DNS Traffic Forecasting Using Deep Neural Networks

Diego Madariaga[1], Martín Panza[1], and Javier Bustos-Jiménez[1]

NIC Chile Research Labs, Universidad de Chile
{diego,martin,jbustos}@niclabs.cl

Abstract. With the continuous growth of Internet usage, the importance of DNS has also increased, and the large amount of data collected by DNS servers from users' queries becomes a very valuable data source, since it reveals user patterns and how their Internet usage changes through time. The periodicity in human behavior is also reflected in how users use the Internet and therefore in the DNS queries they generate. Thus, in this paper we propose the use of Machine Learning models in order to capture these Internet usage patterns for predicting DNS traffic, which has a huge relevance since a big difference between the expected DNS traffic and the real one, could be a sign of an anomaly in the data stream caused by an attack or a failure. To the best of the authors' knowledge this is the first attempt of forecasting DNS traffic using Neural Networks models, in order to propose an unsupervised and lightweight method to perform fast detection of anomalies in DNS data streams observed in DNS servers.

Keywords: DNS traffic · Forecasting · Machine Learning.

1 Introduction

Internet has experienced a huge growth during the last decade undeniably, as well as DNS system along with it, which has become a fundamental part of Internet working. Both systems, the whole Internet and DNS, have also become more complex due to the constant technological development, emerging of new needs, and the study and successful correction of failures caused by either errors or by the design of varied attacks against the vulnerabilities of the systems' infrastructure. Unfortunately, it is impossible to assure a complete protection of the systems as new unknown attacks and failures are always possible. However, due to its importance it is indispensable to keep the correct working of DNS, since if this is not achieved, the users would be seriously affected by the system's failures. This is especially important for DNS operators, who are responsible for responding to any existing failure.

The early detection of possible failures or attacks is a powerful aid for keeping the correct working of any system. That is why the automatic detection of anomalous events on computer networks issue has taken big relevance in the last years. However, in addition to the innate difficulty of this problem, when we

refer to real world DNS we are referring to large amount of data, given the huge and growing number of users who are constantly querying the system's servers. This demands automatism and efficiency to the systems that perform anomaly detection in real-time.

Forecasting the system's behavior would contribute to the solution to detect anomalies in network traffic, as finding big differences between what is expected, given the system's past information, and the encountered values would give a quick sign of an anomaly occurring in the data stream. This is specially plausible in DNS, where data behavior is strongly influenced by human patterns which present a strong periodicity.

This work performs DNS traffic forecasting on real DNS data from the Chilean country code top-level domain '.cl' using Machine Learning models. It also proposes the development of an unsupervised and lightweight method, and its usage for fast anomaly detection in DNS data streams of DNS servers.

To the best of the authors' knowledge, this is the first attempt of forecasting DNS traffic using Neural Networks models, with an important future usage for early detection of anomalies on DNS traffic.

2 Related Work

Network traffic prediction is a task that has been continuously studied from a lot of different points of view, focusing on specific portions of the whole network traffic, according to the network protocol, or the target users to study.

Among the important studies about traffic forecast, some that can be mentioned are the prediction of IP backbone traffic using AutoRegressive Integrated Moving Average (ARIMA) models [14], the prediction of data traffic belonging to most popular TCP ports using AutoRegressive Moving Average (ARMA) models [3], and the prediction of TCP/IP traffic using both ARIMA and Neural Networks models [5].

In addition, other research studies have focused on predicting mobile network traffic, implementing models to forecast future call traffic using ARIMA models [1], and based on chaos analysis [9]. Other studies have researched mobile Internet usage, forecasting mobile Internet data traffic using Seasonal ARIMA (SARIMA) models for 2G GSM networks [18], 2G/3G networks[22], and 3G/4G LTE networks [10].

In the area of automatic detection of network traffic anomalies and attacks there exist relevant works, which are closely related to DNS traffic since both general network traffic and DNS traffic are similar, due to how DNS process works. That is why anomalies and attacks are also visible in traffic at DNS level. Most of these studies perform event detection using supervised learning techniques, training the models with specific instances of pre-classified events (attacks or failures) in order to recognize them in future network flows [2].

Also, there are implemented systems that establish certain rules to find different kinds of anomalous events like network intrusion detection [17, 15]. Moreover, the problem of detecting network anomalies by using unsupervised learning

methods has been studied with the purpose of recognizing and detecting anomalies from unknown threats, by performing outlier detection in network traffic [4].

With respect to DNS, there are studies that analyze and propose supervised methods to detect some specific DNS attacks, such as DoS [7], DDoS [11] [6], Domain Fluxing [23], Botnet Domains and Malware [20], and Kaminsky Cache Poisoning [12]. Nevertheless, there are not deep studies with regard to the use of unsupervised learning techniques in DNS data or to the use of big amount of data from real DNS queries. Furthermore, DNS data, that corresponds mostly to UDP packets and uses particular rules, has not yet been exploited to perform analysis, compared to other types of network traffic flows.

Close to the goals of this paper, people from InternetNZ, the registry for the '.nz' ccTLD (Country Code Top-Level Domain) from New Zealand, showed in their blog the use of the Prophet forecasting model [19] to analyze trends on DNS queries for '.nz' domains. Also, they look into their DNS traffic dataset in order to find past anomalies in the stream by visual inspection [16]. Our work proposes the use of Neural Networks models to forecast DNS data traffic and to propose the implementation of an automatic lightweight method to perform early anomaly detection.

3 Data Set Overview

The data used in this work consists of a month of normal operation traffic of one of NIC-Chile's authoritative DNS servers. It starts on 2 October, 2017, until 1 November of the same year. NIC-Chile is the official registry for the '.cl' ccTLD, which is the geographic top-level domain from Chile. Every DNS packet from queries to the server and responses to users are present in the dataset. The server studied belongs to an anycast configuration along with other servers.

A time series of DNS traffic was built by aggregating all the successful server responses into 1-hour intervals. Therefore, each point of the time series corresponds to the amount of DNS packets from server responses with record types 1, 2, 15 or 28 (A, NS, AAAA, MX) obtained in one hour of data. The whole time series represents a total of 249,434,772 DNS packets.

The Figure 1 below shows a portion of the time series used to train and test the model, corresponding to two weeks of the DNS traffic time series. It is possible to observe some regular patterns on it, represented as periods of one day and one week in the data. This makes perfect sense with the fact that the regularity of human patterns [8, 13] is reflected on Internet usage, producing regular patterns as well [21]. Moreover, there are some visible outliers in the time series shown below, and these anomalies in the traffic could be easily detected by comparing the amount of data traffic measured in a real-time flow and the amount expected, obtained by the use of forecast models.

The data was split into training and testing set, leaving the last complete week for testing and the rest for training. That is roughly a 77% / 23% distribution.

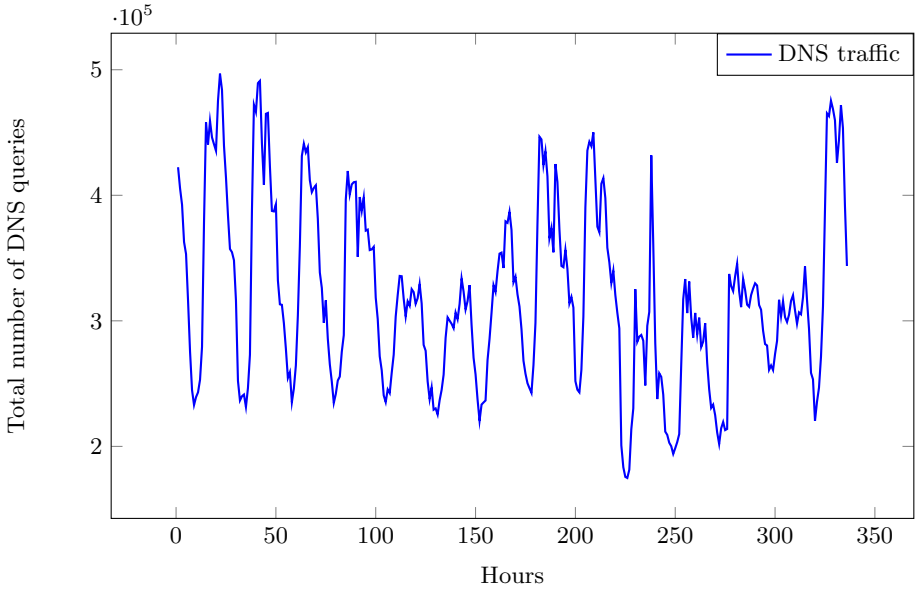

Fig. 1. DNS traffic time series

Since data is real from a normal working of the system, it takes on great importance in the analysis of this work and gives relevance to the results obtained as users patterns are captured in the traffic, showing clear periodicity from day to day.

4 Forecast Models

The selected neural network models follow some commonly used architectures in literature that address time series analysis problems.

To evaluate these forecast models, a more basic forecast model is proposed for establishing a comparison on some forecast errors indicators obtained at predicting DNS traffic. The selected comparison model is Weighted Moving Average.

4.1 Neural Networks

Artificial Neural Network models, based on the natural neural networks of the human brain, consist of the interconnection of several nodes that individually define weights and operations to transmit data over their own connections and thus, over the network. The storage and update of the data perceived on the nodes allow the network training and learning from the input, in order to give logical predictions to future information. The different configuration of the connections between the nodes leads to different types of networks, some of which are commonly used at time series forecasting.

LSTM Long Short-Term Memory is a type of recurrent neural network. That is, a network whose connections contain loops in order to keep information, passing it through the steps of the network. An LSTM-layer's special feature is that it can retain long-term information and learn when to get or not get it into account. The implementation of an LSTM unit consists of three gates: input gate, output gate and forget gate that are related according to the following equations:

$$c_t = i_t \circ tanh(W_c x_t + V_c y_{t-1} + b_c) + f_t \circ c_{t-1} \tag{1}$$

$$y_t = o_t \circ tanh(c_t), \tag{2}$$

where i_t, f_t, o_t are the respective activation functions of each gate:

$$g_t = sigmoid(W_g x_t + V_g y_{t-1} + b_g), \tag{3}$$

where g is the corresponding gate, W and V are weight matrices, b is a bias vector, x_t and y_t are input and output vectors of the step t, and $\circ$ corresponds to the entry-wise product between two matrices.

CNN-LSTM Convolutional Neural Network (CNN) differs from a normal artificial neural network at its use of kernels to apply a convolutional operation over the input data in order to transform it and obtain specific information to focus on. The equation of a convolutional layer for a 2-dimensional input is described below:

$$X_{ij} * K_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{n-1} w_{ab} X_{(i+a)(j+b)} \cdot K_{ab} + b \tag{4}$$

where X is the input and K is the $m \times n$ kernel. w and b are weight and bias respectively.

CNN-LSTM would be the combination of the two networks mentioned above. It will consist of a convolutional layer before an LSTM layer.

4.2 WMA

Weighted Moving Average, commonly used to smooth out short-term fluctuations and highlight longer-term trends in time series, can be defined as a forecasting model where the next forecast values are calculated as a weighted average of past values, where the weights decrease in arithmetical progression. A weighted moving average with linear decreasing weights and a window k is defined as follows:

$$WMA_n = \frac{y_n \times k + y_{n-1} \times (k-1) + ... + y_{n-k} \times 1}{k + (k-1) + ... + 1} \tag{5}$$

5 Accuracy Measures

In order to compare the obtained results by the different forecast models proposed, it is necessary to establish the metrics to consider during the evaluation process. The selected metrics are divided in two groups:

5.1 Prediction Errors

Firstly, is necessary to evaluate the forecast results with regard to the difference between real and predicted values, for which the following measures are considered:

1. **Root-Mean-Square-Error** (RMSE): Square root of the average of the square of vertical distance between each real value and its forecast.

$$RMSE = \sum_{i=1}^{n} \sqrt{\frac{(r_i - p_i)^2}{n}}$$

2. **Mean Absolute Error** (MAE): Mean of the vertical distance between each real value and its forecast.

$$MAE = \sum_{i=1}^{n} \frac{|r_i - p_i|}{n}$$

5.2 Time Series Distances

In addition to the prediction errors, it is important to take into account other factors when performing an evaluation of the forecast models, mainly because the prediction errors mentioned above do not consider important characteristics of the nature of a time series when comparing them, such as the fact that the points of time series have a logical order, given by the time. Therefore, another way to measure accuracy of forecast results is needed, in order to capture the similarity between the shape of real data and predicted curves or the data distribution. The following distance measures, commonly used for time series clustering, are considered:

1. **Edit Distance for Real Sequences** (EDR): Compares the two time series in terms of how many edit operations (delete, insert or replace) are necessary to transform one curve to the other. The distance between two points is reduced to 0 or 1, where if the distance between two points r_i and p_j is less than a given ϵ, then the points are considered equal.
2. **Dynamic Time Warping** (DTW): It is a time series alignment algorithm that aims at aligning two sequences by warping the time axis iteratively until an optimal match between the two sequences is found, which minimize the sum of absolute differences for each matched pair of indices. An acceptable match must follow four conditions:

- Every index from the first sequence must be matched with one or more indices from the other sequence, and vice versa.
- The first index from the first sequence must be matched with the first index from the other sequence.
- The last index from the first sequence must be matched with the last index from the other sequence.
- The mapping of the indices from the first to the second sequence must be monotonically increasing, and vice versa.

6 Experimental Results

As mentioned in Section 3, the DNS traffic time series with information about 1-hour aggregated data is forecasted sequentially hour by hour until completing one week of predicted data, i.e. 168 forecast operations.

Figures 2 and 3 show the forecast results of performing DNS traffic forecast using LSTM and CNN-LSTM methods described in Section 4. Also, Figure 4 shows the results obtained by using the Weighted Moving Average as basic forecast model for comparison.

The accuracy measures obtained for each forecast model, are presented in Table 1.

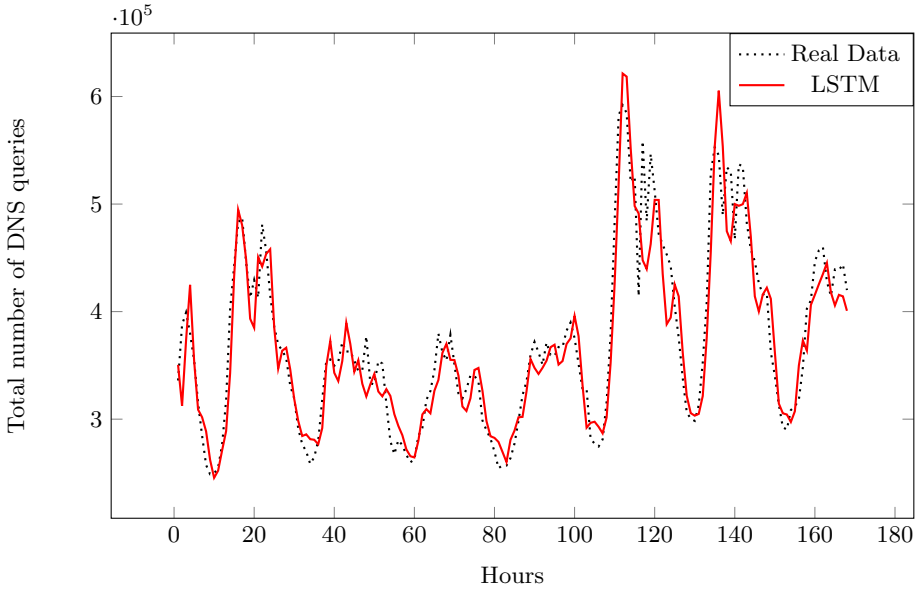

Fig. 2. DNS traffic forecast using LSTM network

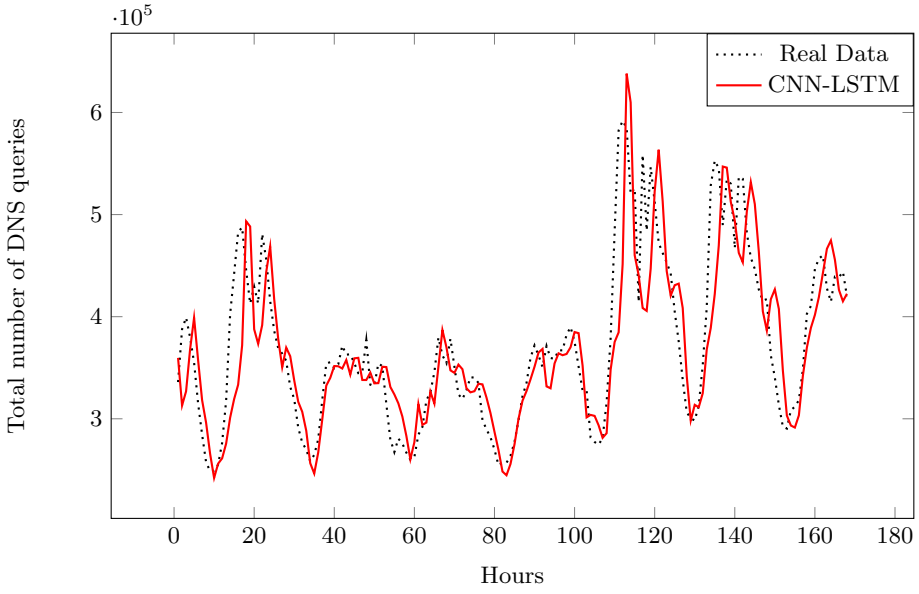

Fig. 3. DNS traffic forecast using CNN-LSTM network

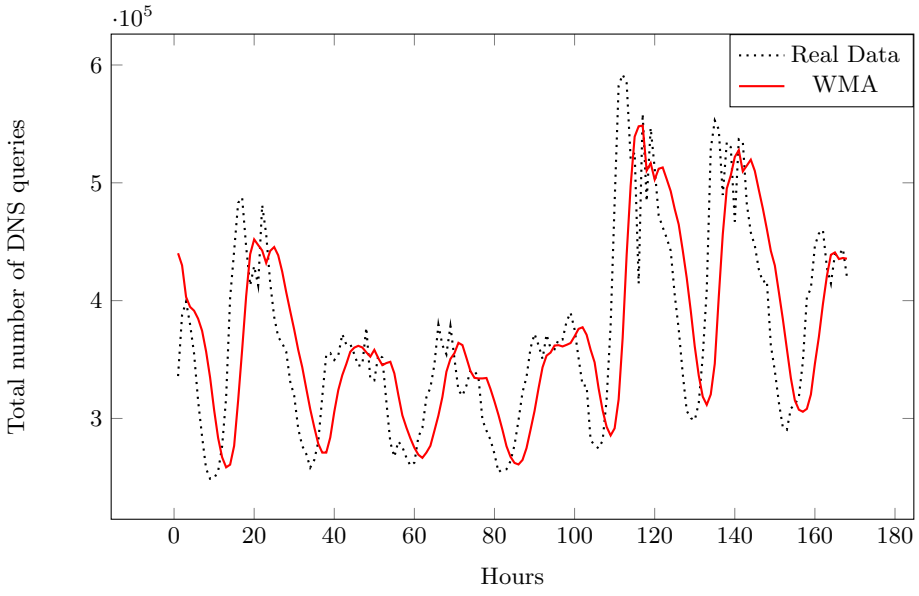

Fig. 4. DNS traffic forecast using Weighted Moving Average

Table 1. DNS traffic forecast errors

Method	RMSE	MAE	EDR	DTW
WMA	68330	50952	68	3910580
LSTM	**28980**	**21902**	60	3710717
CNN-LSTM	48752	34432	**53**	**3531212**

7 Discussion

As Figures at Section 6 illustrate, all models show positive visual results, as the predicted curves captures the basic periodic pattern present in data. Both LSTM and CNN-LSTM models are capable of predict almost all sudden changes in real data curve thanks to their more risky behavior at forecasting, obtaining results quite fit the real values. In the other hand, WMA results are basically presented as a smoothing of the real curve, without foreseeing the most interesting points of real data. Also, it is also remarkable the fact that WMA model takes a couple of points before detecting the change of phase in the day periodicity, presenting clearly a big delay with respect to the real curve. Both neural network models show a similar behavior, as observed in Figures 2 and 3, where LSTM demonstrates a slight improve over CNN-LSTM when detecting changes in the periodicity phase.

With regard to the accuracy measures presented in Table 1, neural network models outperformed WMA model in every distance and error measure, proving that results obtained by LSTM and CNN-LSTM have a better accuracy in average and also, that their curves are more similar in shape to the real one. At comparing both neural network models, LSTM obtains lower prediction error results (RMSE and MAE), while CNN-LSTM obtains lower time series distances (DTW and EDR).

8 Conclusions and Future Work

Basing on the results obtained, this work concludes the feasibility of forecasting DNS traffic using Machine Learning models. Considering real data from a constantly and massively queried domain that is '.cl', this paper states its results as representative for similar DNS traffic. Specially since it comes from normal human activity, which presents a strong periodicity. Since neural networks outperformed WMA model of comparison in the accuracy measures presented in Section 5, their effectiveness to the solution of this forecasting problem is concluded. They were able to capture the periodic patterns in the time series and performed better than WMA when detecting abrupt phase changes. As they

did not focus on outliers, this work also concludes a utility for these results to perform outlier detection on data streams.

An improvement in the results obtained in this work could be achieved firstly, by developing more specialized and complex deep learning architectures. This is a tough work that is out of the reach of this paper as it used well-known models in literature. Secondly, external factors could be added to the models' training to accomplish a better understanding of the traffic behavior.

As mentioned before in this paper, the good performance obtained at forecasting DNS traffic is a very valuable result especially for DNS providers since the forecast models can be easily used to develop a lightweight monitoring tool to continuously analyze traffic in DNS servers, where a big difference between the real amount of DNS data experienced and the forecast levels of traffic could be an early sign of the presence of an anomaly in the data stream produced by an attack or a failure.

Also, as future work it is important to carry out deeper analysis about DNS traffic data, performing smart data aggregation to try to infer more detailed reasons of unexpected traffic anomalies. Possible ways of performing this aggregation is by grouping DNS queries according to their query type (A, AAAA, MX, NS), by applying subnet mask to the sender IP addresses or by grouping queries for similar domains, where domains are considered similar if they are requested in similar way, i.e. their DNS traffic time series are close in distance. By doing this, it could be possible to perform forecasts of different portions of the complete DNS traffic, giving to DNS providers the possibility to detect in which specific portion of the whole stream there is an anomalous event.

In relation to the aforementioned, the first steps have been taken in order to group queries for similar domains, by creating time series with the portion of the DNS traffic according to each requested domain and performing clustering by using time series K-means algorithm with Dynamic Time Warping distance. Figure 5 shows the results obtained by performing *Elbow method* to find an appropriate number of clusters. According to these results, a good approach for a future work would be to use nine clusters, analyze their content, and to use the forecast models described in this paper to perform more accurate and understandable anomalous event detection.

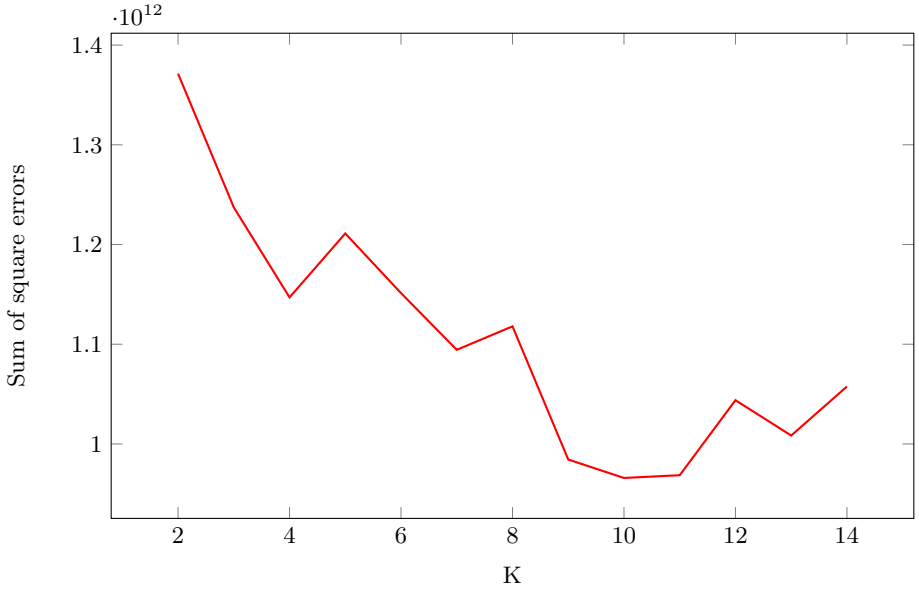

Fig. 5. Elbow method for DNS traffic clustering

References

1. Akinaga, Y., Kaneda, S., Shinagawa, N., Miura, A.: A proposal for a mobile communication traffic forecasting method using time-series analysis for multi-variate data. In: Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE. vol. 2, pp. 6–pp. IEEE (2005)
2. Alsirhani, A., Sampalli, S., Bodorik, P.: Ddos attack detection system: Utilizing classification algorithms with apache spark. In: New Technologies, Mobility and Security (NTMS), 2018 9th IFIP International Conference on. pp. 1–7. IEEE (2018)
3. Basu, S., Mukherjee, A., Klivansky, S.: Time series models for internet traffic. In: INFOCOM'96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE. vol. 2, pp. 611–620. IEEE (1996)
4. Casas, P., Mazel, J., Owezarski, P.: Unsupervised network intrusion detection systems: Detecting the unknown without knowledge. Computer Communications **35**(7), 772–783 (2012)
5. Cortez, P., Rio, M., Rocha, M., Sousa, P.: Multi-scale internet traffic forecasting using neural networks and time series methods. Expert Systems **29**(2), 143–155 (2012)
6. Douligeris, C., Mitrokotsa, A.: Ddos attacks and defense mechanisms: classification and state-of-the-art. Computer Networks **44**(5), 643–666 (2004)
7. Feinstein, L., Schnackenberg, D., Balupari, R., Kindred, D.: Statistical approaches to ddos attack detection and response. In: null. p. 303. IEEE (2003)
8. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. nature **453**(7196), 779 (2008)

9. Hu, X., Wu, J.: Traffic forecasting based on chaos analysis in gsm communication network. In: Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on. pp. 829–833. IEEE (2007)
10. Miao, D., Qin, X., Wang, W.: The periodic data traffic modeling based on multiplicative seasonal arima model. In: Wireless Communications and Signal Processing (WCSP), 2014 Sixth International Conference on. pp. 1–5. IEEE (2014)
11. Mirkovic, J., Reiher, P.: A taxonomy of ddos attack and ddos defense mechanisms. ACM SIGCOMM Computer Communication Review **34**(2), 39–53 (2004)
12. Musashi, Y., Kumagai, M., Kubota, S., Sugitani, K.: Detection of kaminsky dns cache poisoning attack. In: Intelligent Networks and Intelligent Systems (ICINIS), 2011 4th International Conference on. pp. 121–124. IEEE (2011)
13. Oliveira, E.M.R., Viana, A.C., Sarraute, C., Brea, J., Alvarez-Hamelin, I.: On the regularity of human mobility. Pervasive and Mobile Computing **33**, 73–90 (2016)
14. Papagiannaki, K., Taft, N., Zhang, Z.L., Diot, C.: Long-term forecasting of internet backbone traffic: Observations and initial models. In: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies. vol. 2, pp. 1178–1188. IEEE (2003)
15. Paxson, V.: Bro: a system for detecting network intruders in real-time. Computer networks **31**(23-24), 2435–2463 (1999)
16. Qiao, J.: .nz dns traffic: Trend and anomalies (2017), https://blog.nzrs.net.nz/nz-dns-traffic-trend-and-anomalies/
17. Roesch, M., et al.: Snort: Lightweight intrusion detection for networks. In: Lisa. vol. 99, pp. 229–238 (1999)
18. Shu, Y., Yu, M., Liu, J., Yang, O.W.: Wireless traffic modeling and prediction using seasonal arima models. In: Communications, 2003. ICC'03. IEEE International Conference on. vol. 3, pp. 1675–1679. IEEE (2003)
19. Taylor, S.J., Letham, B.: Forecasting at scale. The American Statistician **72**(1), 37–45 (2018)
20. Thomas, M., Mohaisen, A.: Kindred domains: detecting and clustering botnet domains using dns traffic. In: Proceedings of the 23rd International Conference on World Wide Web. pp. 707–712. ACM (2014)
21. Wang, H., Xu, F., Li, Y., Zhang, P., Jin, D.: Understanding mobile traffic patterns of large scale cellular towers in urban environment. In: Proceedings of the 2015 Internet Measurement Conference. pp. 225–238. ACM (2015)
22. Xu, F., Lin, Y., Huang, J., Wu, D., Shi, H., Song, J., Li, Y.: Big data driven mobile traffic understanding and forecasting: A time series approach. IEEE transactions on services computing **9**(5), 796–805 (2016)
23. Yadav, S., Reddy, A.K.K., Reddy, A.N., Ranjan, S.: Detecting algorithmically generated domain-flux attacks with dns traffic analysis. IEEE/Acm Transactions on Networking **20**(5), 1663–1677 (2012)