



ISSN: 2350-0328

**International Journal of Advanced Research in Science,
Engineering and Technology**

Vol. 6, Issue 10, October 2019

Building a k-means Algorithm to Reduce False Positives When Determining a Network Attacks

Gulomov Sherzod Rajabayevich, Kadirov Mirhusan Mirpulatovich, Karimova Nazimakhon Aybekovna, Raximjonov Umidjon Shuxratjonovich

Assistant professor, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

Assistant professor, Department of Information Technologies, Tashkent State Technical University, Tashkent, Uzbekistan.

Senior Lecturer, Department of Information Technologies, Tashkent State Technical University, Tashkent, Uzbekistan.

Student, Tashkent State Technical University, Tashkent, Uzbekistan.

ABSTRACT: The article discusses the methodology for detecting network attacks based on the method of inductive prediction of states; an algorithm for detecting and identifying network attacks is proposed, which allows one to perform not only an exhaustive search for the classification features of network attacks, but to limit itself to a shortened search.

KEYWORDS: algorithm, networks attacks, traffic filtering, protecting information, filtering methods

I. INTRODUCTION

Of particular importance in the world is the improvement of effective methods and means of information protection, the organization of controlled information applications, and the development of models and algorithms for detecting and recognizing network attacks in computer networks [1]. In this regard, in research works, special attention is paid to the following aspects: development of a mathematical model for recognizing network attacks in real time; improving the methodology and algorithm for detecting network attacks based on the inductive state prediction method; development of software packages for detecting network attacks in computer networks.

II. BUILDING THE K-MEANS ALGORITHM

For the primary separation of mixed traffic into reliable and malicious, it is optimal to use the clustering algorithm k -means.

This algorithm allows clustering with a known number of clusters [2]. The algorithm has an acceptable accuracy necessary for primary separation, and a higher speed of operation compared to other algorithms. The essence of the algorithm is to distinguish two clusters and calculate their centers of mass; at subsequent iterations, the clusters are corrected and the centers of mass are recalculated.

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (1)$$

Where

k – is the number of clusters;

S – derived clusters;

$i = 1, 2, \dots, k$;

μ_i – centers of mass of vectors $x_j \in S_i$.

As a result of the algorithm, the mixed traffic will be divided into two clusters corresponding to reliable and malicious traffic.

Thus, at this stage, only three traffic groups are available for analysis and processing:

The corresponding trustworthy traffic preceding the start of the attack is – T .

Corresponding, trustworthy traffic, isolated from mixed traffic – T^* .
Corresponding to malicious traffic, isolated from mixed traffic – H .

III. SUCCESS CRITERIA, CORRECTION OF THE RESULTING CLUSTERS

To evaluate the clustering efficiency, the stationary probability equation is considered:

$$\begin{aligned}
 p_0\lambda &= p_1\mu, \\
 (\lambda + i\mu)p_i &= \lambda p_{i-1} + (i + 1)\mu p_{i+1}, i = 1 \dots K - 2, \\
 (\lambda + (K - 1)\mu)p_{K-1} &= \lambda p_{K-2} + KN\mu p_K, \\
 (\lambda^* + K\tilde{N}\mu)p_K &= \lambda p_{K-1} + KN\mu p_{K+1}, \\
 (\lambda^* + i\mu^*)p_i &= \lambda^* p_{i-1} + (i + 1)\mu^* p_{i+1}, i = K + 1 \dots N - 1
 \end{aligned}
 \tag{2}$$

where
 λ – load intensity;
 λ_L –load intensity created by legal users;
 S – malicious traffic intensity;
 μ – request queue release rate;
 μ^* –rate of request queue when filter is activated;
 E_1, E_2 –errors of the first and second kind;
 K – filter activation boundary;
 N – request queue size;
 $\lambda = \lambda_L + S$ –attack load;
 $\lambda^* = \lambda_L + S(1 - E_2)$ – load with activated filter.

Since query validation reduces server performance, in some cases, the intensity of releasing the request queue can be estimated as $Z\mu$

where
 Z is the deceleration coefficient.
When rationing

$$\sum_{i=1}^N p_i = 1,$$

we get the probability of blocking the request

$$P_{BLK} = \frac{\frac{1}{N!} \left(\frac{\lambda}{\mu}\right)^{K-1} \frac{\lambda}{\mu^*} \left(\frac{\lambda^*}{\mu^*}\right)^{N-K}}{\sum_{i=0}^{K-1} \frac{\left(\frac{\lambda}{\mu}\right)^i}{i!} + \sum_{i=K}^N \frac{1}{i!} \left(\frac{\lambda}{\mu}\right)^{K-1} \frac{\lambda}{\mu^*} \left(\frac{\lambda^*}{\mu^*}\right)^{i-K}}
 \tag{3}$$

Thus, the separation efficiency of malicious and trusted traffic can be estimated as

$$R = (1 - E_1)(1 - p_B)
 \tag{4}$$

At the next step of the algorithm, the correction of the obtained samples is carried out taking into account these criteria:

Dimension criterion for the resulting clusters. If in the current period related to the attack, the number of requests is n , and in similar seasonal periods related to reliable traffic is m , then the number of malicious requests will be approximately $n-m$. The same is true for various properties of network activity.

Criteria for the similarity of reliable samples. The maximum similarity of a trustworthy sample preceding the start of an attack with a trustworthy sample extracted from mixed traffic.

The criterion for the correspondence of the centers of mass. The center of mass of a trustworthy sample extracted from mixed traffic should correspond to a similar seasonal period of trustworthy traffic preceding the start of the attack.

For further clarification, we can calculate the probability of each element belonging to its class. Elements with the least probability are transferred to opposite groups, taking into account the criterion for the dimension of groups [3]. To calculate the similarity of trustworthy clusters and then to classify new incoming requests, you can use the “naive Bayesian classifier” . As a probabilistic model for the classifier, we use the conditional probability over a dependent

variable of class C with a small number of results or classes, depending on several variables F_1, \dots, F_n . Using Bayes' theorem, we write:

$$p(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (5)$$

The conditional distribution over the class variable C can be expressed as follows:

$$p(C|F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^n p(F_i|C) \quad (6)$$

Thus, to classify traffic into two classes, we have:

$$p(T|D) = \frac{P(T)}{P(D)} \prod_{i=1}^n P(w_i|T) \text{ --for a trusted user class;}$$

$$p(H|D) = \frac{P(H)}{P(D)} \prod_{i=1}^n P(w_i|H) \text{ --for a class of untrustworthy users.}$$

At the end of this step, the elements from the set T^* , assigned to the malicious traffic group are interchanged with the elements of the set H .

This step is repeated until all elements of the set T are marked as trustworthy, or until the algorithm reaches the threshold value of iterations. The obtained samples corresponding to reliable and malicious traffic, as well as the mechanism for keeping them up to date allow using them with various classifiers.

IV. FLOWCHARTS OF THE ALGORITHM FOR DETERMINING NETWORK ATTACKS AND THE ALLOCATION OF MALICIOUS TRAFFIC

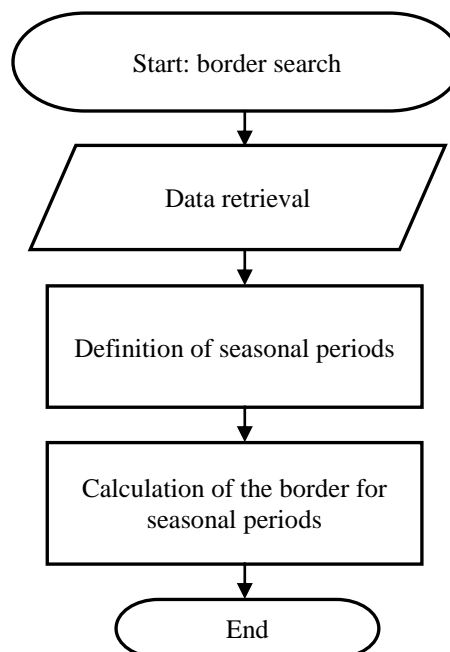
In Fig. 1. The block diagrams of the algorithm for determining network attacks and the allocation of malicious traffic are shown.

The first flowchart explains the algorithm for allocating malicious traffic; the second and third algorithm for determining the start of an attack.

At the first step, subroutines are called to identify seasonal periods, calculate for them the permissible limit of the number of requests, and determine the beginning of the attack.

In the event of an attack, the algorithm splits the mixed traffic into two clusters, one containing malicious requests, the other trustworthy requests. These clusters are being specified. New queries are analyzed for belonging to a particular cluster and, by the result, are added to the corresponding cluster.

1)



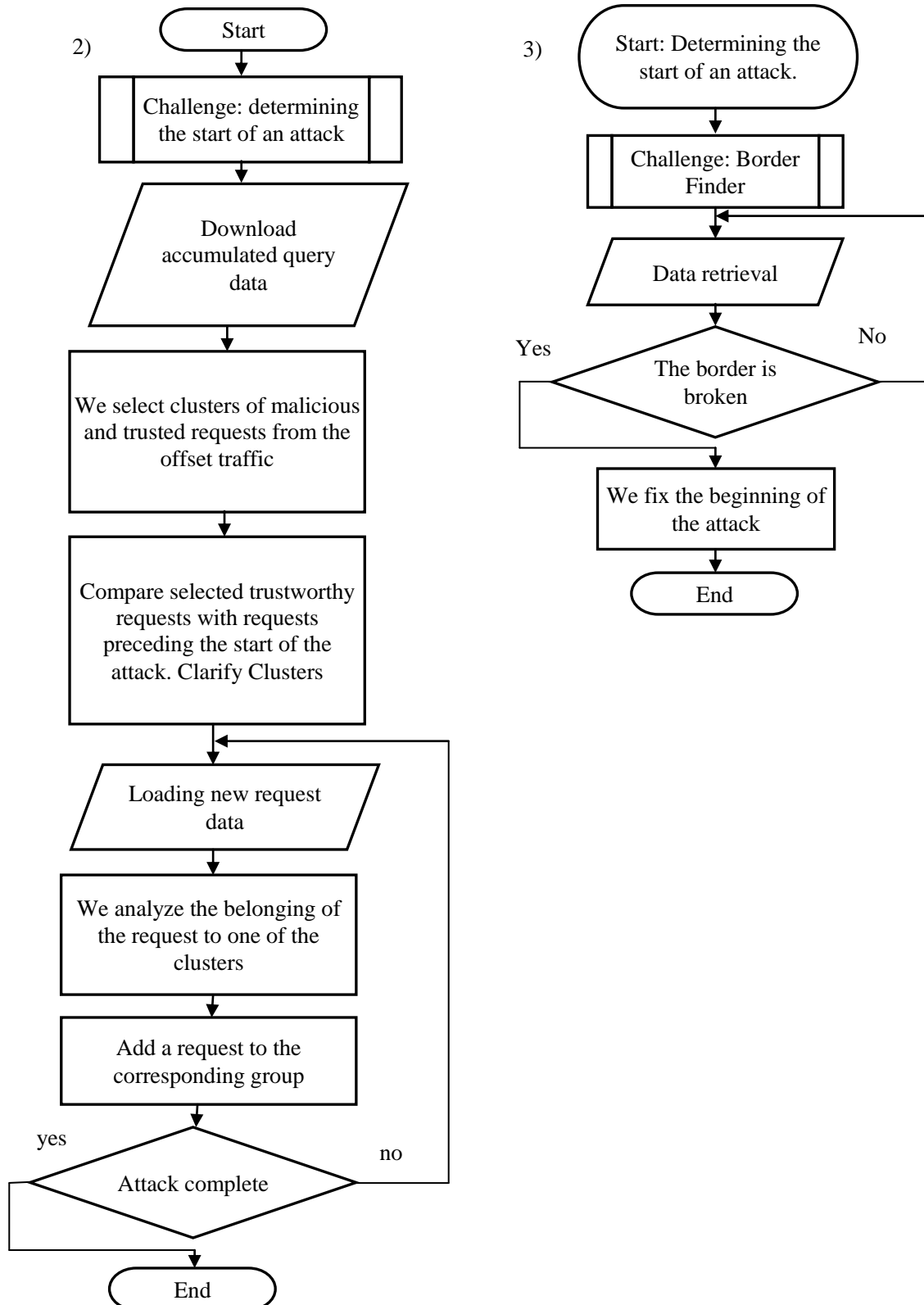
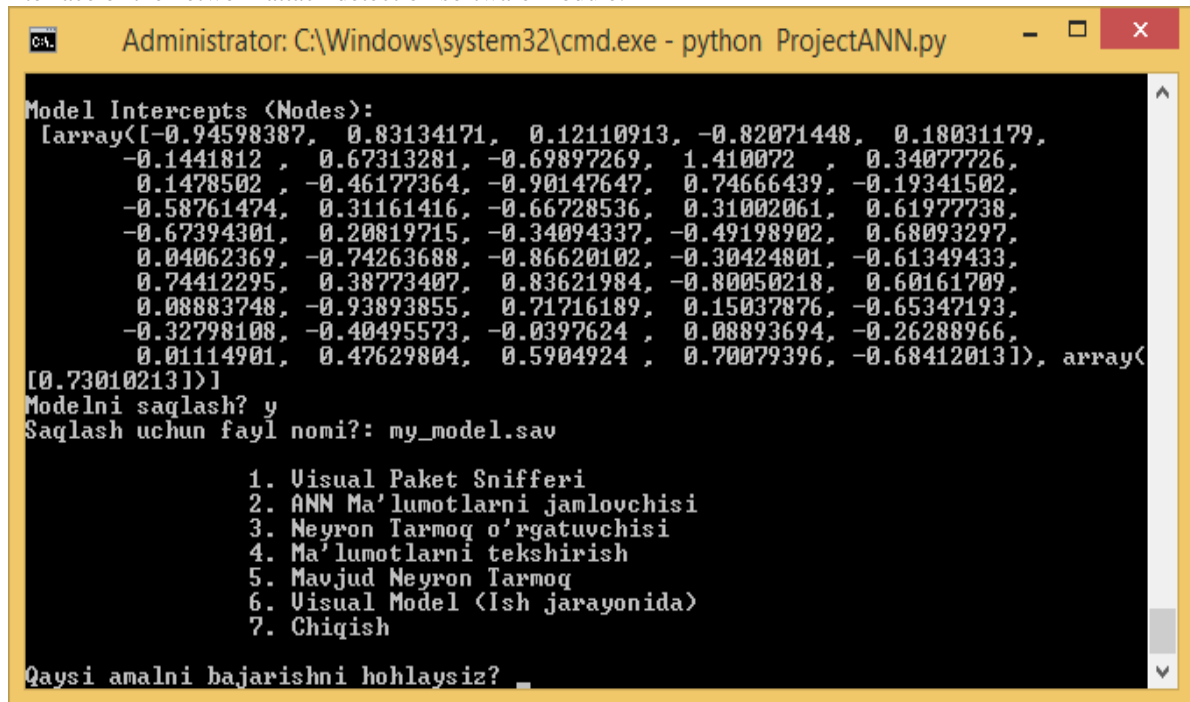


Fig. 1. Flowcharts of the algorithm for determining network attacks and the allocation of malicious traffic

V. RESULTS

Based on the proposed algorithm, a software module is developed.

The developed software module works both in the Windows operating system and in Linux. Figure 2 shows the user interface of the network attack detection software module.



```
Administrator: C:\Windows\system32\cmd.exe - python ProjectANN.py

Model Intercepts (Nodes):
array([-0.94598387,  0.83134171,  0.12110913, -0.82071448,  0.18031179,
        -0.1441812 ,  0.67313281, -0.69897269,  1.410072 ,  0.34077726,
         0.1478502 , -0.46177364, -0.90147647,  0.74666439, -0.19341502,
        -0.58761474,  0.31161416, -0.66728536,  0.31002061,  0.61977738,
        -0.67394301,  0.20819715, -0.34094337, -0.49198902,  0.68093297,
         0.04062369, -0.74263688, -0.86620102, -0.30424801, -0.61349433,
         0.74412295,  0.38773407,  0.83621984, -0.80050218,  0.60161709,
         0.08883748, -0.93893855,  0.71716189,  0.15037876, -0.65347193,
        -0.32798108, -0.40495573, -0.0397624 ,  0.08893694, -0.26288966,
         0.01114901,  0.47629804,  0.5904924 ,  0.70079396, -0.68412013]), array(
[0.73010213])
Modelni saqlash? y
Saqlash uchun fayl nomi?: my_model.sav

1. Visual Paket Snifferi
2. ANN Ma'lumotlarni jamlovchisi
3. Neyron Tarmoq o'rgatuvchisi
4. Ma'lumotlarni tekshirish
5. Mavjud Neyron Tarmoq
6. Visual Model (Ish jarayonida)
7. Chiqish

Qaysi amalni bajarishni hohlaysiz? _
```

Figure 2. Creating a network attack detection model from a dataset

VI. CONCLUSION AND FUTURE WORK

Block diagrams of the k-means algorithm have been built to reduce false positives when determining a network attack and isolating malicious traffic that allows clustering with a known number of clusters. The developed attack detection software module allows you to identify attacks spaced in time, as well as predict future events of operating systems.

It was established that detection of the probability of skipping suspicious packets allows optimizing the filtering of network packets, which ensures a high level of network security.

This work on the subject of the grant EOT-Arex-2018-168 "Improving methods and means of detecting attacks in computer networks".

REFERENCES

- [1] M.M. Karimov, Sh.R. Gulomov, B.K. Yusupov, "Approach development accelerate of process special traffic filtering", Journal of Computer and Communications, vol. 3, no. 9, pp. 68-82, September 2015.
- [2] Erman J., Arlitt M., Mahanti A. Traffic classification using clustering algorithms // Proceedings of the 2006 SIGCOMM workshop on Mining network data. – ACM, 2006. – C. 281-286.
- [3] O.S. Ternovoy, A.S. Shatoxin, Metodiki sredstvaviyavleniya i protivodeystviyaugrozam informatsionnoy bezopasnosti v konteksteregionalnix web-resursov // Regionalnyye aspektitexnicheskoyipravovoyzashitiinformatsii. – Barnaul: Izdatelstvo AltGU, 2013.
- [4] Gulomov Sherzod Rajaboevich, Nasrullayev Nurbek Bakhtiyorovich. Method for security monitoring and special filtering traffic mode in info communication systems // 2016 International Conference on Information Science and Communications Technologies (ICISCT). Tashkent University of Information Technologies. Tashkent, Uzbekistan. Applications, Trends and Opportunities 2nd, 3rd and 4th of November 2016.



ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Issue 10 , October 2019

AUTHOR'S BIOGRAPHY



Gulomov Sherzod Rajaboevich. Assistant professor. was born February 26, 1983 year in Shakhrisabz city, Republic of Uzbekistan. In 2009 graduated «Information technology» faculty of Tashkent University of Information Technologies. Has more than 120 published scientific works in the form of articles, journals, theses and tutorials. Currently works of the department «Information Security» in Tashkent University of Information Technologies.



Kadirov Mirhusan Mirpulatovich. Assistant professor. was born May 22, 1985 year in Tashkent city, Republic of Uzbekistan. Has more than 90 published scientific works in the form of articles, journals, theses and tutorials. Currently works at the department of “Information technologies” in Tashkent State Technical University.



Karimova Nazimakhon Aybekovna, Senior Lecturer. was born 04.01.1971 year in Tashkent city, Republic of Uzbekistan. Has more than 20 published scientific works in the form of articles, journals, theses and tutorials. Currently works at the department of “Information technologies” in Tashkent State Technical University.



Raximjonov Umidjon Shuxratjonovich, student. was born 30.01.1996 year in Tashkent region, Republic of Uzbekistan. Student of group 6-18 of the faculty of mining and metallurgy. Study in geodesy cartography and cadastre, Tashkent State Technical University.