Chapter 23
The Improved LSTM and CNN Models for DDoS Attacks Prediction in Social Media

Rasim M. Alguliyev
Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

Ramiz M. Aliguliyev
https://orcid.org/0000-0001-9795-1694
Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

Fargana J Abdullayeva
https://orcid.org/0000-0003-2288-6255
Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

ABSTRACT

Automatic identification of conversations related to DDoS events in social networking logs helps the organizations act proactively through early detection of negative and positive sentiments in cyberspace. In this article, the authors describe the novel application of a deep learning method to the automatic identification of negative and positive sentiments in large volumes of social networking texts. The authors present classifiers based on Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to address this problem domain. The improved CNN and LSTM architecture outperform the classification techniques that are common in this domain including classic CNN and classic LSTM in terms of classification performance, which is measured by recall, precision, f-measure, train loss, train accuracy, test loss, and test accuracy. In order to predict the occurrence probability of the DDoS events the next day, the negative and positive sentiments in social networking texts are used. To verify the efficacy of the proposed method experiments is conducted on Twitter data.

DOI: 10.4018/978-1-7998-5348-0.ch023

1. INTRODUCTION

Recently, cyber-attacks have become widespread, targeting giant corporations such as Sony, Verizon, Yahoo, Target, JP Morgan, Ashley Madison, and government agencies. Cyber-attacks are cause of leakage of sensitive information of users, loss of lives, the destruction of critical infrastructures.

The most common cyber-attacks are DDoS (Distributed Denial of Service) attacks (Kaur et al., 2017), uses multiple compressed systems to cut or stop the services of hosts connected to the Internet (Carl et al., 2006). Usually, web servers of the bank or credit card payment networks are the target of such attacks. Therefore, a single attack may cause considerable loss (Matthews, 2014). Detecting and predicting DDoS attacks is a challenging task (Bleakley & Vert, 2014; Imamverdiyev & Abdullayeva, 2018). The purpose of the traditional DDoS detection system is to distinguish malicious packet traffic from normal traffic (Mirkovic & Reiher, 2004). The malicious traffic in the network occurs after the DDoS attack takes place. In the detection of DDoS attacks prior to occurring the data of the social network have a great importance. On the basis of social media data, it is possible to track the traces of subjects targeted by the object.

Most information security experts believe that hacking attacks on businesses will be carried out through social media channels. Facebook, LinkedIn, Twitter are the most widely used networks. Social networks, besides allowing people to connect with each other, but also become a powerful political tool (Hua et al., 2013). Social media is regarded as the next big cybercrime vector (George, 2014).

At usual social media is considered as a sensor that collects information about various social events such as, disease epidemics, protests, elections and so on. The exponential growth of data containing the society opinion in the Web environment led researchers to focus on opinion mining and sentiment analysis of social media data (Ebrahimi et al., 2016). Among social media websites, the Twitter is a site that publishes more information on social issues, natural disasters, incidents and DDoS attacks planning. By analyzing Twitter, it is possible to identify the discussed events that will be occurred and analyze the trajectories (sources) of these events. Additionally, when analyzing the sentiments of the peoples related to the events which will be occurring, it is possible to get a lot of information about a certain event. Analysis of the sentimental traces allows to conduct the sentiment analysis by space and time, and predict the sentiments of the users in advance.

In (Liu & Zhang, 2012), the review of the various approaches related to opinion mining and sentiment analysis is provided. In (Jiang et al., 2011), the method for the providing classification of the sentiments in social media discussions into positive, negative, and neutral classes is proposed. This is a targeted sentiment analysis.

Another application area of the targeted sentiment analysis is to determine what do think people of one country about people of another country. In (Chambers et al., 2015), in order to model the relations between states, the "country-to-country sentiment data" are used. The data classification here is provided based on Bootstrapped classifiers.

The subject of the sentiment analysis is a text. There are two methods of sentiment analysis:

1. **Dictionary-based methods:** In (Taboada et al., 2011), sentiment analysis method, named as SO-CAL (Semantic Orientation CALculator) is proposed. Here in order to classify positive and negative sentiment, the dictionary is used. In this approach, each word is assigned a numerical value;
2. **Machine learning methods:** In the machine learning based sentiment analysis method, by using the statistical method called word embedding, each word is assigned values as a vector form and

the model trains these digitized sentences using machine learning or deep learning methods. SVM, Random Forest, and Naïve-Bayes are traditional machine learning methods, but CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit) are deep learning methods. In (Yoo et al., 2018), a system that analyzes and predicts the sentimental trajectory of the users, based on the events, recorded in real time discussions is proposed. Here the trajectory analysis and the sentiment analysis are both practically tested. To analyze and predict the sentiment, a deep learning method is applied and high results are obtained. To detect the events, the words such as crime, disaster, accident are used.

To predict cyber-attacks based on social media data, the extensive research is conducted. In (Khandpur et al., 2017), for the detection of the large-scale cyber-attacks such as DDoS, data breaches, and account hijacking, the supervised detection method based on social media data on Twitter is proposed. In this work, for the feature extraction and semantic structure modeling, the convolution kernels and dependency parses approaches are used. In (Ritter et al., 2015), the detection issue of DoS, data breaches and account hijacking attacks in Twitter discussions is considered. For this purpose, in the proposed approach, the label regularization, constrained semi-supervised EM and one-class SVM are used. The main objective of study (Suarez et al., 2018) is the monthly forecasting of the Twitter discussions about security attacks. For the detection of the incidents, l_1 regularization is used. To create security alarms based on users' sentiment data, the analysis of the Twitter discussions is conducted. This analysis is carried out on the basis of comparison of three supervised training algorithms which are Bayesian classification, Support Vector Machines and maximum entropy for text classification and the classifier with the best classification result is selected as the main classification model. After the implementation of the classification by applying the l_1 regularized regression, the forecasting is conducted. Regression is the best tool for predicting events given as linearly independent observations. In (Lippmann et al., 2017), for the detection of the cyber-attacks discussions provided on the Stack Exchange, Reddit and Twitter page the classifier based on hybridization of the TF–IDF, logistic regression, and linear SVM methods is proposed. In this work, to identify cyber-attack discussions, the keyword-based approach is used. The proposed approach searches on the basis of 200 keywords and phrases and calculates the frequency of these words in the document. A document containing more frequency number of keywords is assumed that document is more relevant to cyber attack topic. To implement the classification, the keywords such as "kit", "ınfected", "checksum", and the phrases such as "buffer overflow", "privilege escalation", "Distributed Denial of Service" are used. Here, the classifier generates a probability value that determines whether the document is related to the cyber attack topic discussion. This probability value allows assigning the document into the cyber or non-cyber classes.

Existing studies are based on the idea of training the system on classified samples and a fixed number of features. Such approaches cannot detect the cyber-attacks in dynamic nature.

The purpose of this paper is by means of self-learning methods to predict the likelihood of cyber-attack related words, interpreted in text type social media discussions. Attackers create unpleasant or negative sentiments in the social network texts against target object. The purpose of this paper is to predict the next day occurrence probability of the DDoS event, based on the social media discussions. Here as the time interval, a daily forecast is taken, however, the proposed method can make predictions for different time periods too. The input data of the model is text streams. For the converting words into vector form, embedding method is used.

The difference between this work from the existing ones is that there is no need to know the class of samples in advance.

The main contributions of this paper are:

- To predict the DDoS attack occurrence probability based on social media data, CNN model with 13 layers and improved LSTM model is proposed;
- In the classification of the data, the class labels are not used;
- In the detection of positive and negative sentiments, feature extraction and selection are not performed.

This paper consists of the following sections: In section 2, Backgraund study presented. Section 3 describes the problem statement formulation. Section 4 summarizes some of the methods used in the DDoS prediction based on social media data. In section 5, an improved CNN model is provided. In section 6, an improved LSTM model is provided. In section 7, the results of the comparative analysis of the proposed method with existing methods are described. Section 8 presents the conclusion of this work.

2. BACKGROUND STUDY

2.1. Information Security Events

2.1.1. Denial of Service Attacks (DoS)

The DoS is designed to deny the liveness properties (e.g. uptime) of a web service to other users. These attacks are most often accomplished by an agent who amplifies requests for a network service with no other intention but to saturate the service beyond some capacity of the resources behind that device (e.g. bandwidth, processing or memory).

2.1.2. Data Breach

Data breach is an attack which implements sophisticated techniques to pilfer a collection of personal or digital credentials. The effects of a data breach if unmitigated may result in the fraudulent use of personal information. However, early detection may alert affected users to monitor for fraud and initialize preventative measures such as updating credentials. Data breach attacks my elicit signals from social media such as early discovery and warnings generated by affected users who discover fraud, further this may be useful to other users whose stolen personal information (e.g. credit card) has not yet been exploited.

2.1.3. Account Hijacking

Account hijacking may involve an intruder guessing, cracking, or using default passwords to gain unauthorized access to user accounts or system privileges. Account hijacking usually focuses on the problem of determining an unknown user password by using techniques including brute force attacks, dictionary attacks (using frequently used passwords).

2.2. Cybersecurity Data

Sources of cybersecurity data in the field of information security are divided into two classes:

1. **Formal sources:** For example, the NIST National Vulnerability Database (NVD), United States Computer Emergency Readiness Team (US-CERT) and so on;
2. **Non-formal sources:** For example, developer forums, chat rooms and social media platforms like Twitter, Reddit, and Stack Overflow.

These sources publish information about security vulnerabilities, threats, and attacks. Automatic retrieval appropriate information relevant to the field from OSINT (Open-source intelligence) data is one of the key issues that attract the attention of researchers. OSINT covers data collected from open sources, such as newspapers, magazines, social networking sites, video sharing sites, wiki pages, blogs, and so on.

DDoS Cyber-attacks Data. Statistical information on DDoS type cyber attacks is available at www.digitalattackmap.com. The purpose of this website is to visualize global DDoS attacks and is created with a collaborative effort of the Arbor Networks and Google Ideas organizations.

In this paper, to analyse DDoS events the Twitter data is used.

2.3. Cyber Attacks and Forensic

For disclosure of the cyber-attack crimes, the cyber forensics are used. The forensic of the cyber attacks is the complicated issue (Shackelford, 2009).

Cyber forensic is used for collecting, evaluating and storing evidence from computer-related crimes (Jr & Menendez, 2002). Although the forensic investigation is useful, the ability of this approach to identify the motives behind cyber-attacks is limited.

2.4. Motivations Behind the Cyber-Attacks

Cyber-attacks are viewed as technical and social events (Sakaki et al., 2010). The socio-technical progress of the IT infrastructure of the country and its economy can seriously affect the likelihood of the country being attacked (Mezzour, 2015). Cyber-attacks are closely related to social, political, economic, and cultural (SPEC) motivations (Ghandi et al., 2011). For the prevention of the cyber-attacks effectively, the social and technical progress and the motivation of the cyber-attacks should be taken into account.

3. PROBLEM STATEMENT

The main research question to investigate is whether tweet streams contain useful information for DDoS defense. Our task is to predict the likelihood that a DDoS event will occur to a certain target in the day *d* given the tweet stream over a history period *X* related to the monitored target. *X* is a sequence of N^p days $\left(X = \left\{d^{N^p}, ..., d^2, d^1\right\}\right)$ immediately before *d*, where d^1 is the day before d and $d^i > d^{i+1}$. N^P can be arbitrarily large. The set of tweets posted on d^i is denoted as $d^i = \left\{t^1, t^2, ..., t^{N_i^d}\right\}$ denotes the number of tweets of the day d^i. Each tweet consists of a sequence of words $t^j = \left\{w^1, w^2, ..., w^{N_j^t}\right\}$.

In this work, to transform words into numbers, word embedding method is used. In the input level, we represent each word w^k with a K dimensional embedding, thus mapping a tweet t^j into a matrix:

$$t^j = \left\{e\left(w^1\right), e\left(w^2\right), ..., e\left(w^{N_j^t}\right)\right\} \tag{1}$$

3.1. Primary Goal of Proposal

The primary goal of sentiment prediction in this work we can describe as follows. Assume $D \subset X \times Y$ be the dataset that contains the sentiments, where $X= \{x_1,x_2,\dots,x_n\}$ is the set of sentiments so that $x_i(x_1,x_2,\dots,x_m)^T$ is an m dimension feature vector for i^{th} sentiment. Also, let $Y=\{p,n\}$ be the set of class labels in binary classification problem in which positive and negative sentiments are denoted by p and n, respectively. The goal is to assign the right label from Y to each sentiment.

4. RELATED WORKS

For the predicting DDOS attacks based on social media data, by applying the above-mentioned dictionary-based and machine learning methods, various approaches are proposed.

In (Jiang et al., 2011), social media is used as a crowdsourcing sensor for the getting up insight about cyber-attacks, which willing to occur in the feature. In this work, to detect cyber-attacks such as DDoS, data breaches, and account hijacking, by using seed event triggers, an unsupervised approach is proposed.

In (Chambers et al., 2015), to modeling relationships between states, the Twitter data is used. Here, the experiments are conducted on the state by state sentiment data, and these data is placed on the web page http://www.usna.edu/Users/cs/nchamber/nations/index.html. This dataset consists of information about positive and negative discussions.

In (Sapienza et al., 2018), based on Twitter and Darknet data, a Mirai DDoS attack forecasting method is suggested. To detect attacks, a dictionary with listed attack terms is used. This dictionary contains the names of some malicious ransomware programs, such as wannacry, wannacrypt, petya, wcry, petrwrap and the proposed method can detect those words accurately. Here at first, cyber-security related headers are scanned, then by using the text mining methods, relevant terms are identified and non-relevant terms are removed. The drawback of this approach is that the model is unable to detect new types of attacks when new attack names are not included in the dictionary.

To test the impact of an attack, send attacks and received attacks are taken as dependent variables, and the parameters such as Network Bandwidth, GDP and Internet Users per 100 populations, ICT, CPIA and country-to-country average sentiment score are taken as independent variables, and the correlation between variables is evaluated (Kumar & Carley, 2016a). To perform correlation evaluation, the Pearson correlation method is used. To find the source and target countries of the cyber-attacks the network visualization is used. In this work, by using quadratic assignment procedure (QAP) the correlation between cyber-attack networks was found. From the correlation, it was revealed that countries with high bandwidth and corruption are the best source for DDoS attacks. Because countries with high bandwidth can provide hosting services to any number of computers to implement broadband DDoS attacks. In addition, it is determined that the countries with high Per-capita-GDP indicator and better Information and Communication Technologies (ICT) infrastructure become as targeted countries. In this work along with the attack data, the data showing the country's level taken from the World Bank website, such as the Information and Communication Technologies (ICT) infrastructure, Per-capita-GDP, Country Policy and Institutional Assessment (CPIA) corruption and Internet Users per 100 population data are also used. Here information about international Internet network bandwidth is taken from www.econstats.com, for the creating of the alliance-and hostility network an information from Correlates of War (www.coorelatesofwar.org) web-page is used. For the tracking relation trend of countries toward each other, the USNA (http://www.usna.edu/Users/cs/nchamber/nations/index.html) data is used.

In (Kumar & Carley, 2016b), a cyber-attack detection method based on decision tree algorithm is proposed. Here the attack probability is calculated on the basis of the Bayesian theorem. To track the relationships of the countries toward each other, based on Twitter sentiments (http://www.usna.edu/Users/cs/nchamber/nations/index.html) the sentiment trends are constructed and the comparison of this trend with the trend constructed on the basis of DDoS data, derived from the Arbor Networks (http://www.digitalattackmap.com), is conducted. The experiments suggest that the negative discussions within the discussions conducted against the certain country increase the occurrence probability of the cyber-attacks to that country, and the presence of positive sentiments to the country reduces it. In the paper, the analysis of this landscape in specific countries is described in detail.

In (Mittal et al., 2016), conceptual approach called CyberTwitter is proposed, which analysis cybersecurity-related discussions and generates timely alarms for security analysts. To conduct the experiments, the OSINT dataset is used.

The fact that the data on Twitter has real-time nature, it is enabled researchers to make important decisions from highly influential events. This type of information is used in the analysis of emergency events, such as earthquakes (Sakaki et al., 2010), forest fires (Longueville et al., 2009), terrorist attacks (Oh et al., 2011), natural disasters (Vieweg et al., 2010) and so on. These applications of Twitter have turned it into the most reliable source of OSINT data.

In the Twitter, many companies, such as Adobe (@AdobeSecurity), Github (@githubstatus), WhatsApp (@wa status) publish information about security incidents, related to their products. Here individual users also publish information about encountered new gaps.

The comparative description of the different approaches by various metrics is given in Table 1.

In the methods mentioned in Table 1, attack detection is provided based on the training of the system on classified samples. This approach is not suitable for attack detection in dynamic environments.

To predict the likelihood of cyber-attack related words, interpreted in text type social media discussions self-learning methods are needed.

5. AN IMPROVED CNN MODEL

An improved Convolutional Neural Network (CNN) is employed in this research. CNN is a subset of deep learning which has attracted a lot of attention in recent year. The CNN architecture consists of three different types of the layer: convolutional layer, pooling layer, and a fully connected layer (Goodfellow & Bengio, 2016).

Table 1. Comparative description of existing approaches by various metrics

	Reference	Dataset	Aim	Method	Classes
1.	(Chambers et al., 2015)	Country-to-country sentiment data http://www.usna.edu/Users/cs/nchamber/nations/index.html	Detection of international relations based on social media	Bootstrapped classifiers	Positive tweets, Negative tweets
2.	(Kumar & Carley, 2016b)	DDoS Cyber-attacks Data collected from the website www.digitalattackmap.com; country-to-country sentiment data http://www.usna.edu/Users/cs/nchamber/nations/index.html	Cyber-attack detection	Decision tree algorithm, Bayesian theorem	Cyber-attack (Yes, No)
3	(Mittal et al., 2016)	OSINT (Open–source intelligence) dataset	Analysis of cyber security related discussions		Cyber-attack (Yes, No)
4.	(Khandpur et al., 2017)	Users' status updates and blog posts-based text streams	Extract and encode cyberattacks reported and discussed in social media	Structured query expansion-based retrieval algorithm	Cyber-attack (Yes, No)
5.	(Kumar & Carley, 2016a)	World Bank data www.econstats.com, "Correlates of War" www.coorelatesofwar.org, USNA (http://www.usna.edu/Users/cs/nchamber/nations/index.html)	Analysis of cyber security related discussions	Pearson correlation, Quadratic assignment procedure (QAP)	Cyber-attack (Yes, No)

5.1. Convolutional Layer

It consists of filters (kernels) which slide across the input data. A kernel is a matrix to be convolved with the input data and stride controls how much the filter convolves across the input data. This layer performs the convolution on the input data with the kernel using Equation (2). The output of the convolution is also known as the feature map.

The convolution operation is as follows:

$$y_k = \sum_{n=0}^{N-1} x_n h_{k-n} \tag{2}$$

where x is input data, h is the filter, and N is the number of elements in x. The output vector is y. The subscripts denote the n-th element of the vector.

5.2. Pooling Layer

This layer is also known as the down-sampling layer. The pooling operation reduces the dimension of output neurons from the convolutional layer to reduce the computational intensity and prevent the overfitting. The max-pooling operation is used in this work. Max-pooling operation selects only the maximum value in each feature map and consequently reducing the number of output neurons.

5.3. Fully Connected Layer

This layer has full connection to all the activations in the previous layer. The rectifier linear unit (ReLu) is used in this work as an activation function for the convolutional layers (1, 3, 5, 7, 9, 11, and 12). The activation function is an operation which maps an output to a set of inputs. ReLu function is defined by the Equation (3):

$$\Phi(x) = \max(0,x) \tag{3}$$

The final output decision of the CNN model is based on the weights and biases of the previous layers in the network structure. Hence, the weights and biases of the model are updated with Equation (4) and Equation (5) respectively for each layer:

$$\Delta W_l\left(t+1\right) = -\frac{x\lambda}{r} W_l - \frac{x}{n}\frac{\partial C}{\partial W_l} + m\Delta W_l\left(t\right) \tag{4}$$

$$\Delta B_l\left(t+1\right) = -\frac{x}{n}\frac{\partial C}{\partial B_l} + m\Delta B_l\left(t\right) \tag{5}$$

where *W, B, l, λ*, *x, n, m, t, C* represents the weight, bias, layer number, regularization parameter, learning rate, the total number of training samples, momentum, updating step, and cost function respectively.

The parameters used to train the CNN model are lambda regularization, learning rate, and momentum. These parameters can be tuned according to the dataset in order to achieve optimum performance. The lambda is to prevent overfitting of the data. The learning rate is to control how fast the network learns during training and momentum helps to convergence the data. The parameters lambda, learning rate, momentum is set to 0.04, 0.001, and 0.99, respectively in this work.

5.4. Architecture of The Improved CNN

Figure 1 shows the architecture of the CNN structure with 20808 input sample lengths where the green, blue, and red color signify the kernel size, max-pooling, and fully connected layer respectively. This proposed deep CNN architecture constructed on thirteen layers and includes five convolutional, five max-pooling, and three fully connected layers.

Step 1: The input layer (Layer 0) is convolved using Equation (2) with a kernel of size 6 to produce Layer 1.
Step 2: Then, a max-pooling of size 2 is applied to every feature map (Layer 2).
Step 3: After the max-pooling operation, the number of neurons is reduced.
Step 4: Again, the feature map in Layer 2 is convolved with a kernel of size 5 to produce Layer 3.
Step 5: A max-pooling operation of size 2 is applied to every feature map (Layer 4), reducing the number of neurons.
Step 6: Then, feature map from Layer 4 is convolved with a kernel of size 4 to produce Layer 5.

Step 7: Again, a max-pooling of size 2 is applied to reduce the number of neurons in the output layer (Layer 6).

Step 8: The feature map in Layer 6 is again convolved with a kernel size of 4 to produce the next layer (Layer 7).

Step 9: A max-pooling of size 2 is applied to the feature map (Layer 8).

Step 10: The feature map in Layer 8 is convolved with a kernel of size 4 to produce Layer 9.

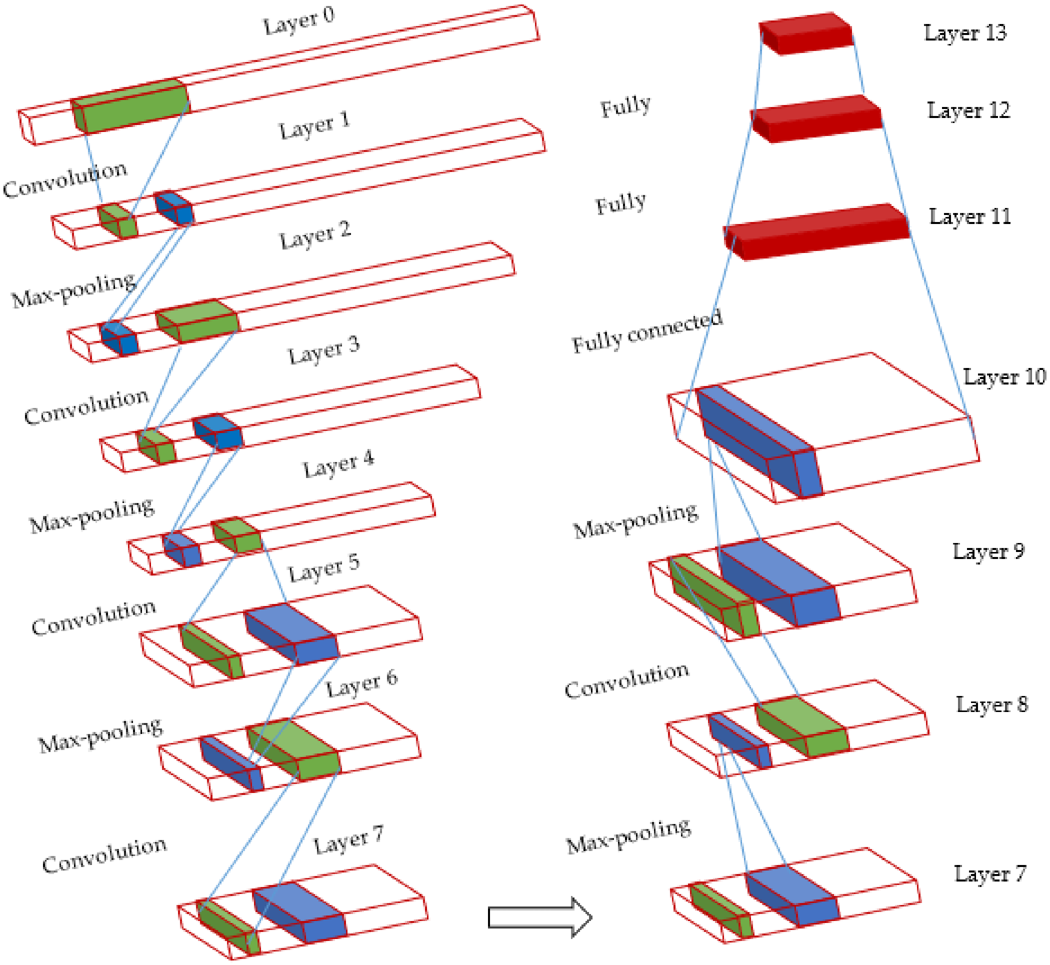
Step 11: Max-pooling of size 2 is applied to every feature map in Layer 10.

Step 12: In Layer 10, the neurons are fully connected to neurons in Layer 11.

Step 13: Layer 11 is fully connected to neurons in Layer 12.

Step 14: Finally, Layer 12 is connected to the last layer (Layer 13) with 2 output neurons. (representing the positive and negative classes).

Figure 1. An Improved CNN



A conventional 1D with a batch size of 10 is employed in this work to train CNN. A batch size is the number of samples used for each training update. The batch size of 10 is chosen in this work.

A total of 140 epochs of training were run in this work. An epoch refers to one iteration of the full training set.

6. AN IMPROVED LSTM MODEL

LSTM (Long Short-Term Memory neural network) is proposed in this study to predict the occurrence probability of the DDoS events based on social media sensor data. LSTM neural network was initially introduced by Hochreiter and Schmidhuber in 1997 (Hochreiter & Schmidhuber, 1997), and the primary objectives of LSTM are to model long-term dependencies and determine the optimal time lag for time series problems. These features are especially desirable for DDoS event prediction in the network domain.

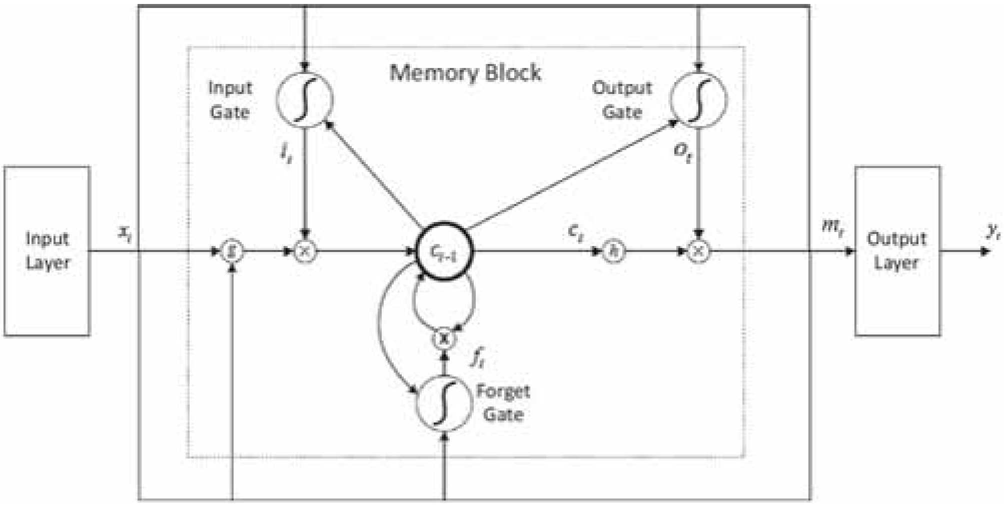
An LSTM is composed of one input layer, one recurrent hidden layer, and one output layer. Different from the traditional neural network, the basic unit of the hidden layer is memory block. The memory block contains memory cells with self-connections memorizing the temporal state, and a pair of adaptive, multiplicative gating units to control information flow in the block. Two additional gates named input gate and output gate respectively control the input and output activations into the block.

The core of memory cell is a recurrently self-connected linear unit called as Constant Error Carousel (CEC). The activation of the CEC represents the cell state. Due to the presence of CEC, multiplicative gates can learn to open and close, and thus LSTM can solve the vanishing error problem by remaining the network error constant.

To prevent the internal cell values growing without binding when processing continual time series that are not previously segmented, a forget gate was added to the memory block. This treatment enables the memory blocks to reset itself once the information flow is out of date, and replaces the CEC weight with the multiplicative forget gate activation.

The above procedure can be visualized in Figure 2.

Figure 2. LTSM neural network architecture



The model input is denoted as x= (x_1,x_2,…,x_t), and the output sequence is denoted as y= (y_1,y_2,… ,y_t), where t is the prediction period. In the context of DDoS events prediction, x can be considered as historical input data (e.g. DDoS incidents), and y is the estimated incident. The objective of LSTM is to predict DDoS event incident in the next time step based on prior information without specifying how many steps should be traced back. To implement this goal, the predicted DDoS event incident time will be iteratively calculated by following Equations (6)-(11):

$$i_t = \delta\left(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i\right) \tag{6}$$

$$f_t = \delta\left(W_{f_x}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f\right) \tag{7}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g\left(W_{cx}x_t + W_{cm}m_{t-1} + b_c\right) \tag{8}$$

$$o_t = \delta\left(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_0\right) \tag{9}$$

$$m_t = o_t \otimes h\left(c_t\right) \tag{10}$$

$$y_t = W_{ym}m_t + b_y \tag{11}$$

where $\otimes$ represents the scalar product of two vectors, and σ(•) denotes the standard logistics sigmoid function defined in Equation (12):

$$\sigma\left(x\right) = \frac{1}{1 + e^{-x}} \tag{12}$$

The memory block is outlined in a box, and consists with an input gate, an output gate and a forget gate, where the outputs of three gates are respectively represented as i_t, o_t, f_t. The activation vectors for each cell and memory block are denoted as c_t and m_t, respectively. The weight matrices W and bias vectors b are utilized to build connections between the input layer, output layer and memory block.

Training LSTM is based on Back-Propagation using the stochastic gradient descent (sgd) method. The common objective function is to minimize the RMSLE (Root Mean Squared Logarithmic Error).

RMSLE is to compare the predictive value with the true value, and is calculated as the square root of the squared bias plus squared standard error:

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(a_i + 1\right) - \log\left(b_i + 1\right)\right)^2} \tag{13}$$

where n is the total number of observations in the testing dataset, a_i is predicted value, and b_i is the actual value. RMSLE is a method to measure the error rate, so smaller RMSLE value indicates more accurate model.

7. EXPERIMENTS

We propose a method for the detection of the DDoS attacks in social media. For this purpose, the proposed method first analyzes and then predicts the sentimental traces in the content, related to DDoS events.

7.1. Experiment Environment

The test process of the model that detects DDoS attacks on social media is provided on the Data Center of Institute of Information Technology of Azerbaijan National Academy of Sciences (AzScienceNet), in the following environment: Ubuntu 16.04.3 LTS amd64 system, 331.2 GB memory, 2933.437 CPU MHz.

The implementation of the method is conducted on the Python and Tensorflow. In this study, for implementation and experiments, Twitter data is used. Twitter, which can be said to be a representative social media site. Here US tweet data is used which is collected by work (Wang & Zhang, 2017).

Various experiments are performed to verify the accuracy of the proposed method. In the used dataset 3048 row of the dataset is positive, and the 17761 row is negative. The prediction of the DDoS attack is proved by using a dictionary with listed attack terms. This dictionary contains the names of some DDoS attacks programs, such as UFONet, Low Orbit Ion Cannon (LOIC), also malicious ransomware programs such as wannacry, wannacrypt, petya, wcry, petrwrap and ather security related words such as hackers, ddos, dos, denial distributed, wikileaks dos, lulzsec, and so on. The proposed method detects these words accurately. In order to find optimal parameters in the proposed model, the neural network is tested at different values of parameters.

The sentimental analysis model is constructed based on improved CNN and LSTM algorithms. The detection accuracy and test results for various metrics of the improved CNN and LSTM models are shown in Table 2.

Table 2. Comparison of the improved CNN and LSTM Models with traditional CNN and LSTM models

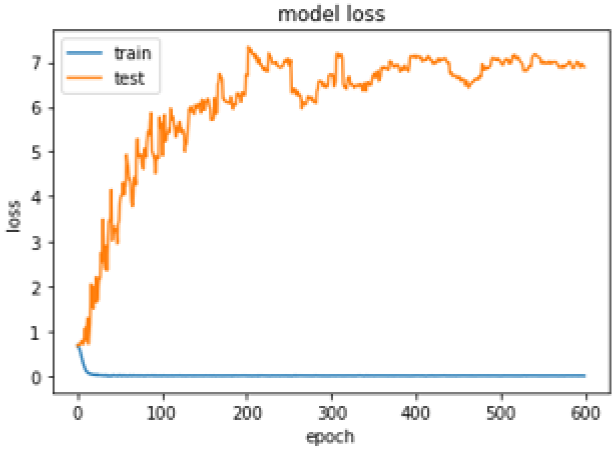
	Recall	**Precision**	**F-Measure**	**Train Loss**	**Train Accuracy**	**Test Loss**	**Test Accuracy**
CNN (proposed)	**0.8455**	0.8923	**0.8683**	0.0919	0.7761	**0.1272**	**0.7744**
CNN (Wang & Zhang, 2017)	0.3469	**0.9297**	0.5053	**0.0126**	**0.9925**	0.8932	0.4026
LSTM (Wang & Zhang, 2017)	0.5364	**0.9154**	0.6764	**0.0025**	**0.9925**	0.2030	0.5487
LSTM (proposed)	**0.7522**	0.8865	**0.8138**	0.1098	0.7090	**0.1354**	**0.6974**

As shown from the results of experiments in Table 2, the improved CNN and LSTM models have produced better results compared to traditional CNN and LSTM models. The experiments are conducted by changing the parameters and the optimal results in LSTM network are obtained in BATCH_SIZE

= 10, EPOCHS = 140, lr = 0.001, momentum = 0.99, decay = 1e-6, nesterov = True values, but in CNN network at BATCH_SIZE = 10, EPOCHS = 500 values and are added to the Table 2. In addition, to improve the results, the kernel regularizer (l2=0.2), BatchNormalization, Weight regularizer (l2 = 0.03) layers are added to the LSTM network, and kernel regularizer (l2 = 0.04) and Dropout = 0.25 layers are added to the CNN network. In each model the optimization function is sgd (stochastic gradient descent), the activation function is relu (rectified linear unit) and the loss function is RMSLE (Root Mean Squared Logarithmic Error).

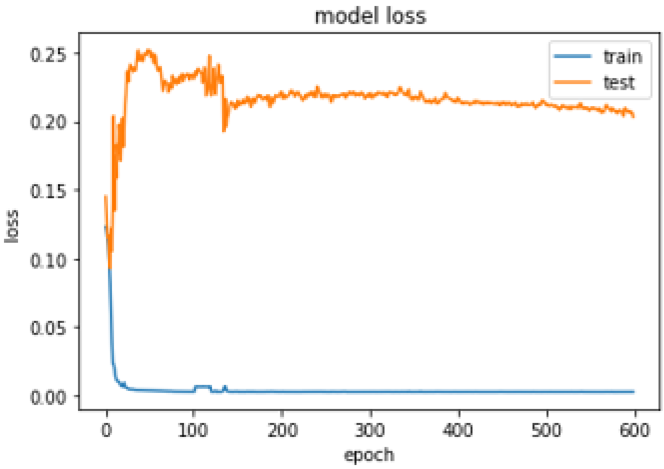
As shown from the Table 2, it is seen that the traditional CNN algorithm has been trained the neural network with little loss and high accuracy (Train Loss = 0.0126, Train acc = 0.9925), but these parameters have been significantly worsened during the testing process. Here the Test loss = 0.8932 and Test acc = 0.4026 show that the neural network has caused a great deal of loss during prediction and almost could not carry out the prediction (Test acc = 0.4026) well. In addition, traditional CNN model has a higher value for precision (0.9297) and low value for recall (0.3469). Better models have to higher values for precision and recall. Weaker models might have high precision (for example, 95%) but low recall (for example, 50%) when it identifies samples of one class largely correct, but it mislabels samples of another class. This landscape can be easily seen from Figure 3.

Figure 3. CNN (Ref. (Wang & Zhang, 2017))



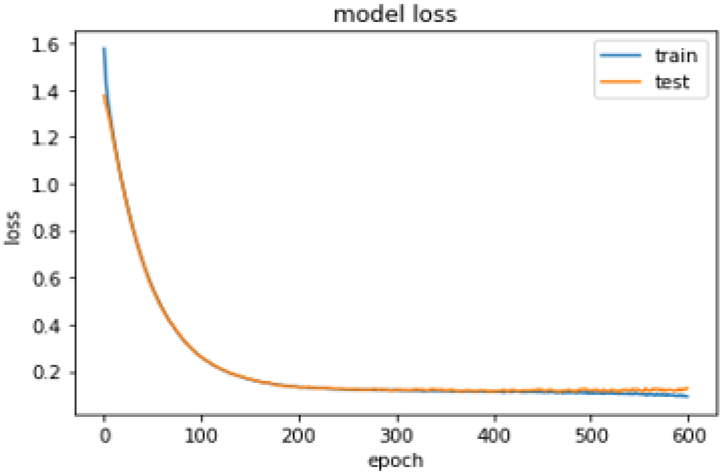
In good prediction models, the dynamics of the test line must be in the direction of the train line and should be as close to it as possible. But here, the opposite landscape is observed. The same landscape can be seen in the traditional LSTM model. Thus, despite the fact that in LSTM model the training loss and training accuracy are Train Loss = 0.0025, Train acc = 0.9925 respectively, but testing loss and testing accuracy of this model were Test loss = 0.2030, Test acc = 0.5487. In addition, traditional LSTM model is also has a higher value for precision (0.9154) and low value for recall (0.5364). This condition is also cannot be considered as a good result. Because here the traditional LSTM model can recognize the 54 percent of points in the dataset, while other points it can't recognize and allows a lot of losses. It can be visualized as follows (Figure 4).

Figure 4. LSTM (Ref. (Wang & Zhang, 2017))



In this work, as mentioned above, by adding various layers to the proposed CNN and LSTM models very high results are obtained in the model. So, the training of the improved CNN model is conducted with low loss and high accuracy, and the loss and accuracy parameters of the model have obtained 0.0919 and 0.7761 values respectively. In this model during testing is also good results are achieved. Thus, the Test loss and Test acc parameters of the model has obtained the 0.1272 and 0.7744 values, respectively. It seems here, that the training and testing of the proposed model are conducted very well. There is not big jumping between the values of the training and testing metrics. As in the training phase, the model is trained with high accuracy, in the testing phase, it predicted the data points very properly. In addition, the precision and recall of the improved CNN model are also has a higher value, e.g. 0.8923 and 0.8455, respectively. The prediction accuracy of the proposed CNN model is visualized in Figure 5.

Figure 5. An improved CNN

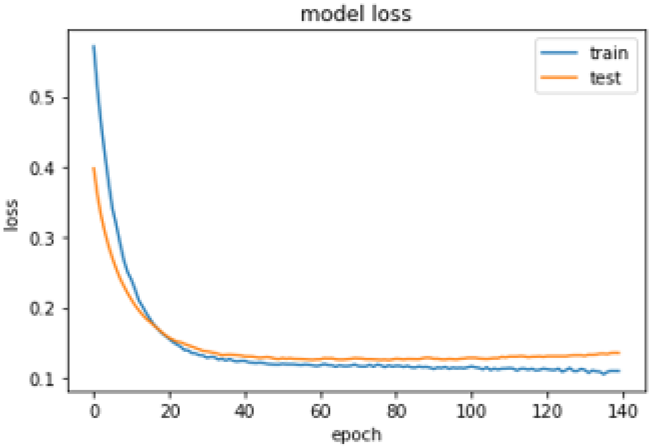


Here the training curve is the almost complete overlap of the test curve. In the proposed LSTM model the good results are also achieved. So, by allowing the low loss in the training of the model, the loss and accuracy parameters of the model have obtained 0.1098 and 0.7090 values, respectively. During the testing, with a slight difference compared to the training phase, also good results are obtained and test loss and test accuracy parameters have obtained the values as Test loss = 0.1354, Test acc = 0.6974. In addition, the precision and recall of the improved LSTM model are also has a higher value, e.g. 0.8865 and 0.7522 respectively.

The predictive accuracy of the improved LSTM model is visualized in Figure 6.

From the visual representation of the LSTM model, it seens that the test curve with the training curve conducted prediction accurately, with very little loss.

Figure 6. An improved LSTM



8. CONCLUSION AND FUTURE WORK

In this study, for the detection of DDoS attacks from the social media text data a deep learning method is employed to automatically identify the two classes of sentiments, positive and negative. To categorize the positive and negative class a 13-layer deep CNN and improved LSTM model is developed. We have performed a comprehensive evaluation of our approach and achieved recall, precision, f-measure, train loss, train accuracy, test loss, and test accuracy of 0.85, 0.89, 0.87, 0.09, 0.78, 0.13 and 0.77, respectively, in detecting DDoS attack related content from social media streams.

Future work is aimed at extending the attack class, as well as modeling the successive dependencies of cyber attacks (from the emergence to reporting). This will help to capture features such as the prevalence of attacks against specific institutions or countries in specific time periods.

ACKNOWLEDGMENT

This work was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – Grant Nº EIF-KETPL-2-2015-1(25)-56/05/1.

REFERENCES

Bleakley, K., & Vert, J. P. (2011). The group fused Lasso for multiple change-point detection. arXiv:1106.4199.

Carl, G., Kesidis, G., Brooks, R. R., & Rai, S. (2006). Denial-of-service attack detection techniques. *IEEE Internet Computing*, *10*(1), 82–89. doi:10.1109/MIC.2006.5

Chambers, N., Bowen, V., Genco, E., Tian, X., Young, E., Harihara, G., & Yang, E. (2015). Identifying political sentiment between nation states with social media. In *Proceedings of the conference on empirical methods in natural language processing*, (pp.65–75). 10.18653/v1/D15-1007

Ebrahimi, M., Suen, C. Y., & Ormandjieva, O. (2016). Detecting predatory conversations in social media by deep Convolutional Neural Networks. (2016). *Digital Investigation*, *18*, 33–49. doi:10.1016/j.diin.2016.07.001

Gandhi, R., Sharma, A., Mahoney, W., Sousan, W., Zhu, Q., & Laplante, P. (2011). Dimensions of cyber-attacks: Cultural, social, economic, and political. *IEEE Technology and Society Magazine*, *30*(1), 28–38. doi:10.1109/MTS.2011.940293

George, T. (2014). The next big cybercrime vector: Social media. *Security Week*. Retrieved from https://www.securityweek.com/next-big-cybercrime-vector-social-media

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Hua, T., Lu, C. T., Ramakrishnan, N., Chen, F., Arredondo, J., Mares, D., & Summers, K. (2013). Analyzing Civil Unrest through Social Media. *Computer*, *46*(12), 80–84. doi:10.1109/MC.2013.442

Imamverdiyev, Y. N., & Abdullayeva, F. J. (2018). Deep learning method for denial of service attack detection based on restricted Boltzmann machine. *Big Data*, *6*(2), 159–169. doi:10.1089/big.2018.0023 PMID:29924649

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (Vol. 1, pp. 151–160).

Jr, A. M., & Menendez, D. (2002). *Cyber forensics: a field manual for collecting, examining, and preserving evidence of computer crimes* (2nd ed.). CRC Press.

Kaur, P., Kumar, M., & Bhandari, A. (2017). A review of detection approaches for distributed denial of service attacks. *Systems Science & Control Engineering*, *5*(1), 301–320. doi:10.1080/21642583.2017.1331768

Khandpur, R. P., Ji, T., Jan, S., Wang, G., Lu, C. T., & Ramakrishnan, N. (2017). Crowdsourcing cybersecurity: Cyber attack detection using social media. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 1049–1057). New York, NY: ACM. 10.1145/3132847.3132866

Kumar, S., & Carley, K. M. (2016a). Approaches to understanding the motivations behind cyber attacks. In *Proceedings of the IEEE Conference on Intelligence and Security Informatics* (pp. 307-309). 10.1109/ISI.2016.7745496

Kumar, S., & Carley, K. M. (2016b). Understanding DDoS cyber-attacks using social media analytics. In *Proceedings of the IEEE Conference on Intelligence and Security Informatics* (pp. 231-236). 10.1109/ISI.2016.7745480

Lippmann, R. P., Weller-Fahy, D. J., Mensch, A. C., Campbell, W. M., Campbell, J. P., Streilein, W. W., & Carter, K. M. (2017). Toward finding malicious cyber discussions in social media. In The AAAI-17 workshop on artificial intelligence for cyber security (pp. 203-209).

Liu, B., & Zhang, L. (2012). *A survey of opinion mining and sentiment analysis. In Mining text data* (pp. 415–463). Springer.

Longueville, B. D., Smith, R. S., & Luraschi, G. (2009). Omg, from here, I can see the flames!: A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the International Workshop on Location based Social Networks* (pp. 73-80). 10.1145/1629890.1629907

Matthews, T. (2014). *Incapsula survey: What DDoS attacks really cost businesses*. Incapsula Inc.

Mezzour, G. (2015). Assessing the global cyber and biological threat [Ph.D. dissertation].

Mirkovic, J., & Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *Computer Communication Review*, *34*(2), 39–53. doi:10.1145/997150.997156

Mittal, S., Das, P. K., Mulwad, V., Joshi, A., & Finin, T. (2016). CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. In *Proceedings of the IEEE/ACM ınternational conference on advances in social networks analysis and mining* (pp. 860-867).

Oh, O., Agrawal, M., & Rao, H. R. (2011). Information control and terrorism: Tracking the mumbai terrorist attack through Twitter. *Information Systems Frontiers*, *13*(1), 33–43. doi:10.100710796-010-9275-8

Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015). Weakly supervised extraction of computer security events from Twitter. In *Proceedings of the 24th international conference on World Wide Web* (pp. 896-905). 10.1145/2736277.2741083

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web* (pp. 851-860). 10.1145/1772690.1772777

Sapienza, A., Bessi, A., Damodaran, S., Shakarian, P., Lerman, K., & Ferrara, E. (2017). Early warnings of cyber threats in online discussions. In *Proceedings of the IEEE international Conference on Data Mining Workshops* (pp. 667-674). 10.1109/ICDMW.2017.94

Shackelford, S. (2009). From nuclear war to net war: Analogizing cyber attacks in international law. *Berkeley Journal of International Law*, *25*(3), 191–251.

Suarez, A. H., Perez, G. S., Medina, K. T., Hernandez, V. M., Meana, H. P., Mercado, J. O., & Sanchez, V. (2018). Social sentiment sensor in Twitter for predicting cyber-attacks using $\ell 1$ regularization. *Sensors (Basel)*, *18*(5), 1–17.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. doi:10.1162/COLI_a_00049

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1079-1088). 10.1145/1753326.1753486

Wang, Z., & Zhang, Y. (2017). DDoS event forecasting using Twitter data. In *Proceedings of the 26th international joint conference on artificial intelligence* (pp. 4151-4157).

Yoo, S. Y., Song, J., & Jeong, O. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*, *105*, 102–111. doi:10.1016/j.eswa.2018.03.055

This research was previously published in the International Journal of Cyber Warfare and Terrorism (IJCWT), 9(1); edited by Graeme Pye and Brett van Niekerk; pages 1-18, copyright year 2019 by IGI Publishing (an imprint of IGI Global).