Chapter 8

Wavelet Transform for Educational Network Data Traffic Analysis

Shwan Dyllon and Perry Xiao

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76455

Abstract

Network monitoring and analysis are very important, in order to understand the performance of the networks, the reliability of the networks, the security of the networks, and to identify potential problems. In this chapter, we present our latest work on university network data traffic analysis by using continuous wavelet transform (CWT). With CWT, you can analyse the data and show how the frequency content of the data changes over time. This time dependent frequency varying information, which is lacking in other techniques, such FFT, is very useful for network traffic analysis. A twelve month's network traffic data, include World Wide Web (WWW) data and Email data were presented in 3D format, by using wavelet transform, we can visualise the hourly, daily, weekly and monthly activities. We will first present the theoretical background, then show the experimental results.

Keywords: wavelet transform, educational network, data traffic analysis

1. Introduction

Network monitoring and analysis are very important, in order to understand the performance of the networks, the reliability of the networks, the security of the networks, and to identify potential problems. With network traffic analysis, network security staff would be able to identify any malicious or suspicious packets within the traffic, whilst network administrators could monitor the download/upload speeds, throughput, etc., and therefore to have a better understanding of network operations. To date, many techniques have been used in network data traffic analysis. The neural network (NN), also known as the artificial neural network (ANN), has been used for prediction, as well as to identify the presence of anomalies

[1–3]. Pattern recognition has been used for traffic data classification [4, 5], and chaos theory has been used for the correlation and the prediction of time series data, and to identify the nonlinear dynamical behaviour of real-time traffic data [6, 7]. The Fourier transform (FFT) and wavelet transform have been used to analyse the frequency components of the traffic data [8–10]. The main difference between FFT and wavelet transform is that wavelet transform is localised in both time and frequency whereas the standard Fourier transform is only localised in frequency. In other words, with wavelet transform, when a certain frequency event happened, we can know both what frequency component was, and when it happened. In this chapter, we will use wavelet transform to analyse the London South Bank University network traffic data, in order to understand and evaluate the network utilisation.

London South Bank University (LSBU) network is based on three-tier network architecture, i.e. CORE layer, distribution layer, and edge layer. The core layer is responsible for the routing protocols, the distribution layer responsible for all the VLAN management as well as spanning tree protocol and loop prevention as well as some level of security i.e. DOS protection, and finally the edge layer responsible for the end user connectivity. The LSBU network traffic raw data were first captured using the Paessler Router Traffic Grapher (PRTG) network-monitoring tool (Paessler AG, Germany), and then many numerical analysis algorithms, including wavelet transform, were developed to analyse the captured raw data.

2. Wavelet data analysis

A wavelet is a small wave. Wavelet data analysis is based on the wavelet transform, which has been used for numerous studies in geophysics, including tropical convection [11]. The wavelet transform can be used to analyse time series that contain nonstationary power at many different frequencies [12]. Unlike traditional sin(t) or cos(t) waves that go from negative infinity to positive infinity, wavelets always begin at zero, increases, and then decreases back to zero. Many types of wavelets exist, most of which are used for orthogonal wavelet analysis [13, 14], which purposefully crafted to have specific properties that make them useful for signal processing. **Figure 1** shows some examples of commonly used wavelets.

Wavelet transform is the convolution of time sequence data and wavelets, and can be generally expressed as:

$$F(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\, \psi^*\left(\frac{t-b}{a}\right) dt \tag{1}$$

Here, a is the scale ($a > 0$), b is the translational value, t is the time, f(t) is the data, and $\psi(t)$ is the wavelet function, and the * is the complex conjugate symbol. Wavelet transform can be generally divided into discrete wavelet transform (DWT) and continuous wavelet transform (CWT).

The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales. DWT decomposes the signal into mutually orthogonal set of wavelets, which is the main difference from the continuous wavelet transform (CWT). DWT can be used for wavelet decomposition and easy and fast denoising of a noisy signal.

http://dx.doi.org/10.5772/intechopen.76455

Continuous wavelet transform (CWT) is an implementation of the wavelet transform using arbitrary scales and almost arbitrary wavelets. The wavelets used are not orthogonal and the data obtained by this transform are highly correlated. To approximate the continuous wavelet transform, the convolution should be done N times for each scale, where N is the number of points in the time series [15]. The wavelet transform is similar to the Fourier transform, but unlike Fourier transform (which is localised only in the frequency space), the wavelet transform localised in both the time space and the frequency space, see **Figure 2**. CWT allows users to have variable resolutions, i.e. either high precision in time and low precision in frequency, or high precision in frequency and low precision in time. Although windowed transform,

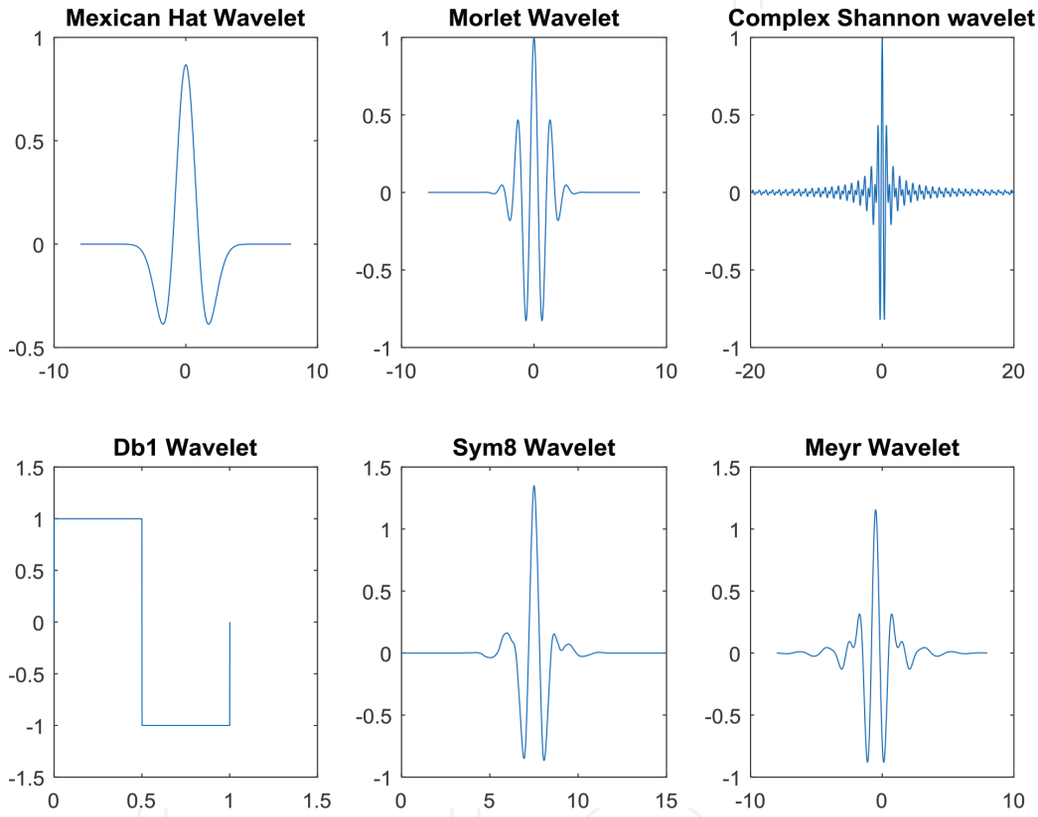

Figure 1. Examples of different types of wavelets.

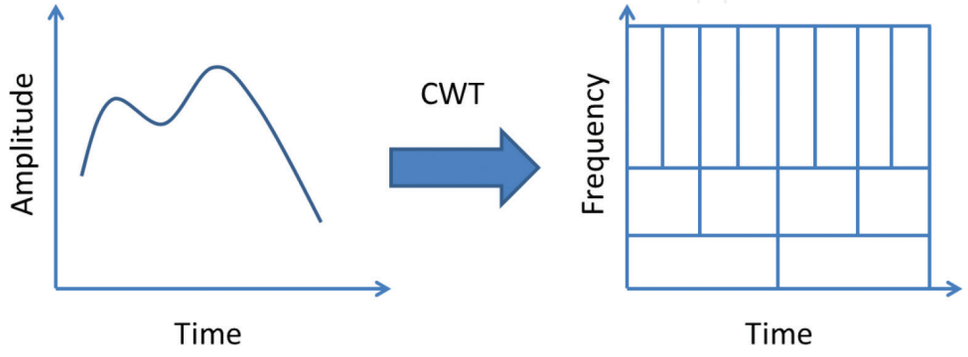

Figure 2. Continuous wavelet transform.

such as Short-Time Fourier Transform (STFT), which is also cable to create a local frequency analysis, the drawback of STFT is that the window size is fixed.

In this study, the wavelet transform used for data decomposition, data denoising and time dependent frequency components analysis by using continuous wavelet transform (CWT).

3. Results and discussions

3.1. Network traffic data 2D and 3D presentation

Figure 3 shows the LSBU 1-year total network traffic data, recorded at 1 h interval, in 3D format (top) and 2D format (bottom), which X axis represents the time of the day, from 01:00 to 24:00, and Y axis represents the day of the year, from 1 to 365, and Z axis represents the total traffic in Gbits per second. The data was recorded at 1 h interval for a period of 1 year, November 2016 to November 2017. By presenting the network traffic data in 2D and 3D formation we can better understand the network usage and characterisations.

The results show that the total network traffic varies from season to season throughout the year, and also varies from time to time throughout the day. By understanding the total network traffic pattern, we can plan better for the network operations, optimise the network usage, and identify potentially suspicious traffics.

Figure 4 shows the LSBU 1-year World Wide Web (WWW) traffic data in 3D format (top) and the corresponding 2D presentation (bottom). The results show that WWW traffic is highly seasonal. It has a strong week day and weekend effect, this agrees well with our previous studies [16, 17]. It also has a strong effect of Christmas, Easter and summer holiday periods. The WWW traffic data varies significantly within a day, with the highest between 10:00 am and 19:00 pm, and lowest between 06:00 am and 09:00 am, not at the midnight! Also, there seems more traffic during the autumn semester (September–January) than spring semester (February–June).

Figure 5 shows the 1-year Email traffic data in 3D format (top) and the corresponding 2D format (bottom). Similar to the WWW data, the Email data also shows week day and weekend effect, as well as seasonal effect. However, different from the WWW data, the major of the Email traffic was between 09:00 am and 18:00 pm, there is very little traffic in the evening and early in the morning. So in these periods, people browsed the web but did not send many emails. The massive peak at the middle of the graph is due to the Email upgrade, where a lot of emails have been sent and received.

3.2. Network traffic data and Fourier transform

Figure 6 shows the original 1-year LSBU total network traffic data (top) and the corresponding Fourier transforms (bottom). **Figure 7** shows the original 1-year LSBU WWW data (top) and the corresponding Fourier transforms (bottom). **Figure 8** shows the original 1-year LSBU Email data (top) and the corresponding Fourier transforms (bottom).

http://dx.doi.org/10.5772/intechopen.76455

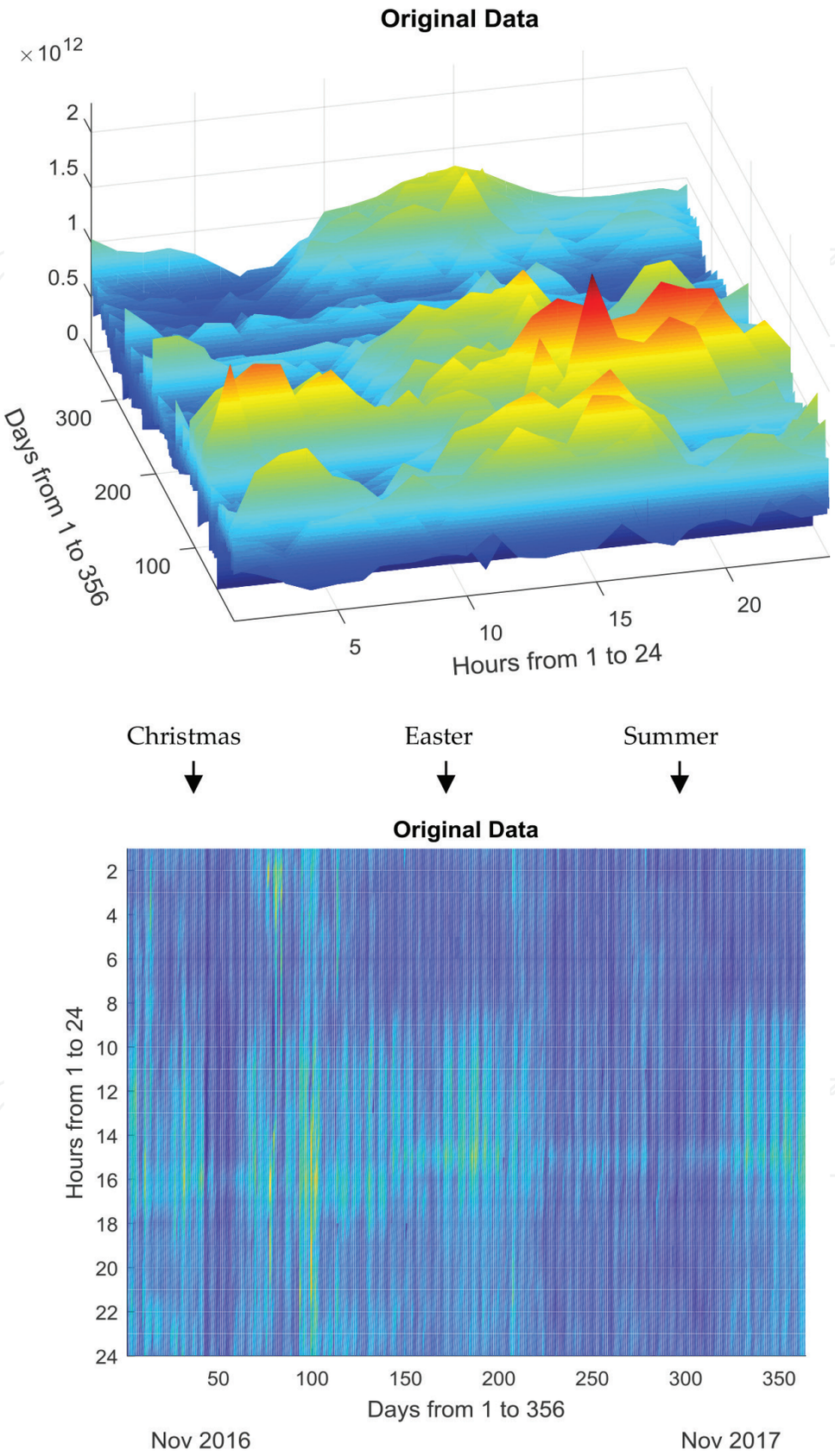

Figure 3. The 3D presentation (top) and the corresponding 2D presentation (bottom) of 1-year total network data in a daily usage pattern (Nov 2016–Nov 2017).

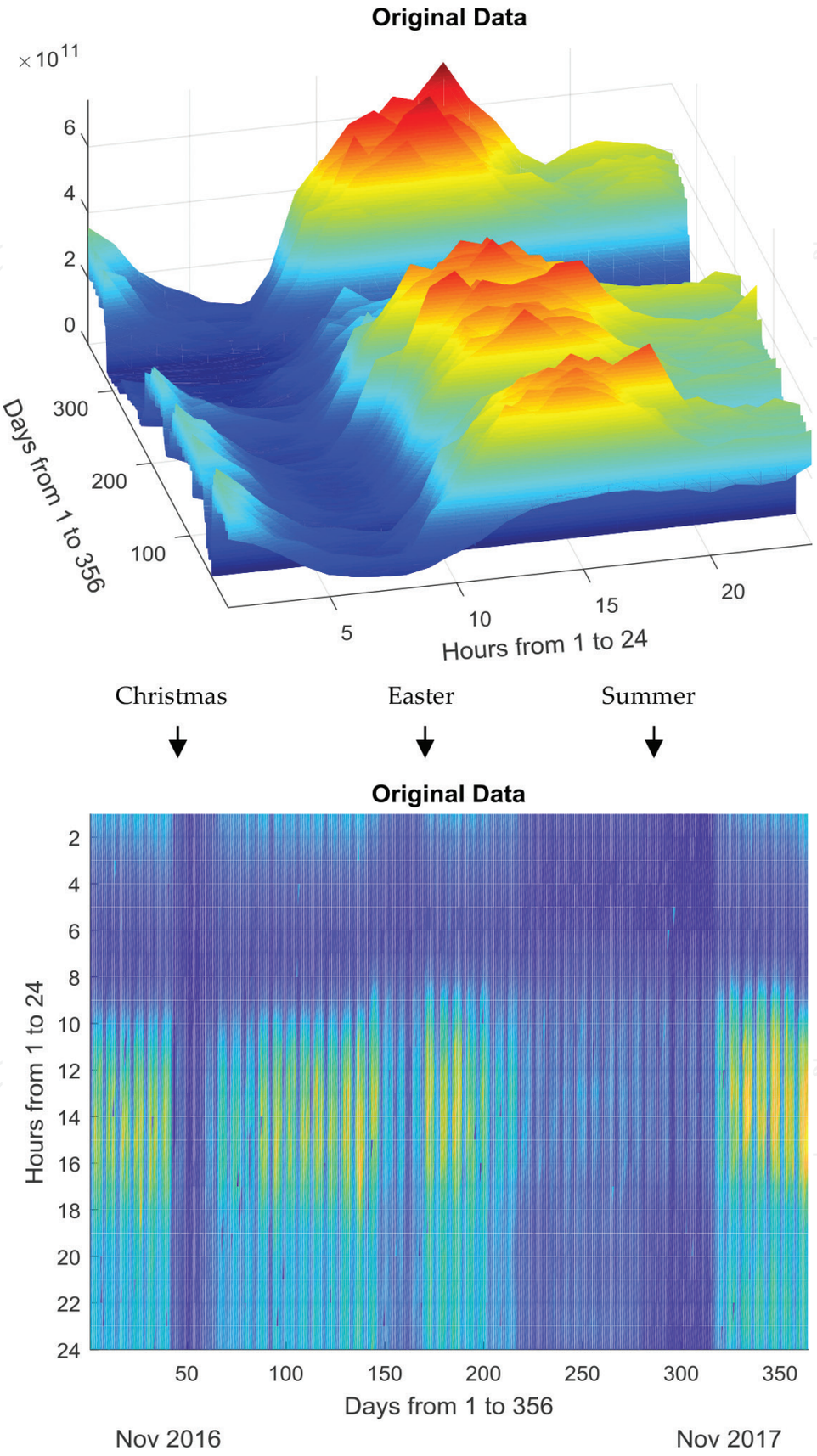

Figure 4. The 3D presentation (top) and the corresponding 2D presentation (bottom) of 1-year WWW traffic data in a daily usage pattern (Nov 2016–Nov 2017).

http://dx.doi.org/10.5772/intechopen.76455

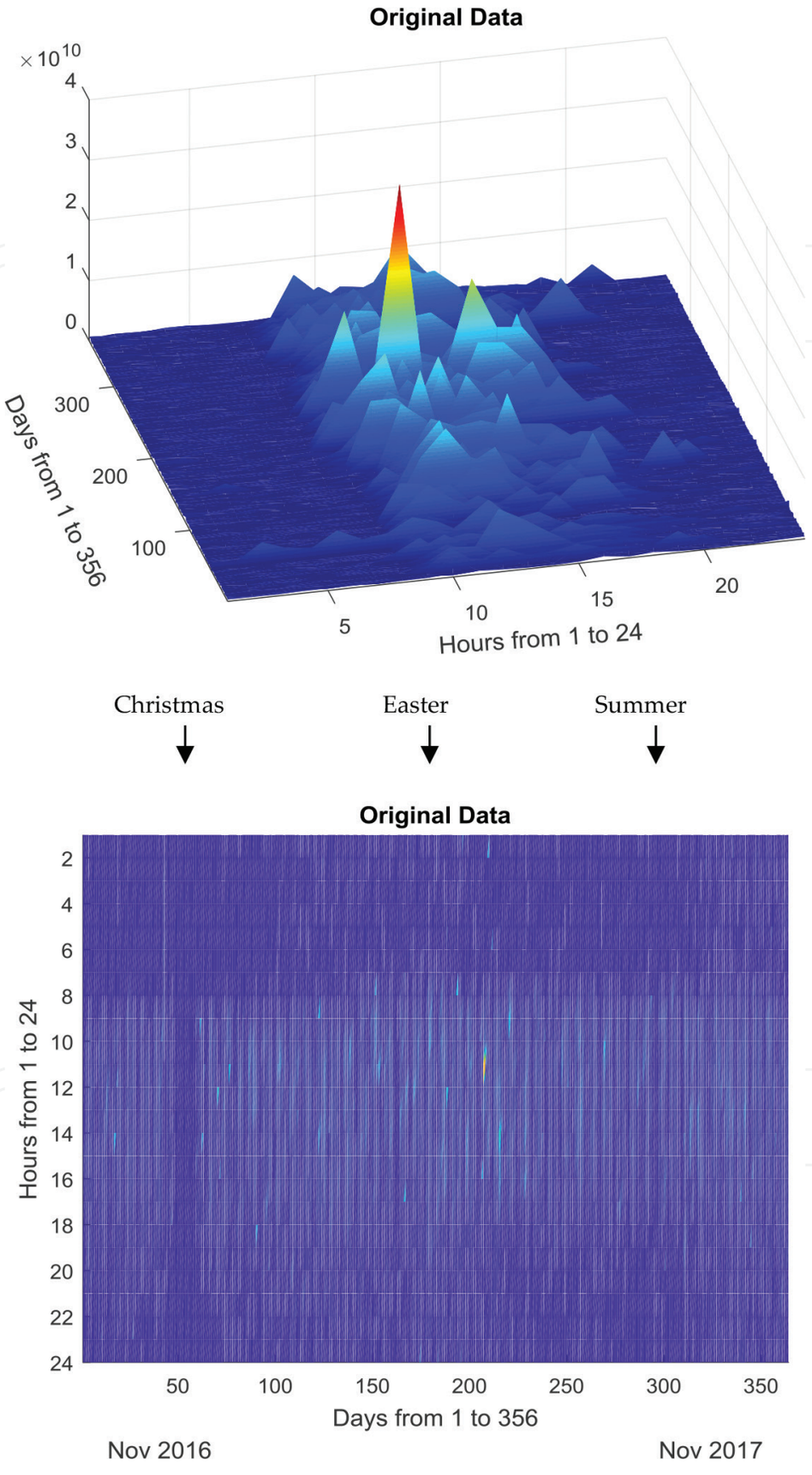

Figure 5. The 3D presentation (top) and the corresponding 2D presentation (bottom) of 1-year Email data in a daily usage pattern (Nov 2016–Nov 2017).

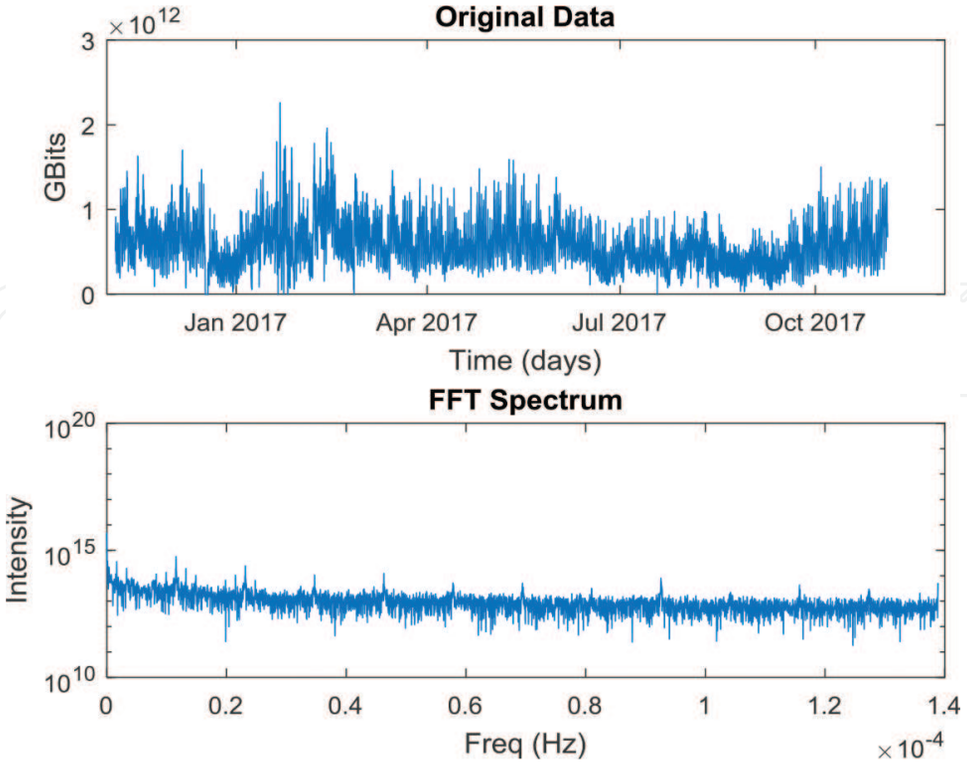

Figure 6. The original 1-year total network traffic data (top) and its corresponding FFT spectrum (bottom) (Nov 2016–Nov 2017).

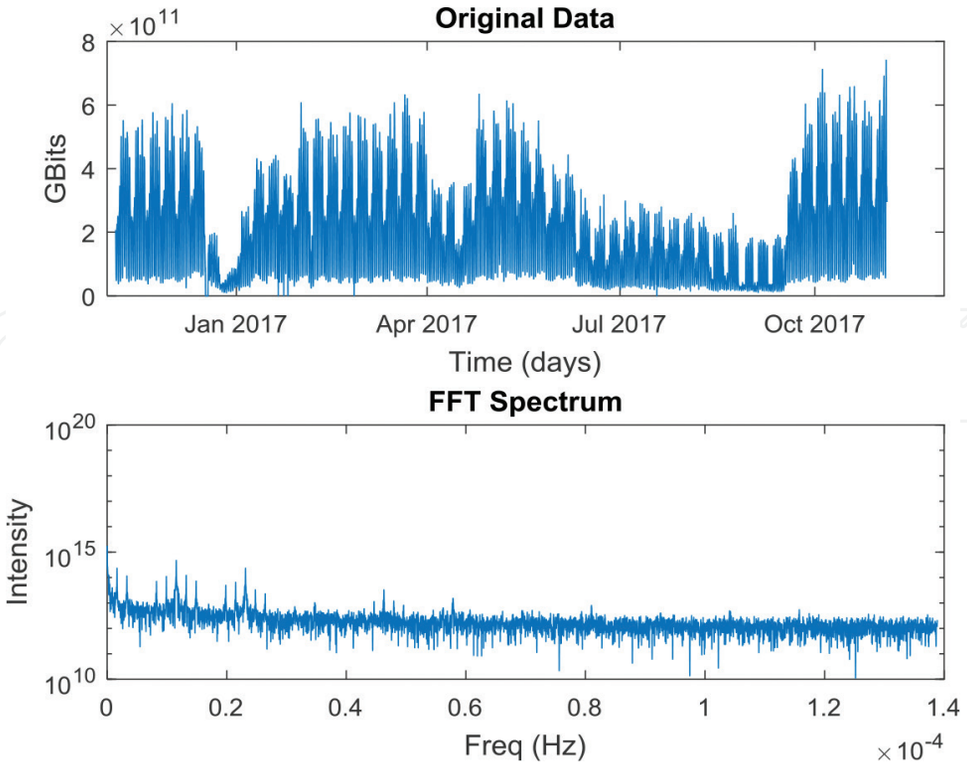

Figure 7. The original 1-year WWW data (top) and its corresponding FFT spectrum (bottom) (Nov 2016–Nov 2017).

The total network traffic data, the WWW data, and the Email data, have completely different patterns, and therefore different FFT spectrum. With the total network traffic data, there are a lot of small peaks throughout the FFT spectrum, indicating there are repeatedly happened events. But with the WWW data and Email, the FFT peaks mainly occurs at lower frequency range. But with FFT, it is not possible to identify when these events happened.

3.3. Wavelet decomposition

Wavelet decomposition [18] is a powerful tool that can decompose the original network traffic into low frequency component (A) and high frequency component (D). The low frequency component (A) is also called approximation coefficient, and the high frequency component (D) is called detail coefficient. By performing decomposition several times, we also have multilevel wave decomposition, see **Figure 9**. The multilevel wavelet decomposition allows us to gradually separate and to eliminate high frequency components, which is mostly noise. Through wavelet decomposition we can reduce the data noise, and therefore observe the data trend better.

Figure 10 shows the wavelet decomposition of the WWW network traffic data, wavelet used is "sym4" wavelet (see **Figure 1**) and the level of decomposition is level 4. The key in wavelet decomposition is to choose the right wavelet and to select the right level of decomposition. The results show that the low frequency component (A4) reflects better the trend of the network traffic data, where the high frequency component (D4) reflects more about the traffic noise.

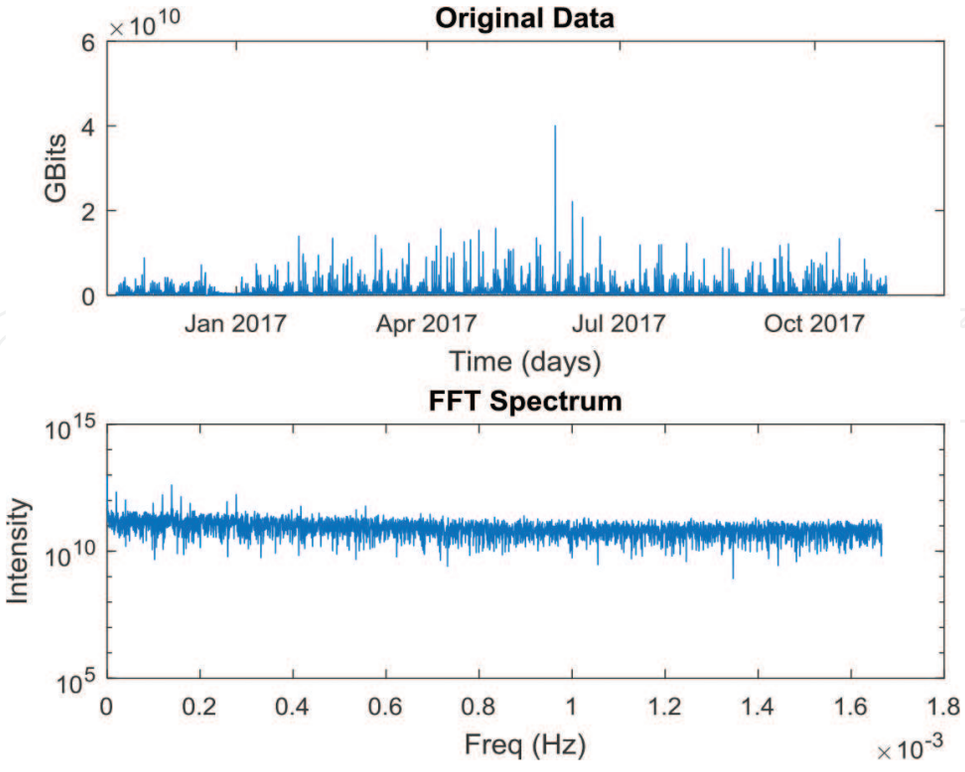

Figure 8. The original LSBU 1 year Email data (top, Nov 2016 – Nov 2017) and its corresponding FFT spectrum (bottom).

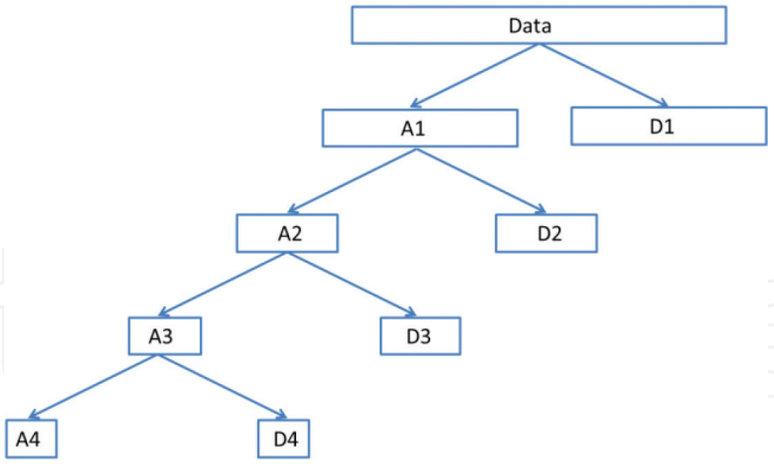

Figure 9. Multilevel wavelet decomposition, where approximation coefficient A is the low frequency component and detail coefficient D is the high frequency component.

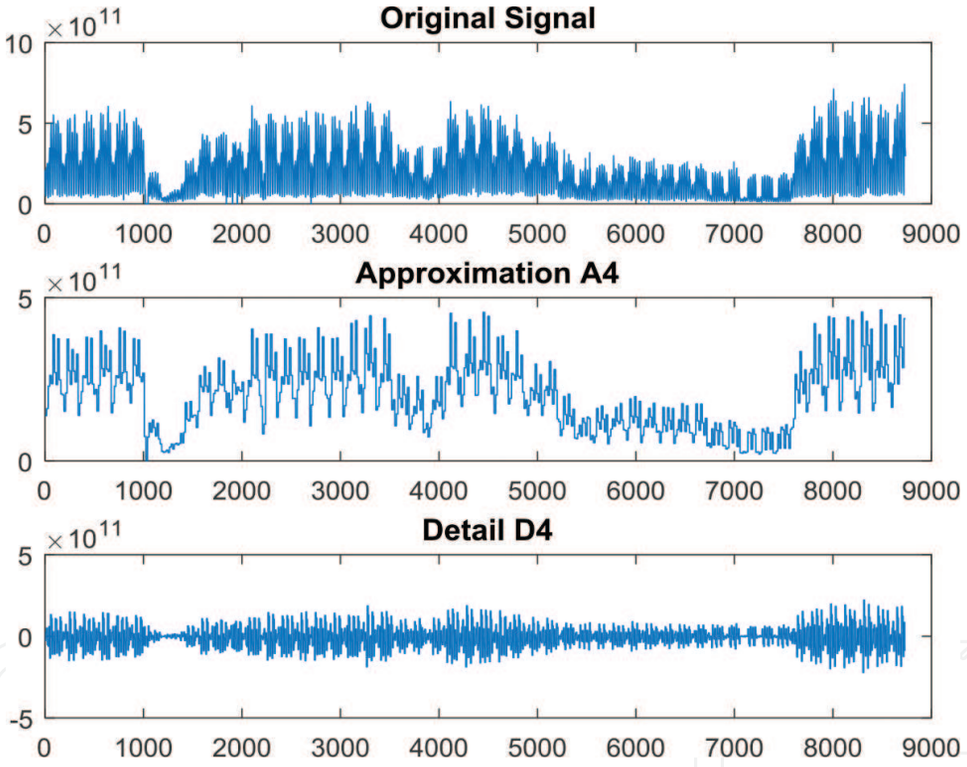

Figure 10. One-year WWW data in an hourly usage pattern (top, Nov 2016–Nov 2017), the corresponding level 4 low frequency component (A4) (middle) and level 4 high frequency component (D4) (bottom).

3.4. Wavelet denoising

Based on wavelet decomposition, a very useful feature of wavelet analysis is denoising, which is very useful for noisy data. The steps are as follows. First, choose a wavelet and a level of decomposition N, and then compute the wavelet decompositions of the data at levels 1 to N. For each level, a threshold is selected and the threshold applied to the detail coefficients (D). Finally, compute wavelet reconstructions using the original approximation coefficients (A) of level N and the modified detail coefficients (D) of levels 1 to N.

Figure 11 shows the 1-year total network traffic data in an hourly usage pattern (Nov 2016–Nov 2017) and the corresponding denoised results. In this case, the wavelet used is "sym8" wavelet (see **Figure 1**), the level of decomposition was chosen as N = 3. Similar to wavelet decomposition, the key in wavelet denoising is to choose the right wavelet and to select the right level of decomposition, in order balance the noise removal and signal integrity. The denoising of the WWW data and Email also yields similar results.

The quality of the denoised results is good. The trends of the original network traffic data are well preserved. To select the right wavelet and right level of decomposition is very important so that we can achieve maximum denoising and preserve the useful information.

3.5. Continuous wavelet analysis (CWT)

With continuous wavelet transform (CWT), we can analyse the data and show how the frequency content of the data changes over time. This time dependent frequency varying information, which is lacking in other techniques, such FFT, is very useful for network traffic analysis. In this CWT calculation, there are several parameters to choose from, i.e. the type of wavelet, the smallest scale (S_0), the space between scales (ds) and number of scales (N_s). The scales (S) can be converted to pseudo frequencies (f_p) by using the following formula,

$$f_p = \frac{f_c}{S\,dt} \tag{2}$$

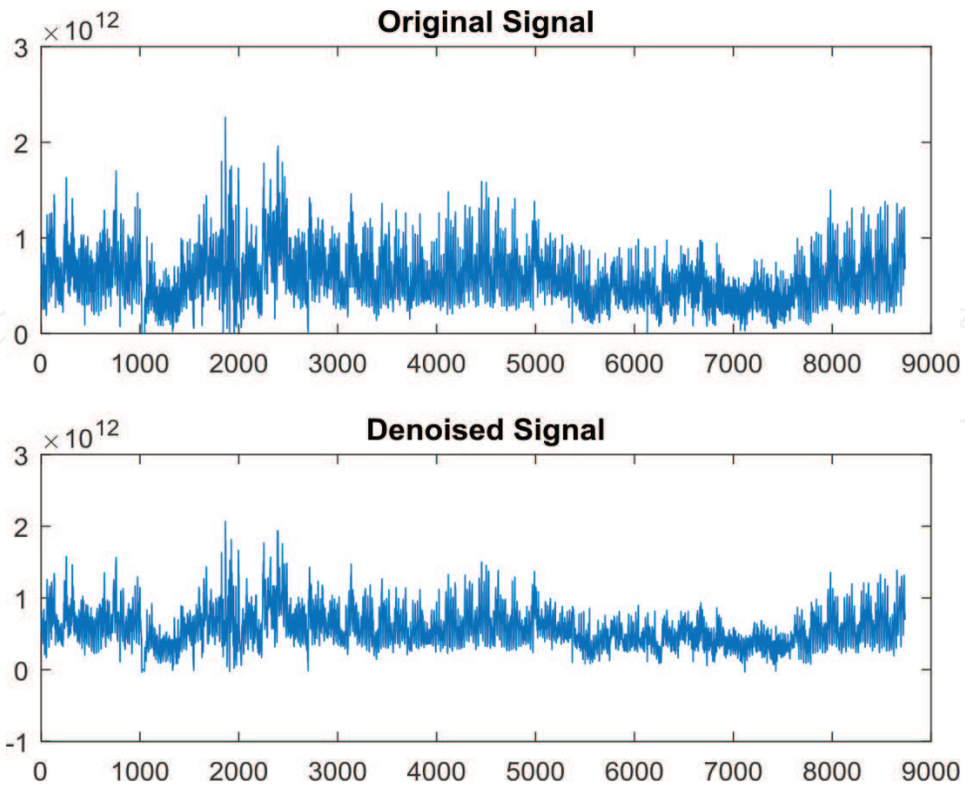

Figure 11. One-year total network traffic data in an hourly usage pattern (Nov 2016–Nov 2017) and the corresponding denoised results.

Where f_c is the centre frequency of the wavelet, and dt is the sampling time. Scales are inversely proportional to frequencies, i.e. small scales represents high frequencies, and vice versa.

Figure 12 shows the continuous wavelet transform (CWT) of 1 year total traffic using different wavelets, e.g. Morlet wavelet (analytic), Mexican hat wavelet (nonanalytic), bump wavelet (analytic), and Paul wavelet (analytic). The X axis is time of 1 year, and the Y axis is pseudo frequency. Different wavelet gives different results. Based on the results, we have decided to use Morlet wavelet to analyse the network traffic data, as it can provide more details on daily, weekly and monthly events, more details will be discussed later.

Figure 13 shows the continuous wavelet transform (CWT) using different parameters. The smallest scale (S_0) decides the highest frequency. The ds decides the resolution of the results, N_s decides the range of the frequency. By balancing the result resolution, frequency range and calculation time, we have decided to perform the CWT using the following values, $S_0 = 6 \times 3600 = 21{,}600$, i.e. six-hourly event, ds = 0.025, and $N_s = 300$.

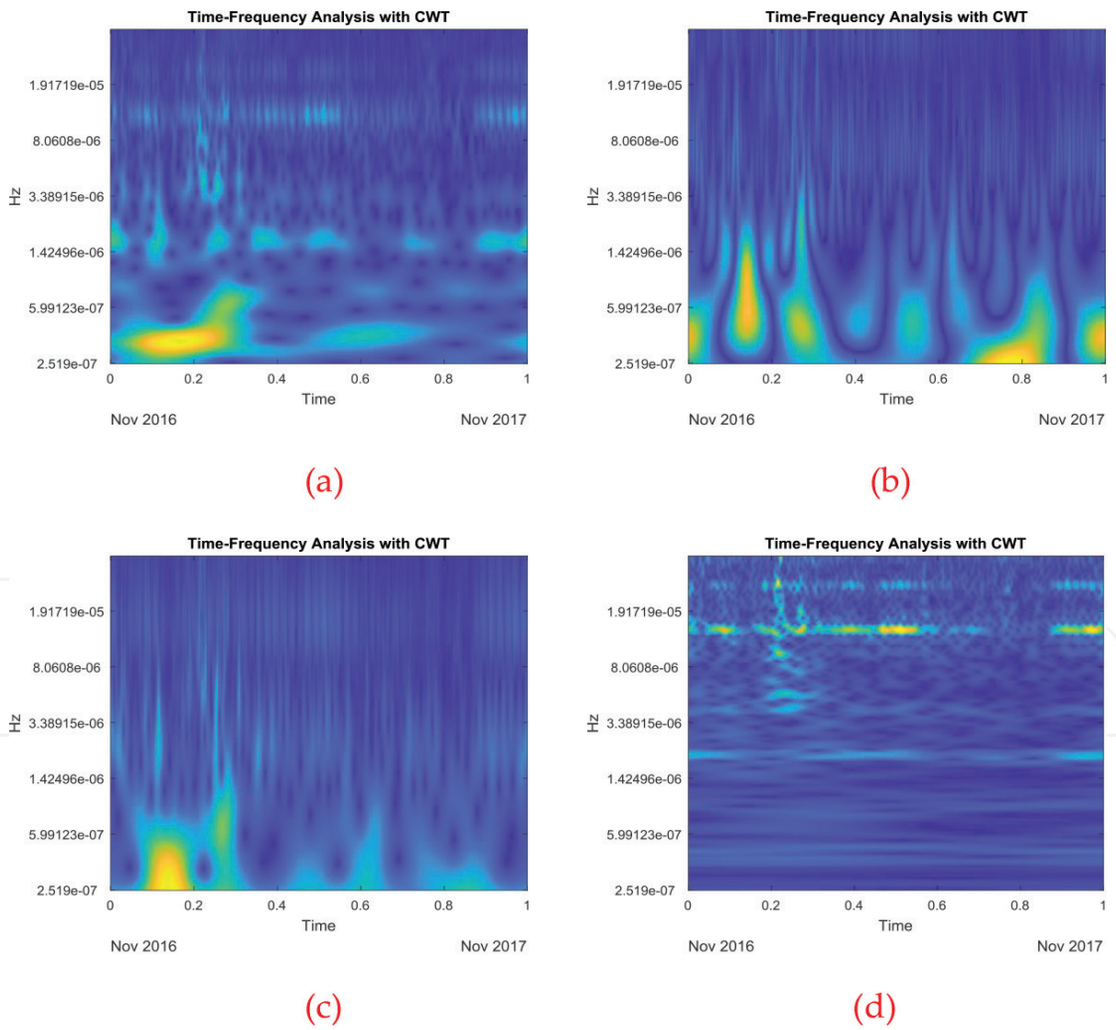

Figure 12. The continuous wavelet transform (CWT) using different wavelets. (a) CWT with Morlet wavelet, (b) CWT with Mexican hat wavelet, (c) CWT with Paul wavelet and (d) CWT with bump wavelet.

http://dx.doi.org/10.5772/intechopen.76455

Figure 14 shows the CWT results of the original 1 year total traffic data using Morlet wavelet, with S_0 = 21,600, ds = 0.025, and N_s = 300 as CWT parameters. The X axis is time of 1 year, and the Y axis is pseudo frequency. We can convert this pseudo frequency into event. **Table 1** shows the pseudo frequencies in Hz of hourly, daily, weekly, two weekly, monthly and quarterly events. Using these pseudo frequencies we can then identify the corresponding hourly, daily, weekly, two weekly, monthly and quarterly events in **Figure 11**. The hot spot at the lower left corner is the when the system is upgraded. By using CWT, we can easily identify the event which is otherwise difficult to identify in the original time domain.

Figure 15 shows the CWT results of the WWW traffic data using the same wavelet and same parameters. The results show that half daily and daily events happen throughout the year. They are highly seasonal, as you can clearly identify the summer, Christmas, and Easter gaps. The half daily and daily events also show clear day and night effects, as well as weekday and weekend effect, while weekly, two weekly and monthly events are patchy, with no seasonal effects.

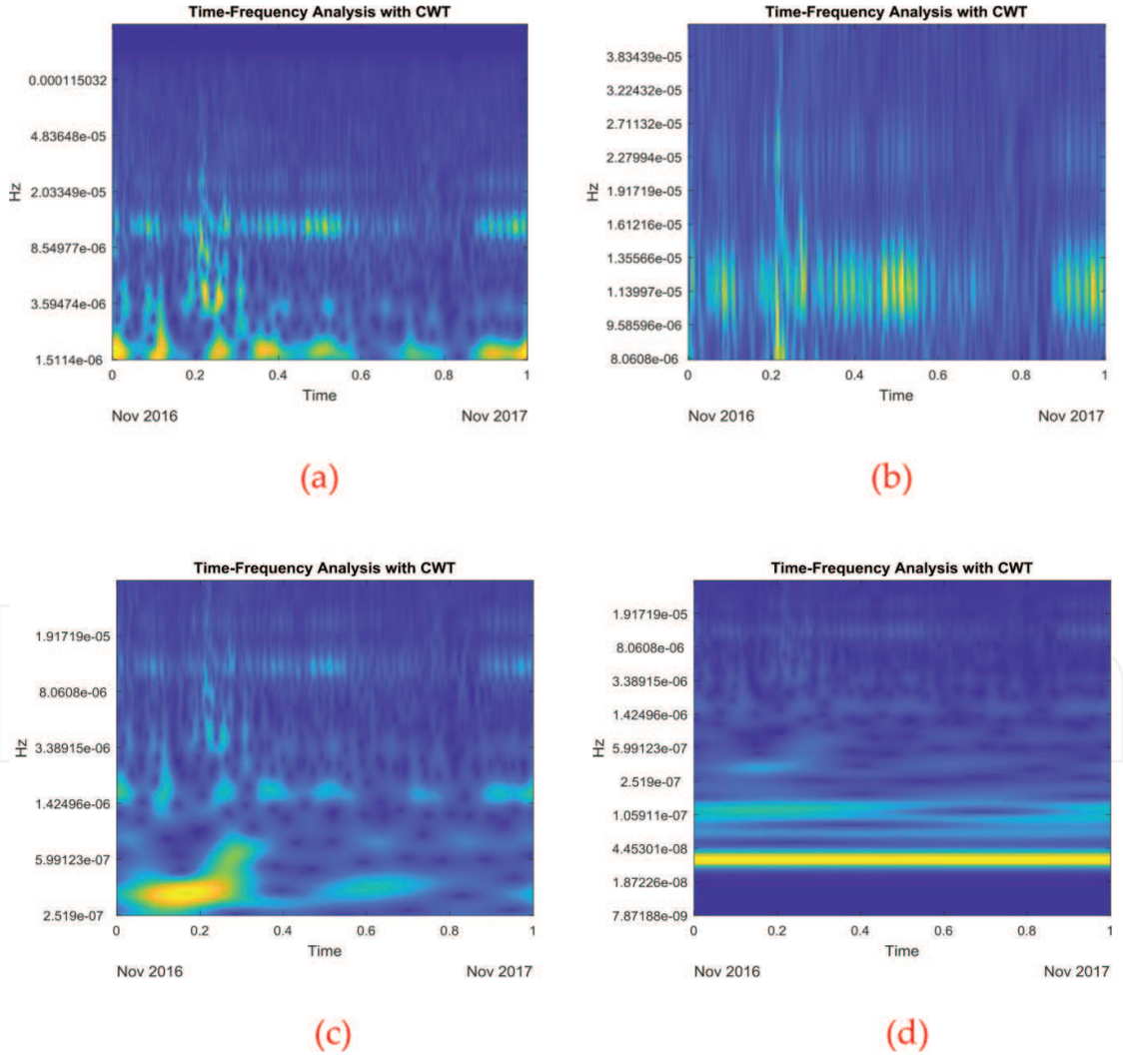

Figure 13. The continuous wavelet transform (CWT) using different parameters. (a) S_0 = 3600, ds = 0.025, N_s = 300, (b) S_0 = 21,600, ds = 0.025, N_s = 100, (c) S_0 = 21,600, ds = 0.025, N_s = 300 and (d) S_0 = 21,600, ds = 0.025, N_s = 500.

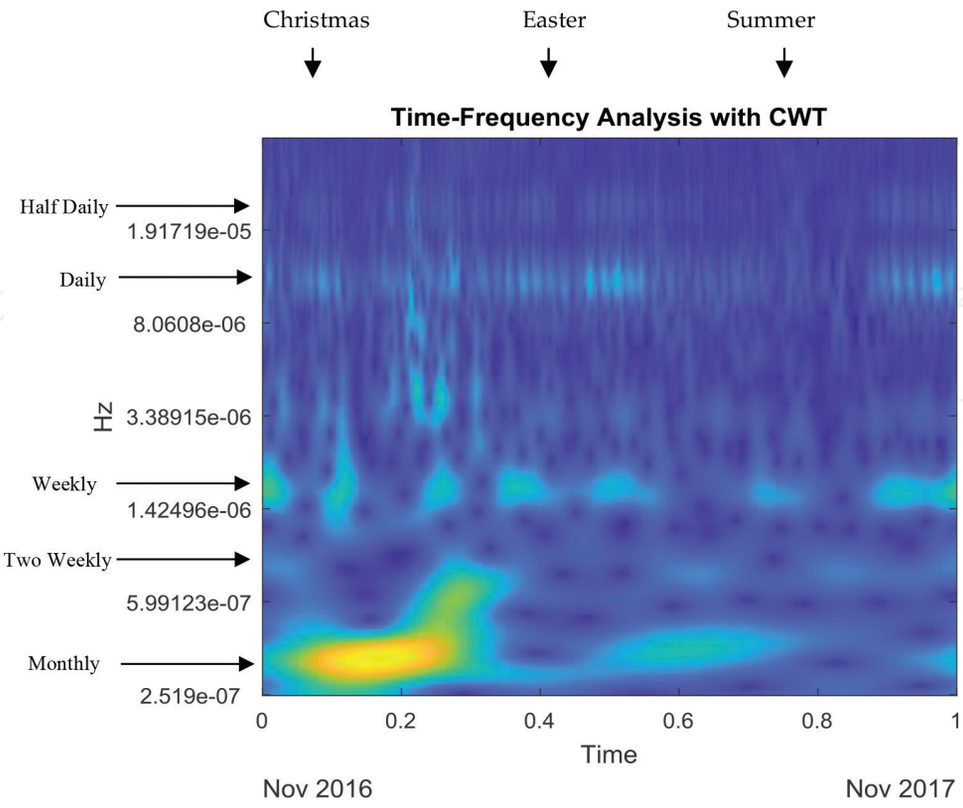

Figure 14. The CWT time-frequency 2D results of the 1-year total network traffic data (Nov 2016–Nov 2017). The hot spot at the lower left corner is the when the system is upgraded.

Time	Pseudo frequency (Hz)
Hourly	2.78E−04
Quarter daily	1.39E−04
Half daily	2.31E−05
Daily	1.16E−05
Weekly	1.65E−06
Two weekly	8.27E−07
Monthly	3.86E−07
Quarterly	1.29E−07

Table 1. The pseudo frequencies (Hz) of different events.

Figure 16 shows the CWT results of the Email traffic data using the same wavelet and same parameters. The results show that hourly event and quarter daily events happen throughout the year. The Christmas gap is obvious whilst the summer and the Easter gaps are not. They also show clear weekday and weekend effects. The half daily, daily and weekly events are very patchy, with no seasonal effects. This kind of time-frequency results can help us to understand the traffic characteristics better.

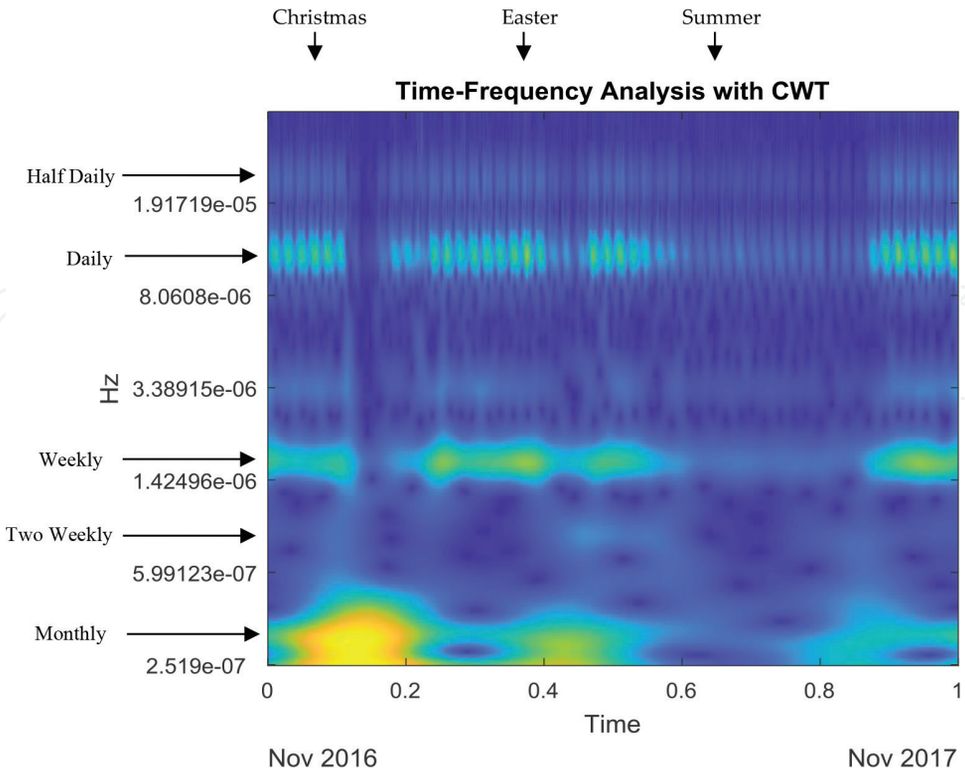

Figure 15. The CWT time-frequency 2D results of the 1-year WWW data (Nov 2016–Nov 2017).

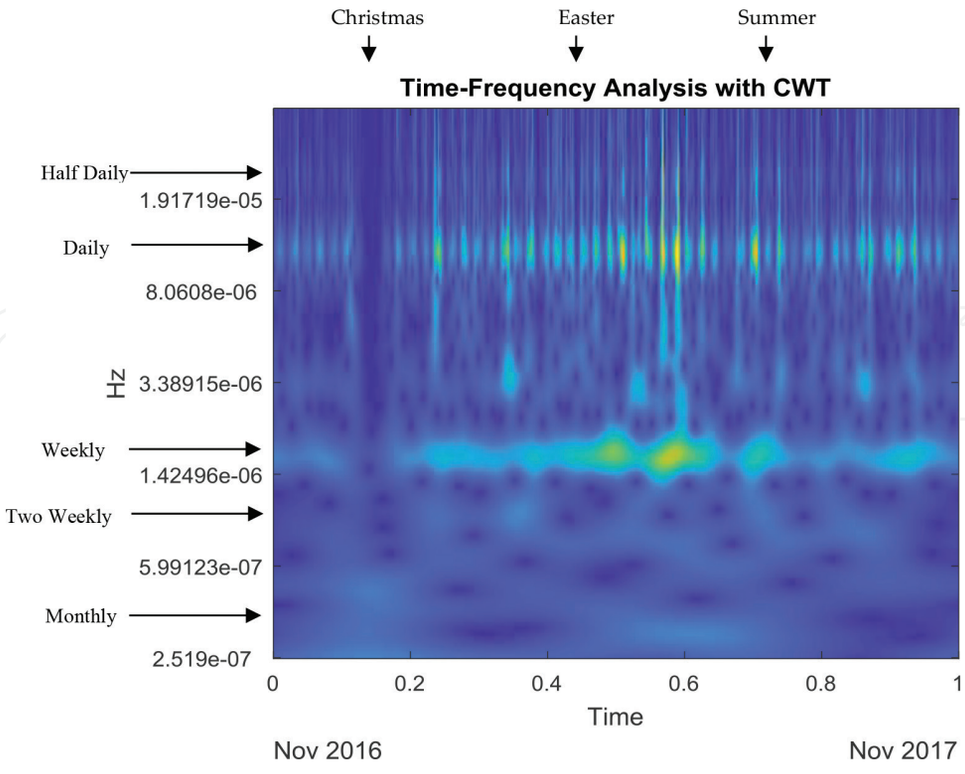

Figure 16. The CWT time-frequency 2D results of the 1-year Email data (Nov 2016–Nov 2017).

4. Conclusions

We present out latest study on using wavelet transform technique for analysing the educational network traffic data. The 2D and 3D presentation network traffic data, i.e. traffic in 24 h and 365 days, helps us to understand better the network traffic pattern. With wavelet transform, we are able to perform network traffic data decomposition and data denoising. With continuous wavelet transform (CWT), we can analyse the data and show how the frequency content of the data changes over time. The CWT analysis shows different characteristics of total traffic data, WWW data and Email data. This time dependent frequency varying information, which is lacking in other techniques, such FFT, is very useful for network traffic analysis. By using CWT, we can easily identify the event which is otherwise difficult to identify in the original time domain.

Acknowledgements

We thank London South Bank University for the financial support of this study.

Author details

Shwan Dyllon[1] and Perry Xiao[2]*

*Address all correspondence to: xiaop@lsbu.ac.uk

1 Department of ICT, London South Bank University, London, UK

2 School of Engineering, London South Bank University, London, UK

References

[1] Zhang J, Song C, Hu Y, Yu B. Improving robustness of robotic grasping by fusing multisensor. In: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. 2012. pp. 126-131

[2] Bermeo JP, Castillo H, Armijos X, Jara JD, Sanchez F, Bermeo H. Artificial Neural Network and Monte Carlo Simulation in a Hybrid Method for Time Series Forecasting with Generation of L-Scenarios. 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), Toulouse; 2016. pp. 665-670. DOI: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0110

[3] Guo T, Xu Z, Yao X, Chen X, Aberer K, Funaya K. Robust online time series prediction with recurrent neural networks. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). 2016. pp. 816-825. Available from: http://ieeexplore.ieee.org/document/7796970/

[4] Rocha E, Salvador P, Nogueira A. A real-time traffic classification approach. In: 2011 International Conference for Internet Technology and Secured Transactions (ICITST); December. 2011. pp. 620-626

[5] Han L, Huang L, Hu Q, Han X, Shi J. Fast Fourier transform based IP traffic classification system for SIPTO at H(e)NB. In: 7th International Conference on Communications and Networking in China. 2012. pp. 430-435. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6417521

[6] Shang P, Li X, Kamae S. Chaotic analysis of traffic time series. 2005;**25**:121-128. http://doi.org/10.1016/j.chaos.2004.09.104

[7] Feng H, Shu Y, Yang OWW. Nonlinear analysis of wireless LAN traffic. Nonlinear Analysis: Real World Applications. 2009;**10**(2):1021-1028. Available from: http://www.sciencedirect.com/science/article/pii/S1468121807002386

[8] Drew M. A frequency analysis approach for categorizing air traffic behavior. In: 14th AIAA Aviation Technology, Integration, and Operations Conference; Atlanta, GA; 16-20 June. 2014. Available from: https://www.aviationsystemsdivision.arc.nasa.gov/publications/2014/AIAA-2014-2420.pdf

[9] Song R, Mason PC, Li M. Enhancement of frequency-based wormhole attack detection. In: Proceedings of the IEEE Military Communications Conference (MILCOM). 2011. pp. 1139-1145

[10] Kim SS, Reddy ALN, Vannucci M. Detecting traffic anomalies using discrete wavelet transform. 2004. pp. 951-961. Available from: http://link.springer.com/chapter/10.1007%2F978-3-540-25978-7_96

[11] Weng H, Lau K-M. Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical western Pacific. Journal of the Atmospheric Sciences. 1994;**51**:2523-2541

[12] Daubechies I. The wavelet transform time-frequency localization and signal analysis. IEEE Transactions on Information Theory. 1990;**36**:961-1004

[13] Lindsay RW, Percival DB, Rothrock DA. The discrete wavelet transform and the scale analysis of the surface properties of sea ice. IEEE Transactions on Geoscience and Remote Sensing. 1996;**34**:771-787

[14] Mak M. Orthogonal wavelet analysis: Interannual variability in the sea surface temperature. Bulletin of the American Meteorological Society. 1995;**76**:2179-2186

[15] Kaiser G. A Friendly Guide to Wavelets. Cambridge, MA: Birkhäuser Boston; 1994. 300 pp

[16] Dyllon S, Dahimene1 H, Wright P, Xiao P. Analysis of HTTP and HTTPS usage on the university internet backbone links. Journal of Industrial and Intelligent Information. 2014;**2**(1):67-70. Available at: http://www.jiii.org/index.php?m=content&c=index&a=show&catid=37&id=88

[17] Dyllon S, Saravanan D, Xiao P. The Usage Analysis of Web and Email Traffic on the University Internet Backbone Links. In: Proceedings of the International MultiConference of Engineers and Computer Scientists. Hong Kong; March 18-20, 2015;**I**:18-21

[18] Wavedec. Multilevel 1-D wavelet decomposition. https://uk.mathworks.com/help/wavelet/ref/wavedec.html?searchHighlight=wavedec&s_tid=doc_srchtitle

IntechOpen