Method of analyzing computer traffic based on recurrent neural networks

V A Chastikova and V V Sotnikov
Kuban State Technological University, 2, Moskovskaya str., Krasnodar, 350072, Russia
E-mail: chastikova_va@mail.ru

Abstract. The given paper proposes a method of analyzing network traffic based on recurrent neural networks. There overview of perspective approaches for analyzing network traffic in order to detect attacks is provided. The authors investigated the largest and currently the most relevant CICIDS2018 dataset. The methods of dealing with the class imbalance in a dataset by adapting the Focal Loss function to the problem of traffic analysis are considered. There proposed method provides the effective representation of information characteristics of network packets by means of encoder subnetworks. The resulting embeddings are fed at the input of the recurrent LSTM layer. The designed network meta-architecture is potentially effective for the presented dataset as well as for relevant analogues.

1. Introduction
The protection of computer networks is becoming an obligatory task not only for state and private organizations but for ordinary users as well. This is due to the growing number of offenders, as well as the automation of the network attack process. Under such conditions network intrusion detection systems based on signature methods, which are able to detect only already known attack vectors, are becoming increasingly less effective.

The relevant solution to this problem is various methods of detecting network anomalies, based on data analysis algorithms.

The detection of anomalies in network traffic is an important direction in information security, as it is a tool for detecting zero day attacks. For a long time the task of efficiently generating a training dataset has remained unsolved [1]: existing datasets suffered from the similarity of generated traffic and its small volume [2], some of them covered only a limited range of attacks or offered a limited number of training characteristics [1, 3].

Each year companies suffer from greater damage caused by the incidents connected with information security. They refer to the direct embezzlement and infrastructure downtime as well as to the response operations for these incidents. The number of attacks keeps on growing every year and the amount of damage inflicted on companies increases as well.

In the given paper the authors use CSE-CIC-IDS2018 dataset, which is currently one of the largest open datasets [1]. It was generated in an intentionally deployed network consisting of 50 malicious hosts, 420 client hosts and 30 servers, and included 7 different attack vectors: brute force, Heartbleed vulnerability exploitation, botnet attack, denial of service attack, distributed denial of service attack, attack on web applications, as well as inside network attack via backdoor in one of the client hosts.
2. Relevance and analogues
The task of intrusion detection systems (IDS) is to monitor network traffic in order to detect uncharacteristic patterns caused by network attacks. Depending on the set task, network packets are classified binary (into safe for the system and malicious), or divided into several classes corresponding to different attack vectors.

Intrusion detection systems analyze the logs of communication processes in a network. The stored metadata is similar to natural speech: it is a sequence of data, a structure of which follows the protocol-driven rules [4]. This allows suggesting that the proven effective neural network approaches used in natural speech processing tasks can also be useful in the field of network attack detection. Recurrent neural networks, analyzing not only individual network packets but also the context of the communication process as a whole, have proven their effectiveness in analyzing authentication data [5].

3. Data
CSE-CIC-IDS2018 dataset includes traffic from various network services and protocols: HTTPS and HTTP (traffic consists mainly of them), SMTP, POP3, IMAP, SSH and FTP, implementing various attack scenarios:

- Dictionary attacks via Patator utility to obtain SSH and FTP administrator credentials on the server.
- Heartbleed (Heartleech) vulnerability exploitation, which allows getting random access to a protected memory part of a remote machine.
- A botnet attack controlled via the Ares utility.
- Denial of Service Attack: Slowloris, Hulk, GoldenEye and Slowhttptest utilities are used which enable to attack the HTTP and HTTPS protocols by opening a large number of TCP connections up to exhausting the web server pool.
- A distributed denial of service attack using the Low Orbit Ion Cannon (LOIC) utility on the UDP, TCP and HTTP protocols, with the preceding port scanning via PortScan.
- Attacks on web applications: Damn Vulnerable Web App (DVWA) is used as a vulnerable application. DVWA is a software tool that is a sandbox to work out protection means against attacks on web applications — which was used to handle various attacks, such as SQL injections, remote code execution, uploading of random files onto a remote server, XSS.
- Attack on the network from the inside: it is implemented by sending an infected file to the victim’s computer by e-mail. After a successful exploitation of the vulnerability on the victim's computer the attacker deploys a backdoor on it, by means of which he scans the network for vulnerabilities from the inside and exploits them to the extent possible.

The traffic from the dataset mentioned above was processed via the CICFlowMeter-V3 utility to extract from it 80 training characteristics, such as the number of packets per second, the number of packets with a certain TCP flag, the mean square deviation of the packet size during a session, and others [6].

There are two main approaches to representing network packets for analysis by a neural network: in the form of dyad-hours - sets of network packets within two separate hours recorded per hour - and in the form of single packets. Since the aggregation of packets into dyads proved to be inefficient when using deep LSTM networks [4], it was decided to use network packets as atomic elements of network interaction.

Due to the fact that a sufficiently large amount of information fields of each network packet is encoded via a unitary code, dataset is a sparse matrix. To reduce the dimension of the processed data, some columns of this matrix are displayed as compact vector embeddings by means of a neural network itself, which are further concatenated with the remaining information characteristics.

4. Network architecture
There are many mathematical methods for mapping high-dimensional data into low-dimensional data while keeping important structures and patterns. Neural network embeddings are mainly used in natural
speech processing, based on the distributive hypothesis that words occurring in similar contexts have similar meanings; therefore, they have corresponding similar vectors [7].

Figure 1. Computational graph of the resulting meta-architecture

Due to the problem specifics of recognizing abnormal traffic and sparseness of incoming data, embedding has been implemented as a subnet within the main neural network and has been trained simultaneously with the other layers. After the subnetwork training many parameters of network packets, encoded unitarily, are packed into a compact vector space, next they are concatenated with the remaining information attributes and fed to the input of the LSTM layers. The resulting meta-architecture is shown in Figure 1.

LSTM networks represent one of the most efficient current neural network models for processing ordered data sets (along with GRU networks) [7, 8]. These networks, unlike traditional recurrent ones, allow accumulating information for a long time, avoiding the problem of exploding and decaying
gradients by storing their state in one hidden neurons and controlling this state by means of other hidden neurons [9].

5. Network training
A typical problem for all datasets with abnormal network traffic is the extreme degree of class imbalance: benign traffic can make up to 80% of the total dataset volume, while individual attack vectors can be represented by less than one percent of network packets [4]. Easily classified packets contribute the most to the loss function gradient, unfavorably affecting the learning process.

A traditional method used in similar tasks [8] is to introduce the weight \( \alpha_t \) into the traditional formula (1) for the loss function of cross-entropy:

\[
CE(P_t) = - \log(P_t)
\]

(1)

\[
CE_w(P_t) = - \alpha_t \cdot \log(P_t) \quad (\alpha_t \in [0, 1])
\]

(2)

where \( t \) is a class of an object, \( \alpha_t \) – a weight coefficient for class \( t \), \( P_t \) – is the predicted probability of a class.

To cope with this problem it was proposed to use the loss function Focal Loss [11], initially developed by Facebook AI Research for using in complex pattern recognition tasks, but also successfully tested in other classification tasks under conditions of class imbalance [12]. Unlike the method described above, it applies weights not to classes of network attacks as a whole, but to specific packets, the classification of which caused the greatest difficulties for the model.

Focal Loss is based on the calculation of cross-entropy between the predicted and true distribution, with the standard formula (1) being expanded by a multiplier:

\[
FL(P_t) = - (1 - P_t)^\gamma \cdot \log(P_t)
\]

(3)

where \( \gamma \) – is a hyperparameter of the function.

With \( \gamma > 0 \) the contribution to the loss function of easily classified (mainly benign) packets decreases, meanwhile the learning process of the neural network focuses on those few packets, classification of which causes the greatest difficulty.

Aimed at optimizing the neural network the Adam algorithm (“adaptive moments”) was used, which is a first-order gradient optimization algorithm with the adaptive learning rate. It combines the advantages of such optimization methods as RMSProp and momentum, while having greater computational stability, especially early in the course of training, which is achieved by means of correcting bias of an estimate of the first and second order moments [13].

6. Conclusion
The researched CICIDS2018 dataset is currently the largest and of the highest quality dataset with network traffic. Initially developed for pattern recognition tasks, the Focal Loss function allows dealing with the class imbalance, which is characteristic for all data sets with network traffic. A method for compact vector mapping of information characteristics of network packets using subnetworks-encoders has been proposed. The transformed vectors are fed to the input of the recurrent LSTM layer. The developed network meta-architecture can be used not only for the described dataset, but for relevant analogues as well.

The effective neural network architecture, consisting of built-in embedding-subnets, layers of long short-term memory and one-dimensional convolutions, as well as the adapted loss function – Focal Loss, was empirically developed. The resulting neural network was trained by means of the dataset described above, using the Adam algorithm.

References
[1] Sharafaldin I, Lashkari A H and Ghorbani A A 2018 Proc. of the 4th Int. Conf. on Information Systems Security and Privacy 1 108–16
[2] Revathi S and Malathi A 2013 *Int. J. of Engineering Research & Technology* 2(12) 1848–53

[3] Tharmini J and Shahrazad Z 2017 *IEEE 26th Int. Symposium on Industrial Electronics* pp 1881–6

[4] Radford B J, Apolonio L M, Trias A J and Simpson J A 2018 *Network Traffic Anomaly Detection Using Recurrent Neural Networks* Preprint abs/1803.10769

[5] Tuor A, Kaplan S, Hutchinson B, Nichols N and Robinson S 2017 *The AAAI-17 Workshop on Artificial Intelligence for Cyber Security* pp 224–31

[6] Lashkari A H, Gil G D, Mamun M and Ghorbani A A 2017 *Proceedings of the 3th Int. Conf. on Information Systems Security and Privacy* 253–62

[7] Mikolov T, Sutskever I, Chen K, Corrado G and Dean J 2013 *Advances in Neural Information Processing Systems* 2 3111–9

[8] Sutskever I, Vinyals O and Le V Q 2014 *Advances in Neural Information Processing Systems* 4 3104–12

[9] Vinyals O, Toshev A, Bengio S and Erhan D 2015 *IEEE Conf. on Computer Vision and Pattern Recognition* pp 3156–64

[10] Hochreiter S and Schmidhuber J 1997 *Neural Computation* 9 1735–80

[11] Lin T, Goyal P, Girshick R, He K and Dollár P 2017 *IEEE Int. Conf. on Computer Vision* pp 2999–3007

[12] Kumar S, Haresh S, Baloch B, Rehman A and Syed T 2017 *Focused-Anchors Loss for imbalanced classification* Preprint abs/1708.02002

[13] Kingma D P and Ba J 2015 *3rd Int. Conf. for Learning Representations* (Preprint abs/1412.6980)