Fuzzy Clustering Based Approach to Network Traffic Classification and Anomaly Detection

Julija Asmuss, Gunars Lauks
Institute of Telecommunication
Riga Technical University
Riga, Latvia
e-mail: julija.asmuss@rtu.lv, gunars.lauks@rtu.lv

Abstract— **In this work, we develop network traffic classification and anomaly detection methods based on traffic time series analysis using fuzzy clustering. We compare four fuzzy clustering techniques using different dimensionality reduction methods and validity indices to work out an effective anomaly detection algorithm. The effectiveness of the proposed classification system is evaluated on traffic data with and without traffic attack components.**

Keywords- fuzzy clustering; fuzzy transform; traffic classification; anomaly detection.

I. INTRODUCTION

The ability to classify and identify network traffic is the main area of interest for many network operation and research topics such as traffic engineering, monitoring, pricing, security, anomaly detecting. Anomalous traffic or unwanted traffic definition is still very fuzzy and immensely varies among networks. But it is clear that anomalies (such as Distributed Denial-of-Service (DDoS) attacks [1], for example) may cause significant variances in a network traffic level, and as a result, legitimate user requests can not get through the network. Our work mainly focuses on flood attacks. The most common DDoS flood attacks target the computer networks bandwidth or connectivity. In this context, traffic volume analysis is considered to be a sensitive tool for anomaly detection.

Many monitoring schemes against DDoS attacks have been reported in the literature (see, e.g., [2][3][7][9][12][13][16]), but only a few of them have been applied in a real network environment. In the context of balance between computation speed and classification success, the task of network traffic anomaly detection is still very important [1].

Our research is devoted to anomaly detection methods based on traffic time series analysis using clustering technique. We follow the idea, which is based on anomalous traffic profile deviation from normal traffic profile, defined empirically on the basis of previously collected information on the properties of normal traffic conditions. The traditional approach to clustering can not be effectively used in this case due to the dynamic behaviour of network traffic in its development over time. This dynamic behaviour should be taken into account when solving traffic classification problems. Traffic time series may belong to one cluster during a certain period; afterwards, its profile may be closer to another cluster. This switch from one state to another can be naturally modelled using a fuzzy approach. We show that fuzzy logic based techniques allow us to deal effectively with the vague and imprecise boundaries between normal traffic and different levels of attacks.

The remainder of the paper is structured as follows. Section II introduces objectives and tasks of this research. Section III contains traffic data representation tools. Sections IV and V are devoted to the clustering and classification stages, respectively. Section VI contains the description of experiments and results.

II. OUTLINE OF RESEARCH OBJECTIVES

Our aim in this research is to use the advantages of the fuzzy logic based approach for analysis of network traffic dynamics and to suggest traffic classification and anomaly detection mechanisms based on fuzzy transforms, fuzzy clustering and classification with good computation speed and classification success characteristics. The tasks of the research follow directly from its objectives:

- To develop a fuzzy transform technique for representation of network traffic and to investigate its advantages in traffic data compression;
- To evaluate the performance of different fuzzy clustering methods for solving the traffic classification problem and to identify the best one;
- To develop the mechanism of traffic classification and anomaly detection using traffic fuzzy clusters and self-similarity characteristics;
- To evaluate the performance of the suggested method and to compare it with the existing solutions.

III. TRAFFIC DATA PRE-PROCESSING

We work with time series that represent aggregated traffic and we compare two dimensionality reduction methods. Usually, the dimension of time series representation is reduced using Piecewise Aggregate Approximation (PAA) [17]. In this research, we also apply time series representation based on Fuzzy Transform (for short, F-transform) introduced in [8].

Generally, F-transform depends on a chosen fuzzy partition, which consists of fuzzy sets given by membership functions. We apply uniform fuzzy partition with the triangular form of membership functions and use discrete formulas for F-transform components.

The idea of using fuzzy transforms in analysis of time series is not new (see, e.g., [4][5]). Our research develops the F-transform technique as a special tool for traffic data aggregation and investigates its role in reduction of traffic classification computational resources.

IV. Fuzzy Clustering Technique

We consider four fuzzy clustering algorithms and evaluate their performance for our purposes: Fuzzy C-Means (FCM); Possibilistic C-Means (PCM); Possibilistic Clustering Algorithm (PCA); Unsupervised Possibilistic Fuzzy Clustering (UPFC) [6][11][14][15]. For each of them we consider also Gustafson-Kessel modification GK (see, e.g., [6]).

We apply four validity indices for determining the number of clusters: modified partition coefficient, Fukuyama and Sugeno index, Xie and Beni index, separation and compactness index (see, e.g., [10]). Taking into account that the result obtained by using each index can be interpreted as evaluation done by an expert, we apply the technique of aggregation of expert opinions.

Thus, at the clustering stage we obtain fuzzy clusters and cluster centroids given by their F-transform components or PAA components, correspondingly.

V. Fuzzy Classification and Anomaly Detection

For the traffic classification merit we use the prototypes obtained at the previous stage. Decision making on classification of a new traffic time series is done in the following way. We suppose that we have new infinite time series, therefore the algorithm runs infinite time too. At each moment in time, the classification is done considering a finite number of time series components and their F-transform (or PAA) components, correspondingly.

In each next step, we start with the computation of F-transform (or PAA) components (as compared with the previous step, the first component is removed and one new component is added to the end). Then, we compute the membership degrees with respect to all clusters of normal (in some cases, of anomalous also) traffic. Next, we evaluate self-similarity parameter changing rate. Finally, decision making on the risk of anomalies is done on the basis of the above mentioned membership degrees and Hurst parameter changing rate by using the fuzzy rule based technique.

VI. Experiments and Results

When studying a traffic classification technique with real traces, it is important to have a baseline for traffic classification that will be used as a reference. Because it is very difficult to obtain a dataset that is representative of real network activities and contains both normal and anomalous traffic, attack traffic for numerical experiments was generated and added as additional component of traffic data.

To evaluate the effectiveness of the proposed technique, we consider all major steps:

- Traffic data pre-processing, the main merit of which is to reduce the amount of traffic data and to allow a more effective use of data analysis techniques, both in time and space;
- Extracting the relationship between traffic data by using fuzzy clustering methods to characterize patterns, which identify normal network traffic;
- Detecting traffic anomalies by using fuzzy rule based prototypical classification methods.

The algorithm consists of two stages: fuzzy clustering, which can be executed only once, and real-time classification. When evaluating the computation time per classification, the clustering stage is not taken into account.

The results obtained by comparing different techniques show that the best compromise between computation speed and classification success is achieved using F-transforms for traffic time series representation and applying UPFC-GK clustering algorithm. The detection rate (DR) in our experiments with this technique was greater than 99%; the false alarm rate (FAR) was about 1% in the worst cases. A comparison with methods based on statistical analysis (D-WARD, DCD, T-test model) is done using DR and FAR evaluation results shown in [16]. For the exact comparison and performance evaluation, it is necessary to obtain results for real traffic classification in the same testing conditions.

Acknowledgment

This work has been supported by the European Social Fund within the project 2013/0024/1DP/1.1.1.2.0/13/APIA/ VIAA/045 „Applications of mathematical structures based on fuzzy logic principles in the development of telecommunication network design and resource control technologies”.

References

[1] Computer Emergency Response Team (CERT-EU), “DDoS overview and incident response guide,” 2014. [Online]. Available from: http://cert.europa.eu/static/WhitePapers/ [retrieved: May, 2015]

[2] K. Lee, J. Kim, K. H. Kwon, J. Han, and S. Kim, “DDoS attack detection method using cluster analysis,” Expert Systems with Applications, vol. 34, 2008, pp. 1659–1665.

[3] M. Li, “Change trend of averaged Hurst parameter of traffic under DDOS flood attacks,” Computers & Security, vol. 25, 2006, pp. 213–220.

[4] V. Novák, I. Perfilieva, M. Holčapek, and V. Kreinovich, “Filtering out high frequencies in time series using F-transform,” Information Sciences, vol. 274, 2014, pp. 192–209.

[5] V. Novák, M. Štepnička, V. Dvorák, I. Perfilieva, V. Pavliska, and I. Vavrıčková, “Analysis of seasonal time series using fuzzy approach.,” International Journal of General Systems, vol. 39, 2010, pp. 305–328.

[6] J. V. de Oliveira and W. Pedrycz, “Advances in fuzzy clustering and its applications,” John Wiley and Sons, 2007.

[7] T. T. Oo and T. Phyu, “A statistical approach to classify and identify DDoS attacks using UCLA Dataset,” Int. J. of Advanced Research in Computer Engineering & Technology, vol. 2, No. 5, 2013, pp. 1766–1770.

[8] I. Perfilieva, “Fuzzy transforms: Theory and applications,” Fuzzy Sets and Systems, vol. 157, 2006, pp. 993–1004.

[9] S. N. Shiaeles, V. Katos, A. S. Karakos, and B. K. Papadopoulos, “Real time DDoS detection using fuzzy

estimators", Computers & Security, vol. 31, 2012, pp. 782–790.

[10] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," Fuzzy Sets and Systems, vol. 158, 2007, pp. 2095 –2117.

[11] X. Wu, B. Wu, J. Sun, and H. Fu, "Unsupervised possibilistic fuzzy clustering," Journal of Information & Computational Science, vol. 7(5), 2010, pp. 1075–1080.

[12] Z. Xia, S. Lu, J. Li, and J. Tang, "Enhancing DDoS flood attack detection via intelligent fuzzy logic," Informatica, vol. 34, 2010, pp. 497–507.

[13] Z. Xiong, Y. Wang, and X. F. Wang, "Distributed collaborative DDoS detection method based on traffic classification features", Proc. of International Conference on Computer Science and Electronics Engineering ICCSEE 2013, Atlantic Press, Paris, 2013, pp. 93–96.

[14] M.-S. Yang and K.-L. Wub, "Unsupervised possibilistic clustering," Pattern Recpgnition, vol. 39, 2006, pp. 5–21.

[15] R. J. Almeida and J. M. C. Sousa, "Comparison of fuzzy clustering algorithms for classification," Proc. of International Symposium on Evolving Fuzzy Systems EFS'06, Lake District, United Kingdom, 2006, pp. 112–117.

[16] M. H. Bhuyan, H. J. Kashyap, D. K. Bhattacharyya, and J. K. Kalita, "Detecting distributed denial of service attacks: methods, tools and future directions," Computer Journal, vol. 57, 2014, 537-556.

[17] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," Knowledge and Information Systems," vol. 3, 2001, pp. 263–286.