network activities are stimulated on the client through parameterized and randomized python scripts in multiple scenarios. Additionally, the traffic data is recorded at the router, within the Openstack environment. The python scripts stimulate the user activities representing a typical workplace. To generate the malicious traffic a number of attacks are implemented within the virtual network. The network traffic was recorded at the external server's network so that data, which is generated, can be made as realistic as possible and the networked data is then combined with other traffic. Although the external servers allow the provision of above-mentioned services that are used by clients, it still becomes a victim of up-to-date and real attacks from the web. The preceding approach provides an opportunity to record malicious as well as normal traffic at the external server. The generated dataset, i.e., malicious and normal data are comprehensively discussed in the next section.

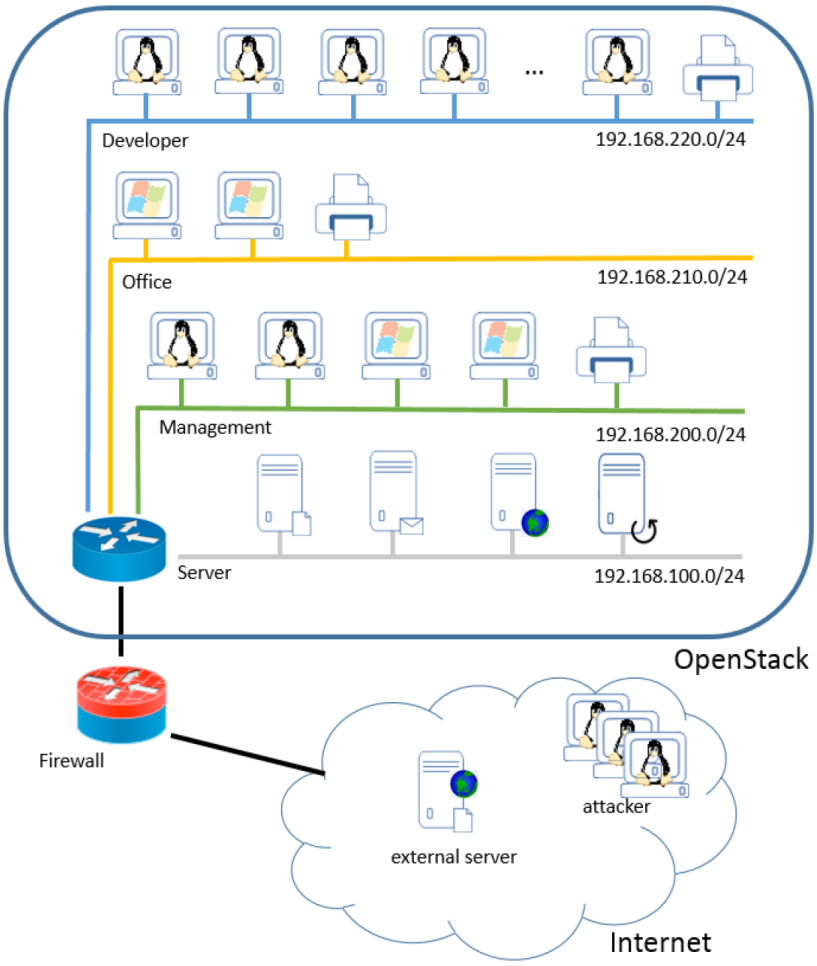


FIGURE 2. General idea of the simulated network anomalies at the external server [14]

4.2. **Network datasets.**

4.2.1. *Generation of normal traffic data.* For the fact that the scripts on the clients are used to copy typical computer consumer behaviour, so we can define two rules concerning the realistic flow of the network traffic. The practical imitation of user behaviour is considered under the first rule. In contrast, the second rule concentrates on the diversity of computer systems. The fulfilment of those rules requires the following characteristics in scripts:

(1) They should be able to run on the various operating systems
(2) They should take into consideration of the routine tasks of staff on computers
(3) Consider the various responsibilities as well as the working techniques of the staff
(4) No regular practice of computer user activities
(5) Consider regular functioning of computerized activities as well as breaks

The purpose of using python was to compose the end user behaviour script. The staff throughout the day deals with multiple task and duties, for example, preparing and giving the presentation, delivering documents via network shared printers, conducting (either small or private business searches) online, and composing the emails. In order to emulate these tasks concerning the potential varied qualities of staff members, every computer user possesses a configuration file. These setup files regulate the frequency and track each customer. Therefore, it is possible that different profiles are assigned to a varied computer user. However, in order to print and exchange the documents, it is

fundamental to reconfirm that size and types of corresponding documents are different. Moreover, the variety of attachment while sending the email messages must be observed.

It is noteworthy that, recognition of behavioural characteristics of a computer user is not possible just by assigning a list of activities every now and then. Instead, it is critical to randomize the temporal sequence of working activities of a computer user in addition to making changes in the type of work activities. However, the working activities must not be entirely illogical and stick to a possible division that is based on uniform functioning for several hours. Generally, the workers are not performing a job that generates required network traffic. Thereby, the scripts should concentrate on functioning hours to avoid tasks inside the staff's non-productive hours and nighttime. Centred on these requirements, the sets of information are quite reasonable, regardless of the fact they do not emulate computer user behaviour perfectly.

4.2.2. *Generation of malicious dataset.* A detailed realistic dataset normally consists of regular data and malicious data. The normal traffic is produced by the Python scripts, which were described in 4.2.1. The creation of the malicious data was based on two types of features. The first one, two clients do attacks internally that are port scans, denial of service, ping scans and brute force attacks utilizing the Linux software tool such as *nmap* and also Python scripts. The attacker's operating systems launch attack on the external server with Brute Force attacks and port scan attacks. In the meanwhile, the external server is accessible from the web, and hence it is exposed to real and latest web attacks. The main characteristics of the dataset, which was used in this research, are summarized in Table 1. The column in Table 1 provides the names or variable names in the dataset files.

The folder called 'traffic' of the CIDDS-001 dataset collected from the official website (see data website in the reference section) contains the ExternalServer folder which comprises various CSV files containing the network traffic data that were captured in an unidirectional Net-Flow format type. The names of these files that are found in these

Table 1. Main attributes of the CIDDS-001 dataset [14]

Column number	**Name**	**Description**
1	Src_IP	This is the address of the source IP
2	Src_Port	This represents the source port
3	Dest_IP	This represents the address of the destination IP
4	Dest_Port	This represents the Port for the destination
5	Proto-	This represents the Protocol used for transport such as UDP, TCP and ICMP
6	Date_first_seen	This represents the start time flow which is first seen
7	Duration	This variable represents the duration of the flow
8	Byte	These are the transmitted Bytes
9	Packet	These are transmitted Packets
10	Flag	This flag variable represents all the TCP flags for the OT concatenation
11	Class	This is the Class label such as attacker or normal or suspicious/unknown, or victim
12	AttackType	These represent the types of Attack used such as Brute Force, Denial of Service and port scan

sub-folders are built based on the format described as: *CIDDS-version-origin-period.csv.* All files start with CIDDS-001, and the data network traffic, which is recorded at the external server, is denoted as the external origin. The last part of that format is the *period*, which gives information about when the data traffic was recorded such as week 1, week 2, week 3 and week 4. For this research, we will focus on week 2, which has visible network anomalies, and signal processing algorithms will be applied to seeing how each mother wavelet performs for the analysis of the lengthy traffic signal of network data. Figure 3 describes the change in network flow, which is measured in *number of bytes* against the duration of week 2 network activity while Figure 4 describes the network traffic data measured in *number of bytes per hour* against the duration of the collected data in week 2. The presence of peaks and troughs in Figure 3 and Figure 4 happen because of the high activity or sudden high activity which causes high peaks, and besides the high peaks, corresponding troughs are formed which may be due to unusual anomalies in the network data.

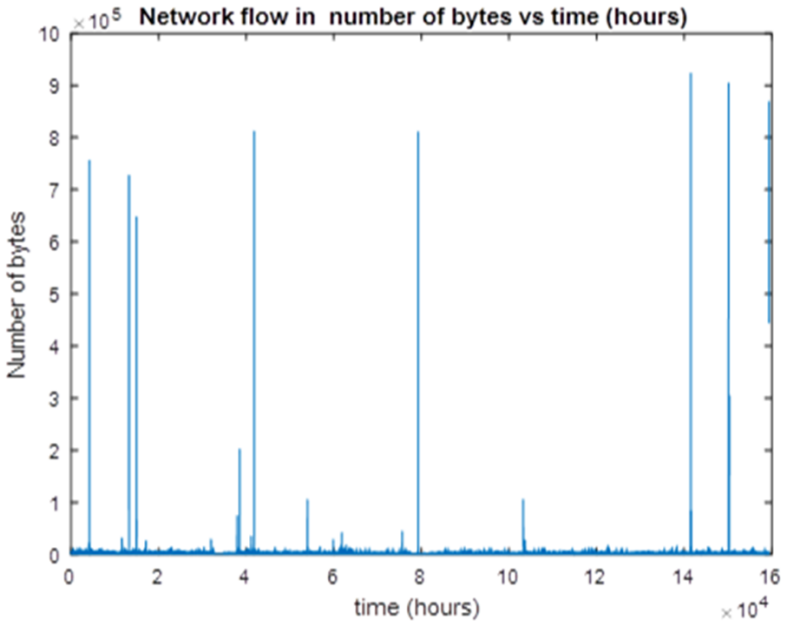


FIGURE 3. Network flow for week 2 with y-axis representing the **number of bytes** of the network activity against duration (x-axis) measured in hours (graphical plot from Matlab)

4.3. **Continuous wavelet transform – Morlet, Daubechies, Haar, Coiflet.**

4.3.1. *Wavelet function.* The function $\psi \in L^2(\mathbb{R})$ represents a wavelet with a mean value of zero. This function is represented by the mathematical expression $\int \psi = 0$ and this expression is normalized where $\|\psi\| = 1$, as well as it is located centrally at the region $t = 0$ [15]. The time-frequency family groups are defined based on the positive scaling of the mathematical function ψ with a scaling magnitude s, and a translation magnitude $u \in \mathbb{R}$.

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \quad \text{where } u \in \mathbb{R} \text{ and } s > 0 \tag{1}$$

Given a function $f \in L^2(\mathbb{R})$, the continuous wavelet transform of the function f at scale s and time u is:

$$Wf(u,s) :=< f, \psi_{u,s} \geq \int_{-\infty}^{+\infty} f(t)\psi^*_{u,s}(t)dt \tag{2}$$

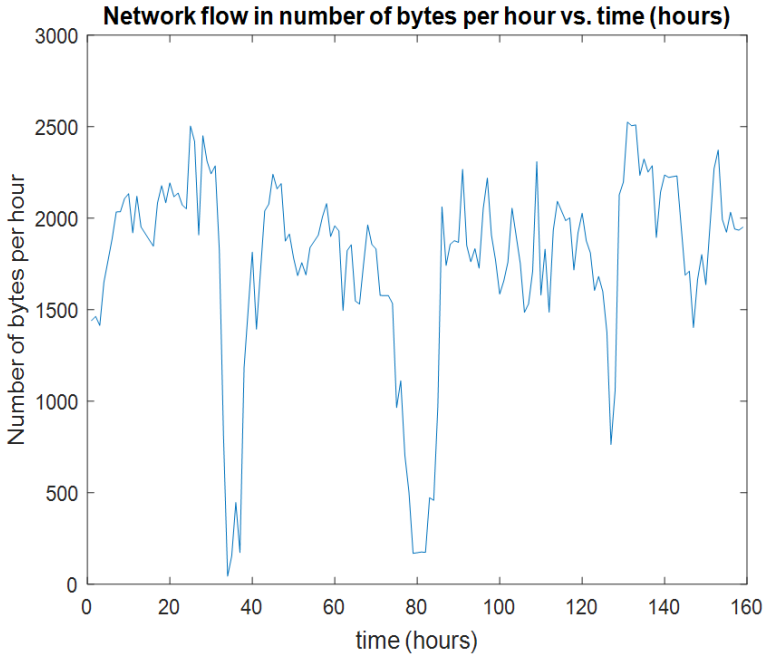


FIGURE 4. Network flow for week 2 whereby y-axis represents **the number of bytes per hour** vs. x-axis, which represents duration measured in hours (computed graphical plot from Matlab)

And Equation (2) provides the frequency component of the function f which corresponds to the time location t and the scale (s). The interesting evolution of the wavelet theory emanates from the fact that it controls two important parameters that are scale (s) and time (u) of the continuous wavelet transform (CWT) which makes it possible to investigate a signal both in time and frequency domains simultaneously and it focuses on the resolution.

4.3.2. *Continuous wavelet transform algorithm.* Currently, the CWT analysis is a well-known time frequency transformation. In mathematical terms, a wavelet series represents square-integrable complex-valued and real-valued function of orthonormal series produced by the wavelet function. Both the mother (base) wavelet and the original signal are required to perform CWT. The following mother wavelets are chosen for the network data analysis: (i) Haar, (ii) Morlet, (iii) Coiflet, (iv) Daubechies and (v) Mexican Hat.

For example, the equation for a Morlet wavelet is given by $\psi(t) = e^{i2\pi f_0 t} e^{-\left(\frac{\alpha t^2}{\beta^2}\right)}$. The CWT of a signal $x(t)$ is obtained by applying the inner product notation (Equation (3)):

$$CWT_x^{\psi}(\tau, s) =< x, \psi_{s,\tau} >= \frac{1}{\sqrt{s}} \int x(t)\psi^* \left(\frac{t-\tau}{s}\right) dt \tag{3}$$

The $CWT_x^{\psi}(\tau, s)$ represents the wavelet transform coefficients for a provided τ and s. The transformed signal $CWT_x^{\psi}(\tau, s)$ consists of two variables that are τ and s. The parameter s represents the scaling magnitude parameter and it finds the time as well as the frequency resolutions of the scaled mother wavelet function of $\psi\left(\frac{t-\tau}{s}\right)$.

In general, frequency is inversely proportional to the parameter values of s. In the meanwhile, the symbol τ represents the shifting variable, which translates the scaled wavelet through time. The symbol $\psi^*(.)$ represents the complex conjugation of the mother wavelet. For example, by including the parameters scale s and time τ, the Morlet wavelet

can be expressed as:

$$\psi\left(\frac{t-\tau}{s}\right) = e^{i2\pi f_0\left(\frac{t-\tau}{s}\right)} e^{-\alpha\frac{(t-\tau)^2}{s^2\beta^2}} \tag{4}$$

In order to produce the continuous wavelet transform, one can obtain the wavelet coefficients from Equation (3). The original wavelet is located at the start of the analyzed signal and then the scaling parameter s is set to 1. The wavelet function at this set scale is multiplied by the signal $x(t)$, and then integration is performed over the duration of the signal, and multiplied by the factor of $1/\sqrt{s}$. Consequently, the wavelet is moved through time to $t = \tau$, and at that point, the wavelet coefficient is then produced at time t with magnitude τ and scale s with magnitude 1. This procedure is repeated until it reaches the end of the signal. Following a similar procedure, the scale s is incremented by 1 and the coefficients are computed for all values for the parameter s. For every computation, a particular scale s fills a specific wavelet coefficients' row of the time scale graphic of the original signal (See Figure 1 for illustration). The various processes that are involved during the wavelet transform are depicted in Figure 5 with emphasis on scale s and time shift t. Figure 6 illustrates a 'zoomed' view on the characteristics of the Morlet wavelet, which would be exploited for depicting anomalies in the network data signal.

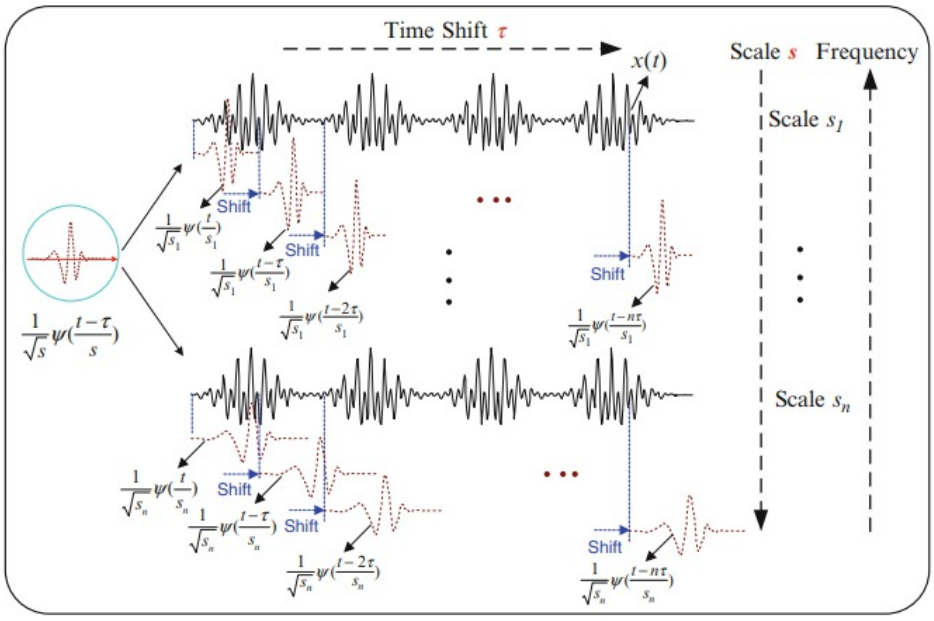


FIGURE 5. Schematic diagram of the processes found in the wavelet transform [16]

4.3.3. *Performance analysis of the different wavelet methods using percentage deviation (PD).* To compare the performance of various types of mother wavelet in analyzing the traffic signal of data for detecting anomalies the percentage deviation evaluation metric is used. In order to contrast and compare the analyses' features, the percentage deviation of the wavelet coefficients obtained after the wavelet transform is computed for each analysis. The median is calculated for the sample points in a particular coefficient. Then, the percentage deviation (PD) for a particular value x is computed using Equation (5):

$$PDx = (x - median) * 100 \tag{5}$$

The reason behind the use of percentage deviation is that those wavelet coefficients that are obtained after the wavelet transform of the traffic signal displaying a lower percentage deviation value are better because deviation magnitude from the origin indicates

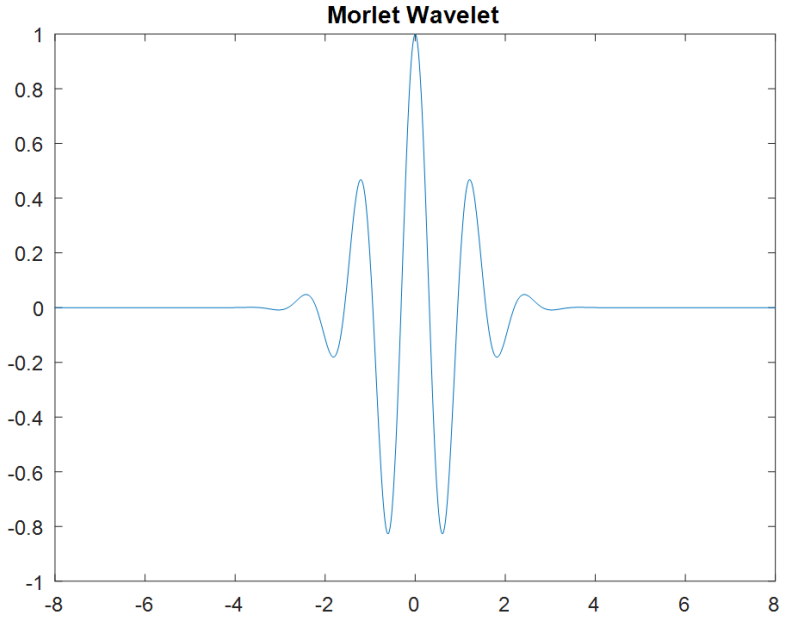


FIGURE 6. Morlet wavelet (simulated in Matlab)

an anomaly in the signal data. Specifically, a good wavelet must demonstrate a larger deviation amount at the start and end positions of an anomaly whereas it should demonstrate lower deviations at other regions in the data signal, in order to create a region of contrast so that this particular contrast is bigger making the anomaly more identifiable. As each trace consists of few anomalies, a good wavelet candidate for the data analysis would have the least deviation in comparison to other wavelet candidates [8].

4.4. **Algorithm steps for detecting of anomalous activities.**

Step 1: Extract relevant data from the CCIDS dataset for week 2 (bytes data) where visible anomalous peaks were found.

Step 2: Perform continuous wavelet transform 1-D using the 5 types of mother wavelets on the one-dimensional traffic signal data.

Step 3: Obtain the continuous wavelet coefficients and save coefficients.

Step 4: Compute the evaluation metric using percentage deviation applied on wavelet coefficients (for each level wherever applicable $2^7 = 128$ scales and 7 is the number of levels).

The average of the percentage deviation (PD) for each applied mother wavelet based on the number of levels is computed and compared for performance purposes of detecting changes in the network traffic signal.

5. **Results.** The percentage deviation was used as an analysis metric to compare the wavelet coefficients that are obtained after the wavelet transform has been applied to the data signal. In this research, the experimental setup includes an external server, which gives two online services that are synchronization service and a public webserver. The python scripts, within the OpenStack environment, simulated different types of realistic user network activities that are similar to a work environment. Then different types of attack were generated within the virtual network to create malicious traffic, and it was also exposed to real and up-to-date attacks from the web. This allows the recording of both normal and also malicious traffic data at the external server. The different types of wavelets that were used in this research are depicted in the following Figure 7. Each wavelet has a different transition change. For instance, Haar wavelet has an abrupt

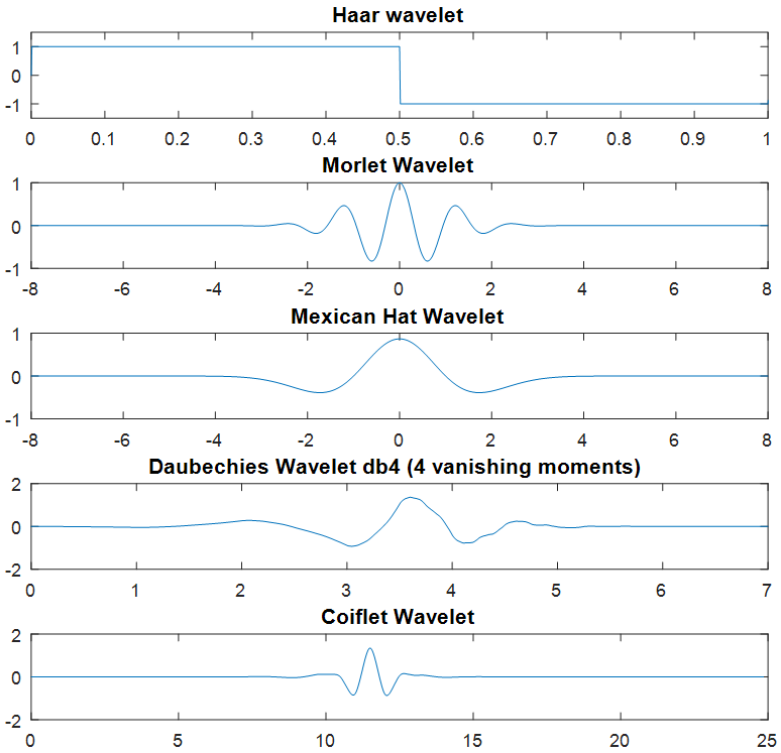


FIGURE 7. The different types of mother wavelet to analyze the traffic signal data (simulated in Matlab)

TABLE 2. The characteristics of the computed continuous wavelet transform

Name of wavelet	Size of CWT coefficients	Mean of CWT coefficients	Median of CWT coefficients
Haar	$*7 \times 159373$	-1.6012	0
Morlet	$*7 \times 159373$	0.2828	0
Mexican Hat	$*7 \times 159373$	0.5984	-5.2719e-10
Daubechies	$*7 \times 159373$	-1.6012	0
Coiflet	$*7 \times 159373$	0.5767	0.0139

transition at the point $x = 0.5$ while for example Mexican Hat has a smooth transition peak at $x = 0$. These transitions will be exploited and used to detect abrupt changes in the network activity.

5.1. **Characteristics of the CWT coefficients.** In Table 2, the features of the computed continuous wavelet transform that were obtained after its application to the network traffic signal using various types of wavelet are illustrated. The first column of Table 2 describes the types of mother wavelet used, the second column describes the size of the computed continuous wavelet transform coefficients whereby 7×159373 represents the number of levels and the length of network signal data respectively. The third column describes the computed average of the continuous wavelet coefficients followed by the fourth column, which summarizes the median values of the continuous wavelet coefficients.

5.2. **Percentage deviations.** As shown in Table 3, the average percentage deviations of the continuous wavelet coefficients were computed based on the median value obtained in Table 2. The first column of Table 3 describes the 5 types of mother wavelet used in this research followed by column 2 which shows the computed mean sum percentage

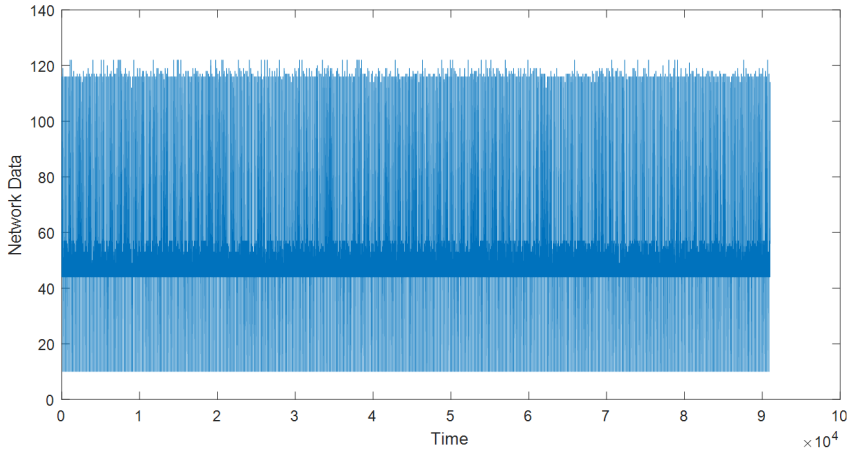
TABLE 3. The computed average percentage deviations of the CWT coefficients from the median value at the various levels

Wavelet name	**Mean sum percentage deviations from the median for each respective wavelet for the whole length of data at each level (%)**	**Relative performance to Haar wavelet**
Haar	160.1	1
Morlet	**28.3**	**5.65**
Mexican Hat	59.8	2.67
Daubechies	160.1	1
Coiflet	56.3	2.84

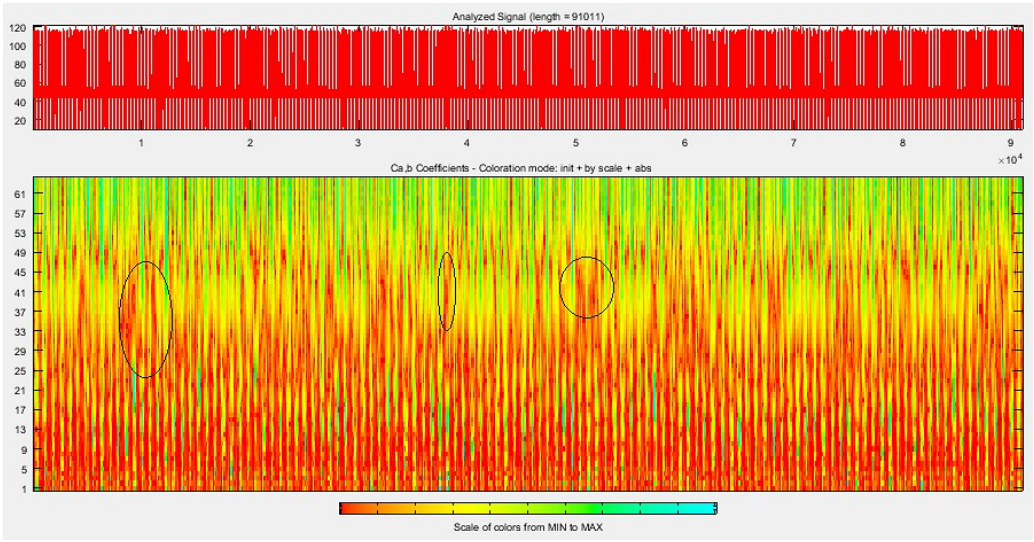
deviations from the median value for each type of wavelet. The third column was added to compare the performance of each type of wavelet against one of the poorly performed mother wavelets such as Haar wavelet (or Daubechies wavelet) as both produce very high percentage deviation values.

5.3. **Validity of findings.** In order to validate the findings in Section 5.2, the continuous wavelet transform was applied to a real network dataset using Morlet wavelet. The DARPA99 dataset was used as benchmark data to observe if Morlet wavelet can detect abrupt changes in the data of length (91) and also to see if those abrupt changes correspond to network intrusion as shown in the following noisy network data. It was found that 1-D continuous wavelet transform using Morlet wavelet could find minute changes in the signal through detailed illustrative figures as shown in Figures 8(a)-8(c).

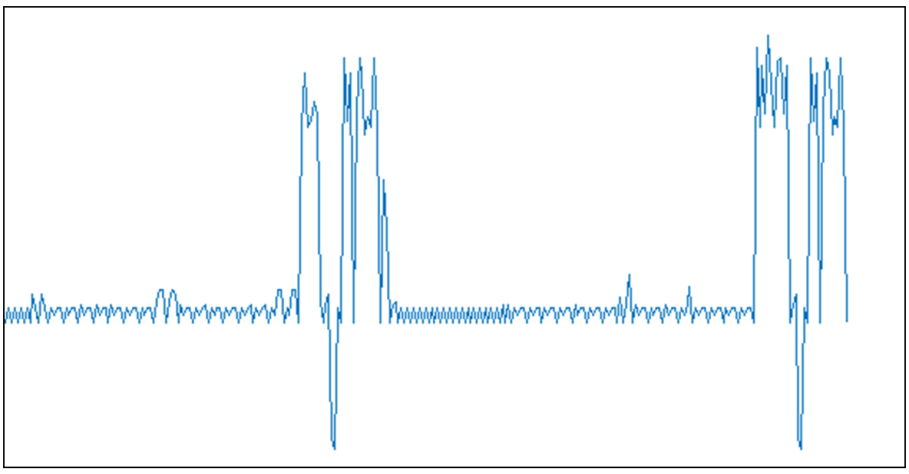
6. **Discussions.** This research employed an existing dataset that was generated from previous studies [14]. Advanced signal processing technique using *continuous wavelet transform* 1-D in the field of network traffic was implemented which considered all detailed changes in the network activity and at various scales or frequencies of the one-dimensional signal. Previous network anomaly detection research in [17] combined both wavelet approximation and system identification theories to achieve high-detection rates in terms of attacks and types of attack. The wavelet analysis analyzes both frequency and time scales and this is why it is robust to detect any change in time as well as frequency respectively. It was found that the Morlet wavelet outperforms the other methods considerably in detecting anomalies in the network data (with a performance index of 5.65) based on the lowest average percentage deviation value of the continuous wavelet coefficients which was found to be 28.3% for the Morlet wavelet as compared to Coiflet wavelet 56.3%, Mexican Hat wavelet with 59.8% and Haar and Daubechies (db1) wavelets with 160.1% each. In some studies, Coiflet wavelet was found to be a good filter for detecting seismic activities [18]. Also, according to the study in [8], it showed that Mexican Hat wavelet performed better than Coiflet wavelet, Daubechies wavelet and Morlet wavelet using wavelet packet analysis whereas in our study Mexican Hat is a second choice while using the 1-D continuous wavelet transform. Furthermore, they pointed out the idea in [4], by emphasizing on the point that a common technique in making the right decision to choose a particular wavelet for a particular signal is not trivial as network traffic data are non-stationary. Thus, there is no one single wavelet to accurately match signal behavior especially in the case of experiencing different types of anomalous traffic activity. Here, this research study demonstrated that the best mother wavelet in detecting anomalies



(a)



(b)



(c)

FIGURE 8. (color online) (a) Plot of the network data (which suffers from attack such as Neptune or mailbomb attacks) vs. time; (b) wavelet analysis of the large dataset to show regions of attacks and (c) zoomed version of the left circle drawn in (b)

in real-time data is the Morlet wavelet and that this technique can be implemented for anomaly detection.

7. **Conclusion.** In the field of network intrusion, in order to detect anomalies in long duration network traffic data, Morlet wavelet continuous wavelet transform seems a very promising continuous wavelet candidate. This research set the duration limit to one week, approximately 160,000 data points, and the Morlet wavelet demonstrated its superior performance as compared to other mother wavelet families such as Mexican Hat, Coiflet, Daubechies, and Haar. Therefore, a network intrusion system can employ a continuous wavelet transform with Morlet as the mother wavelet to analyze real-time data which will give unbiased results in terms of signal anomaly detection. Furthermore, such a novel network intrusion system using continuous wavelet transform of the traffic signal can prevent any sudden attacks from network intruders.

REFERENCES

[1] J. P. Anderson, *Computer Security Threat Monitoring and Surveillance*, Technical Report, James P. Anderson Company, Fort Washington, 1980.
[2] D. E. Denning, An intrusion detection model, *IEEE Trans. Software Engineering*, vol.13, no.2, pp.222-232, 1987.
[3] A. M. Lopes and J. A. T. Machado, Tidal analysis using time-frequency signal processing and information clustering, *Entropy*, vol.19, no.8, 2017.
[4] P. Barford, J. Kline, D. Plonka and A. Ron, A signal analysis of network traffic anomalies, *Proc. of ACM SIGCOMM Internet Measurement Workshop*, 2002.
[5] B.-S. Liu, Y.-J. Li, Y.-P. Hou and X.-S. Sui, The identification and correction of outlier based on wavelet transform of traffic flow, *International Conference on Wavelet Analysis and Pattern Recognition*, vol.4, pp.1498-1503, 2007.
[6] S. Rawat and C. S. Sastry, Network intrusion detection using wavelet analysis, *Proc. of CIT*, Hyderabad, India, pp.224-232, 2004.
[7] A. Soule, K. Salamatian and N. Taft, Combining filtering and statistical methods for anomaly detection, *Proc. of IMC*, 2005.
[8] P. Huang, A. Feldmann and W. Willinger, A non-intrusive, wavelet based approach to detecting network performance problems, *Proc. of Internet Measurement Workshop*, 2001.
[9] M. S. Kim, T. Kim, Y. S. Hin, S. S. Lam and E. J. Powers, A wavelet-based approach to detect shared congestion, *Proc. of the ACM SIGCOMM'04*, 2004.
[10] C.-T. Huang, S. Thareja and Y.-J. Shin, Wavelet-based real time detection of network traffic anomalies, *International Journal of Network Security*, vol.6, pp.309-320, 2008.
[11] C. Callegari, S. Giordano, M. Pagano and T. Pepe, Combining sketches and wavelet analysis for multi time-scale network anomaly detection, *Comput. Secur.*, vol.30, no.8, pp.692-704, 2011.
[12] S. Novakov, C.-H. Lung, I. Lambadaris and N. Seddigh, A hybrid technique using PCA and wavelets in network traffic anomaly detection, *International Journal of Mobile Computing and Multimedia Communications*, vol.6, no.1, pp.17-53, 2014.
[13] M. Salagean, Real network traffic anomaly detection based on analytical discrete wavelet transforms, *Proc. of OPTIM'10*, pp.926-931, 2010.
[14] M. Ring, S. Wunderlich, D. Grudl, D. Landes and A. Hotho, Flow-based benchmark data sets for intrusion detection, *Proc. of the 16th European Conference on Cyber Warfare and Security (ECCWS'17)*, pp.361-369, 2017.
[15] M. Cesari, J. Mehlsen, A.-B. Mehlsen and H. B. D. Sorensen, A new wavelet-based ECG delineator for the evaluation of the ventricular innervation, *IEEE J. Transl. Eng. Health Med.*, vol.5, 2017.
[16] R. X. Gao and R. Yan, From Fourier transform to wavelet transform: A historical perspective, in *Wavelets: Theory and Applications for Manufacturing*, Berlin, Germany, 2011.
[17] W. Lu and A. A. Ghorbani, Network anomaly detection based on wavelet analysis, *EURASIP J. Advances in Signal Processing*, vol.2009, 2009.
[18] V. Joevivek, N. Chandrasekar and R. Jayangondaperumal, Evaluation of optimal wavelet filters for seismic wave analysis, *Himalayan Geology*, vol.37, no.2, pp.176-189, 2016.