I. J. Computer Network and Information Security, **2020, 6, 1-13**
Published Online December 2020 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2020.06.01

Statistical Techniques for Detecting Cyberattacks on Computer Networks Based on an Analysis of Abnormal Traffic Behavior

[1]Zhengbing Hu, [1,2]Roman Odarchenko, [1,2]Sergiy Gnatyuk, [1]Maksym Zaliskyi, [1]Anastasia Chaplits, [3]Sergiy Bondar, [3]Vadim Borovik
[1]National Aviation University, Kyiv, Ukraine
[2]Yessenov University, Aktau, Kazahstan
[3]International Research and Training Center for Information Technologies and Systems, Kyiv, Ukraine
E-mail: drzbhu@gmail.com, odarchenko.r.s@ukr.net, s.gnatyuk@nau.edu.ua, mzaliskyi@nau.edu.ua, orangearrows@bigmir.net, vadymborovyk@gmail.com

Received: 30 June 2020; Accepted: 09 September 2020; Published: 08 December 2020

Abstract: Represented paper is currently topical, because of year on year increasing quantity and diversity of attacks on computer networks that causes significant losses for companies. This work provides abilities of such problems solving as: existing methods of location of anomalies and current hazards at networks, statistical methods consideration, as effective methods of anomaly detection and experimental discovery of choosed method effectiveness. The method of network traffic capture and analysis during the network segment passive monitoring is considered in this work. Also, the processing way of numerous network traffic indexes for further network information safety level evaluation is proposed. Represented methods and concepts usage allows increasing of network segment reliability at the expense of operative network anomalies capturing, that could testify about possible hazards and such information is very useful for the network administrator. To get a proof of the method effectiveness, several network attacks, whose data is storing in specialised DARPA dataset, were chosen. Relevant parameters for every attack type were calculated. In such a way, start and termination time of the attack could be obtained by this method with insignificant error for some methods.

Index Terms: Anomaly Detection, Cyberattacks, Information Security, Data Analysis, Technology Architecture, Abnormal Traffic Behavior, Vulnerability, Security, Threat Model.

1 Introduction

Certainly, during the last 10 years, the concept of the information security is significantly strengthened not only in software manufacturer's heads, who want to protect their personal data and not suffer moral losses but also consumers, who want to protect their own personal data in case of private person, or commercial confidentiality, if participant doing his job as hired employee.

Fast growing of node quantity and their calculation power increase data quantity that transpierce through them that causes necessity of cyber security support approaches and methods improving therewith unceased dependable support of the whole enterprise. The most significant aspect of given problem solving is automatical gathering and analysis of all network traffic traversed through the node. The speed of request processing in network confunction is important, because of response delay to be within reasonable limits.

Outgoing data, in other words traffic, usually represents as standartised network packages with some parameters array that allow to classify this package as normal or abnormal [1]. Nonetheless, network attack starts forecasting ambiguates by poor level of raw data structuredness occasioned by noizes existence.

The most widespread employment of abnormal traffic behavior capturing methods is capturing of unknown attacks and invasions [2, 3]. There are two basic ways to educe network attacks that are not excluding each other: network traffic analysis and content analysis. In the first case just network package headlines are analyzing, in the second – their contents.

Statistical analysis relates to behavioral methods of violation capturing in networks and it based on juxtaposition of current state of network infrastructure to some previously definited criteria, that characterizes normal functioning of network infrastructure. Statistical analysis methods have different interpretation based on different dynamical characteristics of the network traffic, but basic principles are practically identical for all of them. Incontestable advantage of statistical methods usage is possibility of definition of first-ever negative influence employed methods on the object of attack by intruder. Also, would be necessary to certify that usage of statistical analysis methods is the most widespread type of abnormal traffic behavior detection technology implementation. In this article, while analyzing the experimental results, a synthesis of an algorithm for detecting cyberattacks based on the Neyman-Pearson criterion will be designed, which in the general case can be reduced to the CUSUM algorithm. Introducing the Neyman-Pearson criterion and corresponding CUSUM algorithm can facilitate the use of statistically optimal properties during the analysis of abnormal traffic behavior, since with such traffic behavior can be considered as a trend of a non-stationary random process. Therefore, the use of statistical techniques is reasonable.

2. Related Works Analysis

The most widespread employment of abnormal traffic behavior capturing methods is capturing of unknown attacks and invasions [4-6]. There are two basic ways to educe network attacks that are not excluding each other: network traffic analysis and content analysis. In the first case just network package headlines are analyzing, in the second – their contents.

Statistical analysis relates to behavioral methods of violation capturing in networks and it based on juxtaposition of current state of network infrastructure to some previously definited criteria, that characterizes normal functioning of network infrastructure [7-9]. Statistical analysis methods have different interpretation based on different dynamical characteristics of the network traffic, but basic principles are practically identical for all of them. Incontestable advantage of statistical methods usage is possibility of definition of first-ever negative influence employed methods on the object of attack by intruder.

Statistical analysis methods usage is the most widespread type abnormal traffic behavior detection technology implementation. Information about typical object behavior is gathering by statistical sensors and forming in the way of profile [3, 9-11]. The profile in such case is an array of parameters that characterizes object typical behavior. There is a period of the profile training and forming. The profile forming in reliance on objet statistics and mathematical statistics standard methods could be used for that, for example sliding window method [12] or weighted sum method [13].

In the context of traffic analysis, fixation methods of changes for traffic anomaly detection are of interest [13-15]. Purported that reason of traffic anomalies is significant change of traffic characteristics. However, the detection results quality not only depends on changes detection method. Much more significant is choosing of indexes that are more sensitive to operational events and events of network administration, such as: network failures, harmful traffic attacks etc. Oppositely, indexes need to be sensitive enough to traffic changes and failures, caused by legal and harmless traffic. Otherwise, large amount of false alerts could be obtained.

Statistical methods using systems are having large quantity of advantages [16]. Firstly, they do not need constant updating of the attack signature base that simplifies the problem of system data accompaniment. Secondly, they could detect unknown attacks, which signatures have not been written yet, that allows them to be the sort of breaking buffer as long as expert system patterns would be developed. Thirdly, such methods allow to detect much more complex attacks than other methods. Also, it helps to detect attacks, apportioned by time or by attack objects and adapt to user behavior changing, so statistical methods are much more sensitive to violations than people. So, detection depends on the level of adjustment to specific changes

3. Network Traffic Analysis Using CUSUM Algorithm

Commonly, network traffic contains large amounts of data, which, along with the need for real-time analytics, places stringent demands on the effectiveness of search methods and the detection of network anomalies. Therefore, statistical analysis of network traffic and the search for informative signs that shape its normal and abnormal behavior is an important task.

The thesis deals with the statistical method based on CUSUM control charts [17] as an effective statistical method for traffic anomalies detection.

CUSUM control cards, also named cumulative sums algorithm (CUSUM cumulative sum), are based on that $S_1 = s$ ($y1$, ..., yt) has negative value during normal conditions and positive value during changes. Solving at the algorithm of cumulative sums admits by comparison of g_t function to h limit:

$$g_t = S_t - \min_{1\le i\le t} S_i = \max(0, s(y_t) + g_{t-1}) = [g_{t-1} + s(y_t)] \ge h_i g_0 = 0 \tag{1}$$

The alarm occurs is gt crossing the limit h. to reset the algorithm turn g_t back to zero.

In terms of hypothesis veryfing the CUSUM control card continually accomplishes verisimilitude equation, where every solution takes into account as much consequential observations as needed for H_0 and H_1 adopting. The CUSUM control cards begin from initial SPRT, if it equals to H_0 and stops and calls the alarm in case of H_1. The h limit allows to gain the compromise between the average time of detection and the frequency of false alerts. If the Y distribution is unknown, logarithmical verisimilitude correspondence $s(y_t)$ needed to be replaced by statistical characteristics $u(y_t)$ with comparable features: average value $u(y)$ need to be negative to H_0 and positive after H_1. Such method also could be named CUSUM nonparametric algorithm [18].

Corresponding statistical characteristics for the average value positive shift detection could be witten as:

$$u^{+}(y) = y - (\mu_0 + K) \tag{2}$$

Parameter *K* called a bearing term. For the negative shifts detection another statistical characteristics needed

$$u^{-}(y) = (\mu_0 - K) - y \tag{3}$$

As a result, two functions would be obtained:

$$g_t^{+} = [g_{t-1}^{+} + y_t - (\mu_0 + K)] \geq h\,, \tag{4}$$

$$g_t^{-} = [g_{t-1}^{-} - y_t + (\mu_0 + K)] \geq h\,. \tag{5}$$

Typical settings $K = \sigma/2$ and $h = 4\sigma$ or $h = 5\sigma$, where σ is standard deviation of Y_t. CUSUM reveals slender but constant changes with higher probability, because its accumulation after a while.

DDoS-attack detection using the CUSUM algorithm

Distributed intrusion is an intrusion that satisfies the “many violators - one victim” ratio [19]. An invasion is considered as distributed if its various steps are performed on behalf of different sources in the network.

Distributed intrusion events are coordinated. The sources on whose behalf the distributed invasion (attack) is performed are interrelated.

There are different approaches of distributed denial of service attacks detection [1, 2, 17]. Some of them involve the operation of sensors around the perimeter of the network and centralized processing of data coming from them, while others use the installation of additional software at intermediate nodes on the way from the violator to the victim.

All common DDoS attacks detection approaches monitor the traffic sent from an external network to a secure station or service, and identify the attack as a noticeable deviation from some can see the characteristics of the traffic. There are three main groups of methods for detecting DDoS attacks.

The methods of the first group are based on building the activity profile of the remote stations activity profile relatively to the information protected station (service) during training and traffic characteristics comparison with profile characteristics in the detection mode. When deviations from the profile are detected, an alarm is generated. In many developments and studies, the main indicator is the average number of packages received by the station or individual network services. Statistical criteria (standard deviation, chi-square, deviation from the standard normal distribution, significant increase in entropy, etc.), clustering, etc. can be used to detect anomalies. These methods allow you to detect DoS and DDoS attacks.

To the second group, common CUSUM statistical method, based on “change point” detection was taken. In this method analysed traffic on the basis of destined IP-adress, port or protocole divides initially. After it, values of some monitored traffic parameter (quantity of new stations contacted to the server, package quantity difference with set SYN checkbox and packages with set SYN-ACK checkboxes, difference in quantity of established and closed connections etc.) will transform to some subsequence elements.

In the presence of an attack, the current sequence element values will be significantly different to the previous members of the sequence. This method has several significant advantages over other methods. Firstly, the advantage of the method is the high operation speed, which allows it to be applied in real time. Secondly, the method adapts to different network loads, upon the condition of their constance. Additionally, this algorithm is considered to be the best among algorithms with a certain level of false positives, since the algorithm parameters are calculated on the basis of a formula that relates to the algorithm threshold value for the attack detection and detection time.

The third group of methods includes spectral analysis and wavelet analysis based on the spectral characteristics analysis of traffic to detect DDoS attacks.

Fore DDoS-attack detection “traverse point” detection based method has been chosen because of its operation speed and insignificant memory losses. For DDoS detection has been proposed tracing of such traffic characteristics as:

- the quantity of new stations called to the service;
- the quantity of different stations called to the secured service;
- the difference between established and closed connections quantity.

DDoS attack detection based on the concordance between SYN and FIN checkboxes

Concordance verifying between packages with setted SYN checkbox and setted FIN (RST) for detection is proposed in [20]. Upon that, connection closing probability by RST sending needed to be taken into account. Normally, there are two variants of TCP-protocol normal behavior: package sending with setted FIN checkbox and connection failure by RST.

RST could be received in two cases: during the package receiving by the server on the closed port (passive) and during the clear connection closing in case of failure (active). During the gathering traffic analysis, there is no possibility to detect in which of two indicated cases RST has been received. So, there are two mutually exclusive variants: all the RST packages are passive or all the RST packages are active. Quantity of false alerts increases in the first case. Sensitivity during the attack detection decreases in the second case.

As the detection method of SYN-FIN (RST) couples from the limit value in [21] CUSUM (cumulative sum) statical method is proposed. CUSUM allows to monitor changes of commanded parameter iterationally, detecting “traverse points”. Herewith, traffic parameters gathering accomplishes within the equal time segments.

CUSUM algorithm parameters selection

Limit N in the algorithm has no user-friendly sense, so it can not be established antecedently. In [22] following approach to the definition of his value is proposed. The time between the false alerts increases expotentially to N. Let us set detection time as $\tau_n = inf\{n: dN(\cdot) = 1\}$, and normalized attack detection time as:

$$p_n = \frac{(\tau_n - m)^+}{N}, \tag{6}$$

where m is attack start time.

Limit value can be calculated by the formula with preset parameters a and h, which selection is not strictly regulated. During the attack, the value of the subsequence element is rising sharply, that causes his significant deviation from the average value. Parameter a is the top limit value of elements during the normal functioning, so it must be very small number in absolute value (a <1). Parameter h, as had been mentioned before, is the low limit of $\{Z_n\}$ subsequence elements deviation from an average value during the attach occurring, than it is significantly greater than c. In this case detection is senseless to the selection of a. In [23] parameter h is proposed to select as 2 | a |, such condition to the selection of the h is reasoned theoretically. In [24] during the CUSUM usage whilst the new IP-adresses monitoring for the DDoS-attacks detection, values a = 0,05 and h = 0,1 were selected, during the realization tests with such parameters there was no significant quantity of false alerts distinguished.

In the case of an attack h is dramatically overpasses (that is right for given realization), so for the detection of all three parameters a was chosen as a small positive number and h value is equal to $2|a|$. After the problem a, h, value γ is obtaining if $c = 0$. The next step is to obtain τn value. Usually τ_n is a little bigger than m, for example in the work [24] $\tau_n = m + 1$. From this equation the value of N limit could be obtained. During the β parameters selection, that is connected to a and N, two basic points are strived: reducing quantity of false alerts and attack detection time minimization. β is using for traverse accomplishing from $\{X_n\}$ to $\{Z_n\}$.

The bigger is β, the less often positive values would be find in $\{Zn\}$. So the y_n subsequence would be receiving big values for attack detection with lower probability. N – limit value for yn subsequence. The bigger is N, the less false alert quantity and the longer is time of an attack detection. Consequently, N could be obtained from a and h values.

As can be seen from the described, the CUSUM algorithm parameter selection for all three cases is based on the fact that if DDoS-attack is occurred, the subsequence element, formed on the basis of detected traffic parameter is significantly different to the subsequence elements during the normal functioning. Formulas, based on the detection time and false alerts quantity are using during the calculation of the algorithm parameters.

Different IP-adresses monitoring in the input traffic

During the IP-adresses monitoring in input traffic (Source IP Monitoring, SIM) CUSUM method using to quantity of new IP-adresses per session. Proposed, that calling of big number, that has not called to station servise earlier could be an evidence of the DDoS-attack. Such method needed preliminarily training – familiar IP-adresses database organization.

New IP-adresses monitoring in the input traffic is composed of two parts: training and detection.

Training is comprised of legitimate IP-adresses addition to the database of IP-adresses (IPAddress Database, IAD) [25]. On the stage of a detection the calculation of the first time calling IP-adresses stations quantity is accomplishing and the deviation of this value from the average value is obtaining by the CUSUM test. Additionally, on the stage of familiar IP-adresses detection, station addresses, that are overpassed quantity of the interaction with service sessions required by the administrator are adding

Non-Parametrical multidementional CUSUM tests for fast DOS-attack detection in computer networks. DOS-attacks and worms are the most common and the most dangerous categories of external violations, that are involve an extensive spectre of network and computer recourses. Typical DOS-attacks and worms could cause significant changes into the statistical traffic model in contrast to the normal traffic. What is more, such changes are taking place at the unknown time moments and they need to be detected “as soon as possible”. So, the natural way of computer attack detection fits into the so-called changepoint detection theory: to detect changes into the distribution with the least delay and to save herewith false alerts rank (FAR) on the low level.

CUMSUM nonparametrical multidimentional algorithm has several ostensive advantages [18].

Firstly, it uses minimum of available data that is very important, because of the permanent distribution (legitimate traffic) and changed (attack) traffic, usually, are not known beforehand. Secondly, the algorithm having controllable calculated complexness and can be used on-line.

For ICMP- flood i UDP- flood attacks processing, we would be monitor quantity of received packages with its dividing by size and type. So, for TCP SYN attacks we would control buffer sizes linked to received and delivered SYN packages.

Let Δ*n* = t_n-t_n-1 – is a retrieval from *n*-interval. For each p_t package type (ICMP, UDP, TCP etc.) package would be classified by size and carried to the Mpt group. In case of UDP and ICMP attacks, general quantity of N_{pt} *n* and p_t-packages type with dimentions in the *i*- th group, received in the *n*-th time interval would be monitored.

Thirdly, the algorithm is adaptive and self-training, that allows it to adapt to different network loads and utilization of different models.

To detect TCPSYN attacks, B_n buffer dimentions in the end of *n*-time interval needed SYN-packages would be backtraced. In the testing mode N_{pt} *n* and B_n statistics are controlled simultaneously.

Suppose, that series of unfortunate values X_1, X_2, ..., exists and that values are choosed for observation and having general solidity of probabilities p_0 (X_1, ..., X_n) for *n* <*λ* and another solidity of probabilities p_1 (X_1, ..., X_n) for *n*> *λ*, where *λ* ∈ {1,2, ...} – is unknown measurement point. Otherwords, in dependence of changepoint *λ* = *k* and vector with (*n*-1) monitorings (x_1, ..., x_n-1) the conditional dispensation solidity of the *n*-th observation x_n would be:

$$p(X_n \mid X_1,...,X_n,\lambda = k) = \begin{cases} p_0(X_n|X_1,...,X_{n-1}), k>n; \\ p_1(X_n|X_1,...,X_{n-1}), k\le n \end{cases} . \tag{7}$$

where p_0 and p_1 are permanent and changed probability density respectively.

The consequential changepoint detection procedure defines by the time of stop *τ* for the monitored consequence {*Xn*} *n*> 1, so *τ* – is an accidental value that depends on observation. The consequential changepoint detection algorithm uses logarithmic likelihood dependence (LLR) for hypothesis H_k: *λ* = *k*, that change taken place in the point *λ* = *k*, and H_∞: *λ* = ∞, that there are no change. It defines in the next way:

$$z_n^k = \sum_{j=k}^{n} \log \frac{p_0(X_n \mid X_1,...,X_{n-1})}{p_1(X_n \mid X_1,...,X_{n-1})}, n \ge k \tag{8}$$

CUSUM procedure defines, when LLR statistical maximum U_n overpasses in the first time limit *h*:

$$\tau_{CU} = \tau_{CU}(h) = \min\{n \ge 1 : U_n \ge h\}$$

Inasmuch as network attacks take place in the unknown moment of time and usually cause the sharp changing of the normal network traffic, its supposed, that anomaly detection by the changepoint method ideally fits to DOS-attack typical scenarios. But highly developed parametrical methods would not be effective enough, because they need too much lacking beforehand information. So, reliable nonparametrical methods are turned out to be very important. Proposed nonparametrical multidimentional CUSUM method has several attractive features. Firstly, it can be easily implemented into the distributed computer networks. Secondly, it is practically does not need beforehand information, so it is very reliable. Thirdly, when it is optimized and configured, this method becomes very effective. Researches showed that given algorithm allows to detect rapidly different DOS-attacks with a low level of false alarms.

4. Experimental Results and Discussions

Input data obtaining and preparing for traffic analysis. In general, the algorithm of analysis of network traffic is as follows (Fig. 1). Usually, general scheme of network trafic analysis has few predefined steps, which are ahown on Fig. 1. First of all mode of data analysis has to be chosen. There are several possible approachs for this: slising – analysis of some prefix of the transferred data; sampling – analysis of the part of all packages; deep packet capture – all data analysis. Then chosen mode is activated and the system starts to collect and analyse transferred data depending on chosen on previous steps mode. The next step of the algorithm is Anakysing process, during which system has to identify traffic characteristics and decide if the behaviour is normal or has anomalies. After these steps results of the algorithm are generating.

After the turning-on of the network package capturing mode, the network adapter switches to the Promiscuous mode and fixates each package traversed through interface. It is conditionned by the information transfer technology in Ethernet networks. Some restrictions in the analysis mode give a possibility to reduce the volume of analysing information by chopping off the obsolete data excess and system performance increasing. SQL-databases and shaped executive text files, that storage number characteristics are using as vaults.

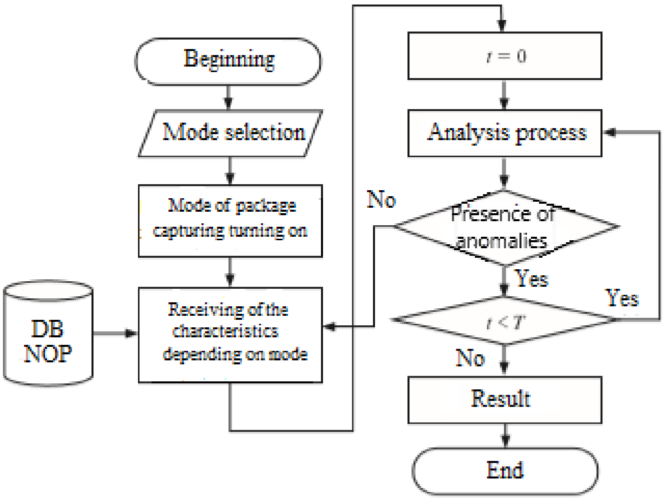


Fig.1. Network control procedure algorythm. DB NOP – databank of normal network operation patterns

An output traffic data, verified to the anomalies availability later on, proposed to receive data obtained by sniffer-program Wireshark (or any other program complex, that accomplishes the network activity dump removal mission).

The Wireshark program complex could be used for “Live” traffic recording from interfaces that are part of the analysis system as well as for analysis of removed and previously saved traffic.

TCPdump – is common program instrument, that allows exploring network data transfer process extensively. The TCPdump conclusion possess the network connections package data in order of appearance of the package at the network. Package data needed to be previously processed before the data gathering procedure. TCPdump data convertors transforms connection records to the infinity of features (or attributes), for example time (time of linking start, or timestamp of the first package), dur (linking duration), src and dst (host source and destination), bytes (the number of sended from source to destination bytes of data), srv (service, or the port of destination), fl ag (how linking responds to the network protocole) etc. These features are summarizing the package level information in the linking limits.

Traffic samples analysis. Make an analysis of datasets that have been received by DARPA Intrusion Detection Evaluation Group at MIT Lincoln Laboratory [29] and by our experiment.

The process of the raw data preparing consists of three steps.

Received traffic dump loading in Wireshark (Fig.2). Making filtration of the dump by the time when attack took place. These data has been taken from the documentation on the laboratory website. So, several attacks that we will consider later on were chosen for the experiment.

Wireshark complex includes variety of analysis methods and statistical data gathering methods that are very necessary for the work. There is an instrument IO Graphs (Fig. 2) for statistical messages reading, that are distributed with some frequency. It is enough to press the key Copy for the statistical data saving and, after the successful registration in clipboard, paste the data into any text file.

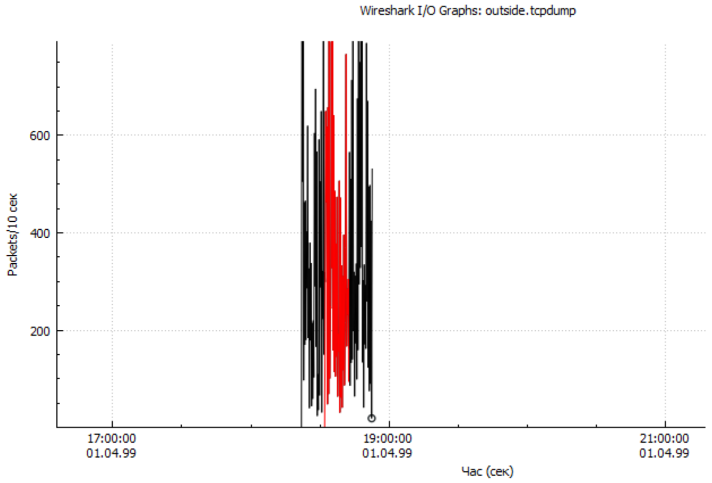


Fig.2. IO-Graphs statistical instrument window with 10 s point

1. The formatting of the received data because for further analysis we need to know just the quantity of packages.
2. Necessary operations for traffic analysis by CUSUM statistical method were designed in the MathCad program environment. Also, for results comparison, an experiment was performed. Stages of the experiment:

1) Running the Wireshark program on the server laptop.

After the beginning of traffic capturing, Wireshark in the online mode captures network packages and presents it on the window of the user interface. The conclusion could be done that information transmits uninterruptedly in the local network. Let’s consider graph of package transmitting in the local network over 5 minutes (Fig. 3).

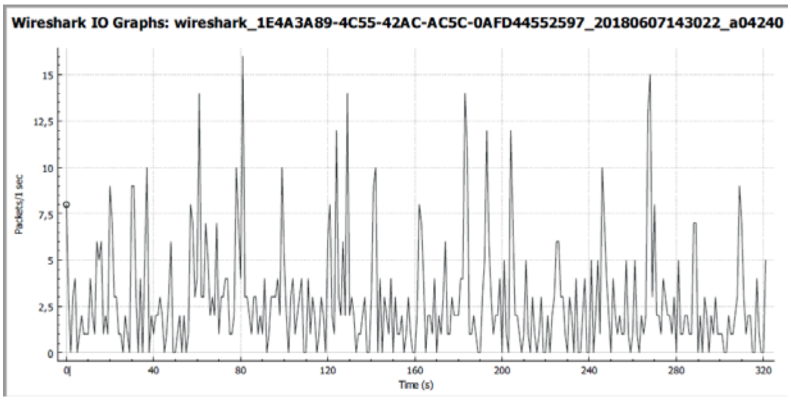


Fig.3. Network traffic analysis within 5 minutes

A conclusion based on this diagram could be done, that during the normal work of the network values are not overpassing 8 packages per second, averagely. Top points are the network anomalies.

It is well known that for latency time in Poissonian process expotential distribution is using. If we expect the event that takes pace every 1/λ units of time (with intesity λ), than latency time distribution is equal to:

$$p(x) = \lambda e^{-\lambda x}$$

The input values were analysed and the bar chart of traffic without an attack subordinated to the expotential distibutional law was built by utilizing the program (Fig.4).

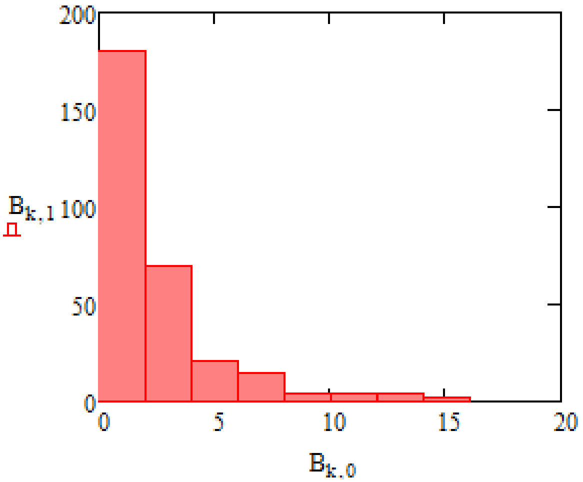


Fig.4. The histogram of the traffic model without an attack

So, the distribution density of probabilities in case of the absence of an attack is equal to

$$f_0(x) = \lambda_0 e^{-\lambda_0 x}, \tag{9}$$

where holding time

$$\lambda = \frac{1}{T}, \tag{10}$$

$$T = \frac{1}{n}\sum_{i=1}^{n} x_i \ . \tag{11}$$

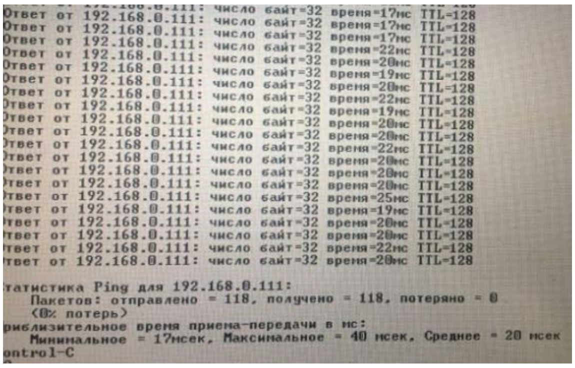
So, we decided to use the CUSUM method for stability control of the model parameters on the whole retrieval.

2) On this stage a PING type attack has been simulated.

For this we would ping our server from four laptops simultaneously for certain distinguishing of the beginning and the end of an attack

PING is the basic assistant of the Windows command line for linking examination in networks, based on TCP/IP. The PING command by the message sending with an echo request, according to ICMP protocole controles the linking on the level of IP protocole with another TCP/IP supporting computer.

DDoS-attach accomplishes in such way: sending the 32-byte package on the server with the IP-adress 192.168.0.111 and receiving the answer, at an average, in about 20 ms. Time to live (TTL) of the package 128. Normally, we have sent 118 packages from each attack laptop (Fig. 5).



```
Ответ от 192.168.0.111: число байт=32 время=17мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=17мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=17мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=22мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=20мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=19мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=20мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=22мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=19мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=20мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=20мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=22мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=20мс TTL=128
Ответ от 192.168.0.111: число байт=32 время=20мс TTL=128
твет от 192.168.0.111: число байт=32 время=20мс TTL=128
твет от 192.168.0.111: число байт=32 время=25мс TTL=128
твет от 192.168.0.111: число байт=32 время=19мс TTL=128
твет от 192.168.0.111: число байт=32 время=20мс TTL=128
твет от 192.168.0.111: число байт=32 время=20мс TTL=128
твет от 192.168.0.111: число байт=32 время=22мс TTL=128
твет от 192.168.0.111: число байт=32 время=20мс TTL=128

татистика Ping для 192.168.0.111:
    Пакетов: отправлено = 118, получено = 118, потеряно = 0
    (0% потерь)
риблизительное время приема-передачи в мс:
    Минимальное = 17мсек, Максимальное = 40 мсек, Среднее = 20 мсек
ontrol-C
```

Fig.5. Accomplished DDoS attack

On the graph, we can see a sharp increasing of the number of packages per second, which means the beginning of an attack and decreasing of the package number, indicating the end of the attack (Fig. 6).

The likelihood function is equal

$$f_1(x) = \lambda_1 e^{-\lambda_1 x} \ . \tag{12}$$

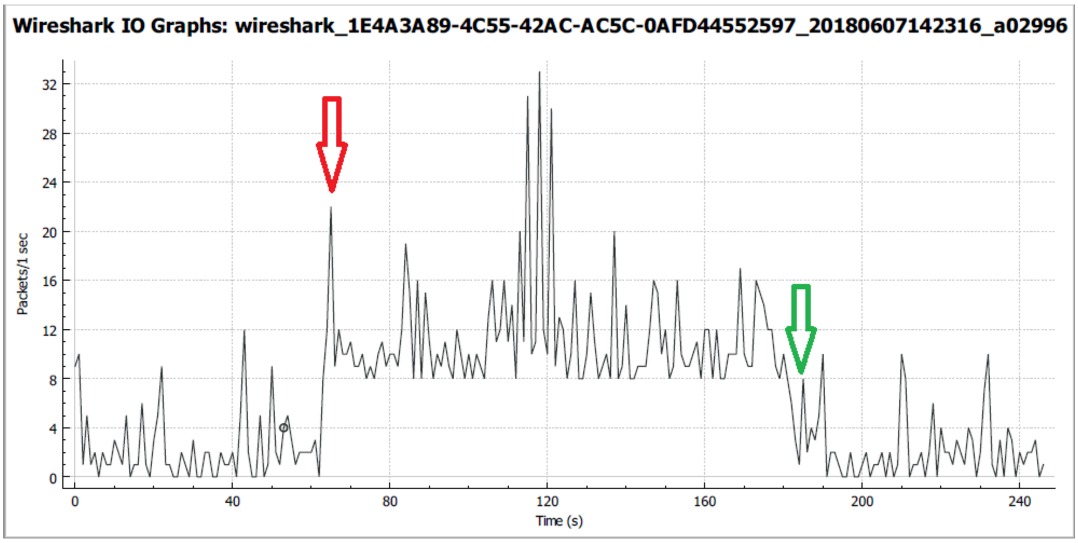


Fig.6. The beginning and the end of an attack

3. Due to the received results, we could calculate the value of h.

For the calculation of this value, let’s find likelihood equilibration, that equals

$$\Lambda_i = \prod \frac{f_1(x)}{f_0(x)} = \prod_{j=1}^{i} \frac{\lambda_1}{\lambda_0} * \frac{e^{-\lambda_1 x_j}}{e^{-\lambda_0 x_j}} \ . \tag{13}$$

To simplify the equilibration, we need to prologify it:

$$\ln \Lambda_i = (n-i-1)\ln\frac{\lambda_1}{\lambda_0} + \lambda_0(1-\frac{\lambda_1}{\lambda_0})\prod_{j=1}^{i} x_j \tag{14}$$

So, the logarythmic likelihood equilibration is the determinative statistics.

The limit value *h* is equal to the first class error function *h*=*g*(*α*). There are two ways of this problem solving: analytical and modelling. By reference to the fact that analytical method needs complex formulas and huge quantity of time, the modelling method has been used in the development.

Let's suppose that *α* = 0.0001. So, after 10000 times calculated likelihood function without an attack and after the decisive statistics finding we obtained the function maximums and that maximums maximum would be the limit value. In such case *h*=20.

After the overpassing of this value we could talk about the attack on the network (Fig. 7).

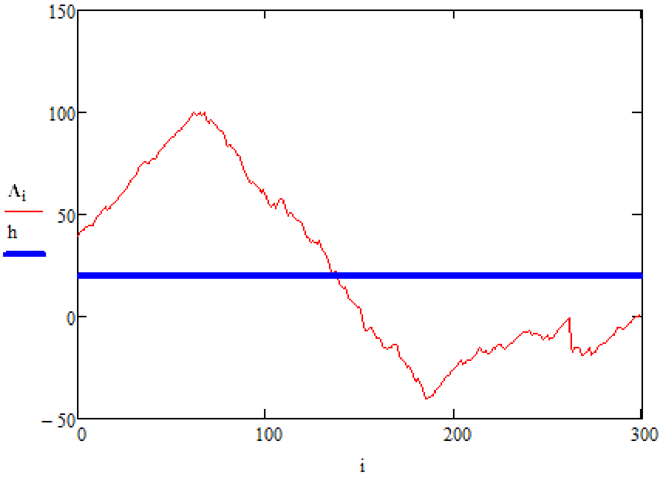


Fig.7. Attack diagram with limit value setting

Let's evaluate moment of the beginning and the end of an attack, that is the maximum and minimum of the function. The result shows that attack started on 62-th second and ended on 185-th second.

The selected method testing for different types of attacks. Let's prove the effectiveness of proposed statistical method for different attack types. Tested attack types are:

1. *Portsweep network attack* – the observer examines the infinity of ports for definition of services that are supported on the singular host [25]. Smurf. Experts relate SMURF attacks to the most dangerous type of DDoS attacks because it has an amplification effect that is the result of straight broadcasting ping requests sending to systems that are obliged to send a response. The request alignes to the network address or to the broadcasting distribution address, but anyway the device must make a transformation of 3 (IP) level transformation to 2 (network) level, because RFC 1812 Requirements for IP Version 4 Routers require that. In the standard C-class network (24-bit extraction address) the network address would be .0 and the broadcasting distribution address would be .255. The broadcasting streightline usually uses for diagnostics and allows to detect working systems without a ping request on each address in range. The smurf attack uses streight broadcasting distribution features and demands three participants minimally: attacking, amplifying and victim. An attacking participant sends mystified package ICMP ECHO to the network amplifying broadcasting distribution address. The source address is changing to the address of the victim for presenting it as target system is the one who has initiated the request. The next event takes place after it: as far as the ECHO package has been sent to the broadcasting address, all amplifying systems retumt to the victim their own answers (if only configuration does not defines another behavior). It uses the IMCP diagnostical protocole. After the one ICMP package sending to the network that consists of 100 systems, an intruder initiates 100 times amplifying of the attack! Amplifying index depends on the network structure, so intruder looks up the huge network that has an ability to kill a victim-network. During the receiving of the requests, all computers of the subnetwork send the echo-responds to the address of the victim that causes ceasing of the work, because of large information amount.
2. *Mailbomb* – one of the simplest network attack type [26]. An intruder sends to the user's computer or to the mail-server of the company one large message or infinity (dozens of thousands) of mail-messages on the victim's mail-server SMTP-port, that causes brakedown of the system, if special blocking programs have not been installed.
3. *Neptune network attack* – an example of DoS attack, that accomplishes flooding distribution (SYN packages) to

one or multiple ports [27-29]. Normal TCP-link installing takes place with respect to the “trilateral handshake” scheme: client sends SYN, server responds SYN-ACK, client responds ACK. If the client’s side sends a huge amount of SYN-packages, the server would be every time forced to initiate linking, that would not be esteblished. The quantity of such linking is limited and when it would achieve the limit, the server will stop to answer the any request.

4. *Ipsweep*. There are two types of port scanning: vertical, when one host is scanning by all of the open ports and horizontal, when the group of hosts is scanning on the some opened port. In the majority of cases, the SYN-package sends on the port and if port is open, he responds SYN-ACK. Also, ast ways are described in [28].

So, taking into account the possible error, the CUSUM method could detect different attack types vaguely certainly. The Ipsweep attack can not be detected, but the complex type of such attack should be taken into account. So, other statistical anomaly detection methods are proposed to test for this type.

Conclusions. Statistical methods are very effective in the field of anomalies detection these days. Metrics have been considered for the CUSUM method that is one of the most effective methods these days [30-33].

Table 1. Attack detection by CUSUM method accuracy comparison

Attack type	Attack beginning time (real)	Attack beginning time (experiment)	Attack duration (real)	Attack ending time (real)	Attack ending time (experiment)	Error beginning/ending (%)
Mailbomb	18:32:17 (9133 s.)	9130 s.	00:10:07	18:42:24 (9740 s.)	9740 s.	0.3% 0%
Portsweep	16:43:15 (2592 s.)	2592 s.	00:03:54	16:47:09 (2826 s.)	2826 s.	0% 0%
Neptune	18:04:04 (11041 s.)	11055 s.	00:06:51	18:10:55 (11541 s.)	11567 s.	1.2% 0.2%
Smurf	18:29:25 (8962 s.)	8962 s.	00:03:01	18:32:26 (9143 s.)	9143 s.	0% 0%
Ipsweep	15:00:16 (12 s.)	Failed to detect	00:15:21	15:15:37 (934 s.)	Failed to detect	-
Ping flood	15:10:00 (62 s.)	62 s.	00:02:03	15:12:03 (185 s.)	185 s.	0% 0%
Mailbomb	18:32:17 (9133 s.)	9130 s.	00:10:07	18:42:24 (9740 s.)	9740 s.	0.3% 0%
Mailbomb	16:54:17 (3254 s.)	3260 s.	00:03:01	16:47:09 (3434 s.)	3444 s.	0.1% 0.2%
Mailbomb	12:32:17 (1937 s.)	1970 s.	00:12:59	12:45:16 (2716 s.)	2730 s.	1.7% 0.5%

5. Conclusions

In the context of traffic analysis, methods of changes fixation for traffic anomaly detection are of interest. Arranged, that the reason of traffic anomalies is a significant change of some traffic characteristics. But the result’s quantity depends not only on selected method of changes detection. The more important are considered traffic indexes that are sensitive to the events relevant to operations and network administration, such as: network failure, harmful traffic attacks etc. Oppositely, indexes need to be sensitive enough to the traffic changing and failures, caused by legal and harmless traffic. In the other case, big amount of false alerts could be received.

Statistical methods using systems have many advantages. They do not require constant updating of the attacks database that greatly facilitates the task of maintaining these systems. They can detect unknown attacks whose signatures have not yet been written, allowing them to be a kind of deterrent buffer until expert systems templates would be developed. Unlike the other methods, more sophisticated attacks can be detected. They can also detect attacks distributed over time or across targets, adapt to changes in user behavior, so they are more sensitive to intrusions than humans.

Detecting depends on the adaptation level of the method to the specific changes of traffic realization statistical features that we can observe. Looking on all that, conclusion can be made that CUSUM algorithm is an effective method for traffic anomalies problem solving.

In the development local network and PING type network attack detection by the help of Wireshark program were experimentally analysed. Bar graphs of the attack model are builded, based on the obtained values. Also, the limit value was obtained by the experiment, overpassing of this value signalizes about the beginning of the attack. For the method effectiveness prooving, several network attacks have been chosen. Data of that attacks storages in specialised DARPA dataset. Corresponding parameters for each type of attack have been calculated. So, by the help of this method, time of the beginning and the end of an attack can be obtained with insignificant fault for some methods. Analysis of accuracy comparison during attack detection by CUSUM method shows that maximum error for different attack types does not exceed 1.7 % that proves accuracy and reliability of the results. Looking on all that, can be said that CUSUM algorithm is an effective method for traffic anomalies problem solving and this method can be proposed for usage in contemporary intrusion detection systems.

Acknowledgement

This scientific work was supported by RAMECS, CCNU16A02015, Young Scientists Association of National Aviation University (Kyiv, Ukraine) as well as International Research and Training Center for Information Technologies and Systems (Kyiv, Ukraine).

References

[1] Ranjan R., Sahoo G. A new clustering approach for anomaly intrusion detection, International Journal of Data Mining & Knowledge Management Process (IJDKP). 2014. vol. 4. No. 2. pp. 29–38.

[2] Barbara D., Wu N., Jajodia S. Detecting Novel Network Intrusions Using Bayes Estimators, Proceedings of the First SIAM International Conference on Data Mining. 2001. pp. 30–49.

[3] Mazurek M., Dymora P. Network anomaly detection based on the statistical selfsimilarity factor for HTTP protocol, Przeglad elektrotechniczny, 2014, pp. 127–130.

[4] Gu Y., McCallum A., Towsley D. Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation, Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. 2005. pp. 32–32.

[5] Barford P., Kline J., Plonka D., Ron A. A signal analysis of network traffic anomalies, Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement, 2002. pp. 71–82.

[6] J. Olivain and J. Goubault-Larrecq. Detecting subverted cryptographic protocols by entropy checking. Research Report LSV-06-13, Laboratoire Specification et Verifi-cation, ENS Cachan, France, June 2006.

[7] D. E. Taylor, "Survey and taxonomy of packet classification techniques," ACM Comput. Surv, vol. 37, No. 3, pp. 238–275, 2005.

[8] Colin J. Bennett, Andrew Clement, Kate Milberry. Introduction to CyberSurveillance. Cyber-Surveillance in Everyday Life,vol. 9, No. 4 (2012)

[9] Callado A., Kamienski C., Szabo G., Gero B., Kelner J., Fernandes S., Sadok D. A Survey on Internet Traffic Identification; Communications Surveys & Tutorials, IEEE Volume 11, Issue 3, 3rd Q 2009, pp. 37-52.

[10] S.-H. Han, M.-S. Kim, H.-T. Ju and J.W. Hong, "The Architecture of NGMON: a Passive Network Monitoring System for High-Speed IP Networks", Accepted to appear in the Proc. of the 13th IFIP/IEEE International Workshop on Distributed Systems: Operations & Management (DSOM 2002), Montreal, Canada, October 21-23, 2002.

[11] Duffield, N.; Lund, C.; Thorup, M., "Learn more, sample less: control of volume and variance in network measurement", IEEE Transactions in Information Theory, vol. 51, No. 5, pp. 1756-1775, 2005.

[12] W. Wu et al., "Sliding Window Optimized Information Entropy Analysis Method for Intrusion Detection on In-Vehicle Networks," in IEEE Access, vol. 6, pp. 45233-45245, 2018.

[13] N. Gupta Gourisetti, M. Mylrea and H. Patangia, "Application of Rank-Weight Methods to Blockchain Cybersecurity Vulnerability Assessment Framework," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0206-0213.

[14] W. Zhang, Q. Yang and Y. Geng, "A Survey of Anomaly Detection Methods in Networks," 2009 International Symposium on Computer Network and Multimedia Technology, Wuhan, 2009, pp. 1-3.

[15] T. Salman, D. Bhamare, A. Erbad, R. Jain and M. Samaka, "Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments," 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud), New York, NY, 2017, pp. 97-103.

[16] C. Callegari, S. Giordano and M. Pagano, "Anomaly detection: An overview of selected methods," 2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, 2017, pp. 52-57.

[17] M. Zaliskyi, R. Odarchenko, S. Gnatyuk, Yu. Petrova and A. Chaplits, Method of traffic monitoring for DDoS attacks detection in e-health systems and networks, CEUR Workshop Proceedings, vol. 2255, pp. 193-204, 2018.

[18] Q. Zhang, C. Rendon, V. M. D. Oca, P. D. R. Jeske and D. M. Marvasti, "A Nonparametric Cusum Algorithm for Timeslot Sequences with Applications to Network Surveillance," 10th IEEE High Assurance Systems Engine-ering Symposium (HASE'07), Plano, TX, 2007, pp. 435-436.

[19] Kruegel C., Toth T. Using Decision Trees to Improve Signature-Based Intrusion Detection, Recent Advances in Intrusion Detection, 2003, pp. 173–191.

[20] Shelukhin O.I., Garmashev A.V. Detection of anomalous emissions of telecommunication traffic using discrete wavelet analysis, Electromagnetic waves and electronic systems, 2012, No. 2. pp. 15–26.

[21] Wang H., Zhang D., Shin K.G. Detecting SYN flooding attacks, Proceedings of IEEE INFOCOM'2002, New York City, NY, 2002, pp. 1530–1539.

[22] Peng T., Leckie C., Ramamohanarao K. Detecting distributed denial of service attacks using source IP address monitoring, Proceedings of the Third International IFIP-TC6 Networking Conference (Networking 2004), pp. 771–782.

[23] Sheluhin O.I., Atayero A.A. Integrated Model for Information Communication Systems and Metworks, Design and Development. IGI Global, USA, 2012. 462 p.

[24] Z. Hassan, R. Odarchenko, S. Gnatyuk et al, Detection of Distributed Denial of Service Attacks Using Snort Rules in Cloud Computing & Remote Control Systems, Proceedings of IEEE 5th Intern. Conf. on Methods and Systems of Navigation and Motion Control, October 16-18, 2018. Kyiv, Ukraine, pp. 283-288.

[25] O. Al-Jarrah and A. Arafat, "Network Intrusion Detection System using attack behavior classification," 2014 5th International Conference on Information and Communi-cation Systems (ICICS), Irbid, 2014, pp. 1-6.

[26] Q. Zhou, W. Hu and W. Zhu, "Detection of Mailbomb Attacks Base on Time Interval Temporal Logic," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 1078-1080.

[27] R. Odarchenko, S. Gnatyuk, T. Zhmurko et al, “Improved Method of Routing in UAV Network”, Proceedings of the 2015 IEEE 3rd Intern. Conf. on Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), Kyiv, Ukraine, October 13-15, Vol. 1, 2015, pp. 294-297.

[28] Jung, J., Krishnamurthy, B., Rabinovich, M.: Flash Crowds and Denial of Service Attacks: Characterization and Implications for CDNs and Web Sites. In: Proc. Int’l World Wide Web Conference, ACM Press, New York, 2002, pp. 252–262.

[29] Cyber systems and technology. DARPA Intrusion Detection Data Sets, Available Online, URL: https://archive.ll.mit.edu/ideval/data/index.html

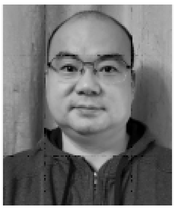
[30] V. Mosorov, A. Kosowski, R. Kolodiy, Z. Kharkhalis, “Data Traffic Modeling During Global Cyberattacks”, International Journal of Computer Networks and Information Security, vol.7, no.11, pp.20-36, 2015.

[31] I. Parkhomey, S. Gnatyuk, R. Odarchenko, T. Zhmurko et al, “Method For UAV Trajectory Parameters Estimation Using Additional Radar Data”, Proceedings of the 2016 4th International Conference on Methods and Systems of Navigation and Motion Control, Kyiv, Ukraine, October 18-20, 2016, pp. 39-42.

[32] F. Adeyinka, E. S. Oluyemi, A. N. Victor, U. C. Uchenna, O. Ogedengbe, S. Ale, “Parametric Equation for Capturing Dynamics of Cyber Attack Malware Transmission with Mitigation on Computer Network”, International Journal of Mathematical Sciences and Computing, Vol.3, No.4, pp.37-51, 2017.

[33] Y. Ghaderipour, H. Dinari. “A Flow-Based Technique to Detect Network Intrusions Using Support Vector Regression (SVR) over Some Distinguished Graph Features”, International Journal of Mathematical Sciences and Computing, Vol.6, No.4, pp.1-11, 2020.

Authors’ Profiles



Zhengbing Hu

Visiting Prof., DSc Candidate in National Aviation University (NAU, Kyiv, Ukraine) from 2019. M.Sc. (2002), PhD. (2006) from the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”. Postdoc (2008), Huazhong University of Science and Technology, China. Honorary Associate Researcher (2012), Hong Kong University, Hong Kong. Major research interests: Computer Science and Technology Applications, Artificial Intelligence, Network Security, Communications, Data Processing, Cloud Computing, Education Technology.

Deputy Director, International Center of Informatics and Computer Science, Faculty of Applied Mathematics, National Technical University of Ukraine “Kyiv Polytechnic Institute”, Ukraine (2017-).

http://www.icics.net/

Roman Odarchenko

DSc, Associate Professor. He received the B.S. and M.S. degrees in telecom-munications from the NAU the Ph.D. and DSc. degrees in telecom-munications systems and networks from the NAU, Kyiv, Ukraine, in 2013 and 2019 respectively.

From 2010 to 2017, he was an associated professor in NAU. Since 2017, he has been a vice-dean of Faculty of air navigation, electronics and telecommunications in NAU. He is the author of more than 70 articles, and more than 50 inventions. His research interests include telecommunication systems and networks, mobile networks, wireless systems, software-defined networking, network security systems etc.

Sergiy Gnatyuk

DSc, PhD, Associate Professor. In 2007 he received MSc degree in information security from NAU. He received PhD in Eng degree in cyber-security from NAU in 2011 and DSc in 2017. In 2014 he received Associate Professor degree.

Vice-Dean of the Faculty of Cybersecurity, Computer & Software Engineering. Scientific Adviser in NAU Cybersecurity R&D Lab http://cyberlab.fccpi.nau.edu.ua/ IEEE Member, Scientific Adviser of Engineering Academy of Ukraine. Research interests: cryptography, quantum key distribution, network & internet security, information security incident management, cybersecurity & CIIP.

Maksym Zaliskyi

PhD, Associate Professor.

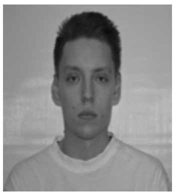
He received his B.Sc. and M.Sc. degrees in radio engineering from National Aviation University, Ukraine, in 2005 and 2007, respectively. He obtained a PhD degree in 2012 in operation and repair of transport means from the National Aviation University.

He has authored 114 published papers on maintenance and operation of radio equipment, statistical data processing and heteroscedasticity analysis. His research interests include design and improvement of the operation system for radio equipment, statistical data analysis.

Anastasiia Chaplits

In 2020 she received Master`s degree in telecommunications from National Aviation University. Besides, in 2018 she has received a grant for education from Erasmus association, so she studied telecommunications for 6 months in Bilbao, Spain.

During her study in university she wrote many research articles in telecommunication systems and networking.



Serhii Bondar

PhD Student (2017-2021). In 2017 he received MSc degree in National Aviation University (Kyiv, Ukraine). He is currently working at International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences (NAS) of Ukraine and Ministry of Education and Science (MES) of Ukraine.

Research interests: Cybersecurity, Telecomunication systems, Unmanned Aerial Vehicles, Data Processing.

Vadym Borovyk

PhD Student (2017-2021). In 2017 he received MSc degree in National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” (Kyiv, Ukraine). Since 2017 he works at International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences (NAS) of Ukraine and Ministry of Education and Science (MES) of Ukraine.

Research interests: computer engineering, cybersecurity, data processing, network security.

How to cite this paper: Zhengbing Hu, Roman Odarchenko, Sergiy Gnatyuk, Maksym Zaliskyi, Anastasia Chaplits, Sergiy Bondar, Vadim Borovik, "Statistical Techniques for Detecting Cyberattacks on Computer Networks Based on an Analysis of Abnormal Traffic Behavior", International Journal of Computer Network and Information Security(IJCNIS), Vol.12, No.6, pp.1-13, 2020. DOI: 10.5815/ijcnis.2020.06.01