ISSN 2347 - 3983
Volume 8. No. 5, May 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter15852020.pdf
https://doi.org/10.30534/ijeter/2020/15852020

Comparative Analysis of Methods Content Filtering Network Traffic

Gulomov Sherzod Rajaboevich[1],Karimova Dilbar[2], Akbarova Shokhida Azatovna[3], Qosimova Gulnora Ismoilovna[4]
[1]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan, sherhisor30@gmail.com
[2]Tashkent State Technical University named after Islam Karimov, Uzbekistan, dilbar.karimova.46@mail.ru
[3]Tashkent State Technical University named after Islam Karimov, Uzbekistan, sohidaakbarova9@gmail.com
[4]Tashkent State Technical University named after Islam Karimov, Uzbekistan, gqosimova@gmail.com

ABSTRACT

In this paper are analyzed the technical methods for filtering network traffic, their advantages and disadvantages. As well as A comparative analysis are carried out of the method of monitoring and filtering network traffic by intercepting network packets implementing deep packet analysis technologies, the machine learning method Random Forest for filtering traffic, which is an ensemble method that works by constructing many decision trees, the method of filtering traffic by calculating the entropy increment for each of the filter attributes, the method of filtering http-packets allows you to reduce the user waiting time for the requested information and the scheme protection of information availability algorithm based on neural network system.

Key words: Distortion, Random Forest, F1-measure, InfoGain, entropy, neural network, sign-class.

1. INTRODUCTION

The development of Internet technologies has led to significant changes in the operation of network resources. At the present stage, a huge number of organizations require optimization of network space. Often, users do not understand how much the channel is “clogs up” with the irrational use of Internet connections. This includes watching online videos, online games, and many other factors of abuse in the workplace.

Therefore, an extraordinary requirement for the system administrator is the optimization of network resources, filtering and traffic analysis. Based on the criteria obtained, it is possible to determine what are the main problems that overload the corporate network:

- uncontrolled users downloading large files from the network;
- security problems caused by the lack of control over which sites the organization’s employees visit;
- irrational use of working time - online games, viewing entertainment resources on the Internet;
- an uncontrolled connection via VPN with production servers, which can lead to viruses on the corporate network.

In line with the statistics from the BrightCloud research center, in the absence of flexible filtering, the share of unnecessary and even dangerous sites in the total traffic of the corporate network is about 42% and only 36% of the resources are considered useful and relevant to the work.Controversial resources include 22% of sites visited. The leaders among the sites are social networks, videos and generating heavy traffic - flash banners.Due to the above problems, filtering issues come first in large companies. To provide security and rational management of corporate network traffic, it is necessary to properly manage network performance.

2. TECHNICAL METHODS FOR FILTERING NETWORK TRAFFIC

There are several ways to filter traffic at different levels of the TCP/IP stack. Each TCP/IP packet is characterized by 4 parameters: source IP address, destination IP address, source and destination ports. Knowing the source IP address, it can determine which of the users sent this packet, and knowing the IP address and destination port, you can understand who this packet is for and whether it is necessary to check this packet and the entire TCP/IP session [1]. By collecting and proxy the traffic of the required TCP/IP sessions, you can get additional information for filtering HTTPS requests, such as: request URL, domain and request body. For this information, you can use various filtering methods.

IP Blocking

When using this method, the server on which the unwanted material is located becomes completely inaccessible to the user. However, taking into account modern technologies, thousands of sites and other services, such as FTP or email, can be located on one IP address, so blocking it will result in all of them becoming inaccessible. Due to the low accuracy of this method, countries use it with caution.

DNS Distortion

When a user accesses any site, the computer sends a request to the DNS server in order to convert the domain name into an IP address. If this method is used, the DNS server returns an invalid address, and the site is inaccessible. Distorting a DNS record can also be implemented without the use of additional equipment. For example, China periodically deprives its users of access to

CNN International due to unwanted news appearing there. Although the purpose of filtering is to block only one page of the news, the remaining pages of the site also become inaccessible.

URL blocking

In the HTTP protocol, the URL contains the domain name of the site, as well as the request parameters. They can be checked with a list of blocked keywords, and if the user matches, the connection with the requested resource is broken, or it is redirected to the block page. This method is more efficient than blocking by IP address and distorting the DNS record, but requires additional equipment, as it uses surface analysis of packets. For example, in China, all requests containing the words “falun” and “gong” are blocked. URL blocking cannot be bypassed with conventional proxies - tools that encrypt traffic, such as VPN or TOR, are required.

File type lock

In response to the HTTP request, the server sets the Content-type header, which describes the type of content being transferred. The value of this header is one of the suitable MIME types. MIME is a standard for transmitting various types of data by e-mail, as well as a specification for encoding information and formatting messages. By comparing the values of the Content-Type header and the file types prohibited for transfer, you can filter requests

HTTPS request filtering

The headers and body of HTTP requests are transmitted in clear text, so the filter can select them and use them to check the site. However, it is not possible for HTTPS site pages to recognize request headers due to traffic encryption. Therefore, for HTTPS requests, filters perform a MITM attack and replace all or part of all site certificates in whole or in part. An important extension of the TLS protocol is the SNI Server Name Indication. With this extension, the domain is available in clear text. This extension is used to organize several HTTPS sites on the same IP address, but it also allows the filter to substitute certificates for specific sites only.

Packet filtering

The most complex and expensive method, since it requires the use of in-depth packet analysis. Currently fully implemented only in China. When using packet filtering, not only the headers of packets containing a URL are studied, but also all their contents. If there are forbidden words, the connection between the user and the server is broken. A significant drawback of this method is that the use of in-depth packet analysis can lead to a significant decrease in Internet connection speed, which, for example, is observed when accessing from China to foreign Internetservers.

Filtering through an HTTP proxy server

This method is most often used by organizations to connect corporate networks to the Internet, but it can be used to filter the Internet throughout the country. A hybrid variant called Cleanfeed is used effectively in the UK and Canada. Each user request is checked against a list of IP addresses containing prohibited content. If there are no matches, then the user request is sent directly. Otherwise, it is redirected to the proxy server of the Internet Watch Foundation.The proxy server receives the requested page and analyzes it. If the page does not contain prohibited materials, then the user gets access to it, otherwise - it appears that the resource is unavailable.Hybrid filtering options through an HTTP proxy server allow precise blocking of narrow categories of content at low cost. At the same time, they are as easy to manage as filtering by IP address.

Search Filtering

In a number of countries, such as China, France and Germany, search engines operating there are required to exclude links to prohibited materials from the search results. For example, in the French and German versions of Google, search results exclude links to groups and other materials prohibited by law. Therefore, users cannot find inappropriate content. Filtering search results is also one of the main methods to combat copyright infringement on the Internet. The method bypasses the use of other search engines - for example, the international version of Google does not exclude group sites from the results and is available from France and Germany.Table 1 describes the advantages and disadvantages of technical methods for filtering network traffic.

Table 1: Advantages and disadvantages of technical methods for filtering network traffic

Method	Advantages	Disadvantages
IP Blocking	Simplicity - it can be implemented using the basic network equipment used by Internet providers.	It can easily get around using various technical solutions, in particular, proxies and VPN.
DNS Distortion	High accuracy - only one site on the server becomes unavailable.	DNS Distortion records is easily bypassed by users - in the operating system settings.
URL blocking	Able to dynamically block new pages if their address contains forbidden words.	It can let in unwanted material or, conversely, allow excessive blocking.
File Type Lock	There is a connection between file blocking filters and the template used to create them, allowing changes to be applied.	Not a reliable means to block.
HTTPS request filtering	- data integrity and the impossibility of third-party intervention; - domain identification.	- when switching from HTTP to HTTPS, the address of all pages of the site changes; - at first, within weeks or a month after the move, the site will experience a decline in traffic from organics.

Packet filtering	Filter unwanted content not only in web pages, but also in all protocols - email, instant messaging services.	The use of in-depth packet analysis can lead to a significant decrease in Internet connection speed.
Filtering through an HTTP proxy server	Reconcile user request with a list of IP addresses.	Does not save from attack through plugins and XSS.
Search Filtering	Search can sort the results by content type.	Some search tools do not work for all languages, and some are only available after signing in to your Google account.

3. METHODS CONTENT FILTERING NETWORK TRAFFIC

3.1. Method for intercepting outgoing and incoming packets

Interception of outgoing packets. The proposed method for intercepting outgoing packets is to intercept a call to a kernel function that transmits a packet to the network and is part of the network device driver. For this, it is necessary to replace the original function of transferring the packet to the network with a specially developed one, whose tasks include analyzing the data transmitted in the packet.

If there is no information undesirable for transmission in the packet, the new function of transmitting the packet to the network calls the original function of transmitting the packet to the network and passes the socket buffer address to it. The address of the original function of transferring a packet to the network is stored in the system table of operations of the network device [2-3-4]. Unlike character and block devices, on Linux, network devices do not create files in the / dev directory. Instead, access to network devices is via network interfaces, a list of which can be displayed using the ifconfig. command). Otherwise, the function destroys the packet by passing its address to the function built into the kernel. Destroying a packet that was not transferred to the original function is necessary in order to avoid memory leak.

Interception of incoming packets. The network device is receiving packets, but there is no function in the network device operation table that receives packets. This is natural, since a request to send a packet to the network is a synchronous event, and a packet arriving from the network is an asynchronous event, processed upon interruption from the network device. The interrupt handler forms a socket buffer and puts it in the kernel queue of received *packets, passing its address to the function built into the kernel.*

3.2. Method for filtering unwanted Internet traffic applications using the Random Forest classification algorithm

The purpose of the method is to evaluate the performance of the Random Forest (RF) algorithm in classification problems for applications in the presence and absence of background filtering of network traffic.Methodology for solving the classification problem using the RF algorithmThe RF algorithm relies on the bagging technique - the use of a composition of independently trained algorithms. As a result, many decision trees are built, each of a separate random subset of the original data sample, and the size of the subsamples coincides with the size of the original sample and has repetitions.

Under implementing the method, a random forest was built and the classification quality was assessed on a given sample [5-6-7]. Empirically, the most acceptable algorithm parameters were selected. The forest consists of 5 trees with the greatest possible depth. Table 2 presents the error matrix for a clean test sample. Real values are indicated vertically, and horizontal predicted by the trained model.

Table 2: Error matrix for test sample

Realprediction	SSL	HTTP	DNS	BitTorrent	Steam	Skype
SSL	295	0	0	0	0	0
HTTP	0	267	0	4	1	0
DNS	0	0	266	1	0	0
BitTorrent	1	0	0	230	0	1
Steam	0	3	0	0	201	0
Skype	6	0	0	1	0	155

The following metrics are used to determine the effectiveness of the algorithm: accuracy, completeness, F1-measure, the values of which are easy to calculate based on the classification error matrix compiled separately for each class. Table 1.5 shows the error matrix for the test sample.

Table 3: Error matrix for test sample

	Real class: *X*	Real class: not*X*
Predicted class: *X*	TP	FP
Predicted class: no*X*	FN	TN

The matrix displays the number of correct and incorrect decisions for a given class.

TP (True Positive) stands for a true positive decision.

TN (True Negative) is a true negative solution.

FP (False Positive) - false positive.

FN (False Negative) is a false negative solution.

$\boldsymbol{Precision} = \frac{TP}{FP}$ – the proportion of correctly classified units of a given class relative to all instances that the algorithm assigned to this class. $\boldsymbol{Recall} = \frac{TP}{TP+FN}$ – the proportion of correctly classified units relative to all instances belonging to this class and the F1 measure calculated by the formula $\boldsymbol{F1} = \frac{Presicion \times Recall}{Presicion + Recall}$

A graphical representation of these metrics obtained experimentally for all analyzed classes is shown in Figure 1.

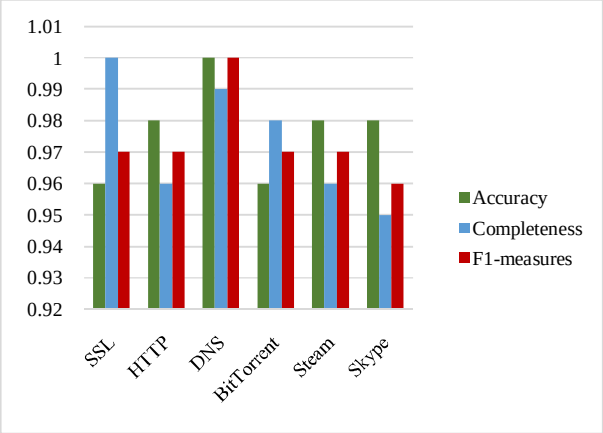

Figure 1: Accuracy, completeness and F1 measures chart for test sample

It can be seen that the algorithm is most efficient for data related to DNS traffic.Checking the operation of the algorithm on a test sample that has the same class composition as the training one, its quality was also evaluated in the presence of background traffic, i.e. when instances of classes that are not in the training set were present in the test set. The composition of this sample is given in Table 3.

Table 3: The composition of the test sample data with impurities

Protocol	Flowsnumber
SSL	295
HTTP	272
DNS	267
BitTorrent	232
Steam	204
Skype	162
LLMNR	169
Quic	95
RTP	19

This situation, when background traffic is present in the classified data, is closer to reality, because the protocols used on the Internet are very diverse. Such a DataSet allows you to get an assessment of the operation of the algorithm in real conditions. Table 4 presents the error matrix for this case.

Table 4. Error matrix for test sample in the presence of background traffic

Real prediction	SSL	HTTP	DNS	BitTorrent	Steam	Skype	LLMNR	Quic	RTP
SSL	295	0	0	0	0	0	0	0	0
HTTP	0	267	0	4	1	0	0	0	0
DNS	0	0	266	1	0	0	0	0	0
BitTorrent	1	0	0	230	0	1	0	0	0
Steam	0	3	0	0	201	0	0	0	0
Skype	6	0	0	1	0	155	0	0	0
LLMNR	169	0	0	0	0	0	0	0	0
Quic	0	0	25	9	0	61	0	0	0
RTP	0	0	0	0	0	19	0	0	0

As can be seen from the table, all the instances belonging to the LLMNR class were classified as SSL by the model, all RTP instances were assigned to Skype, and the Quic class instances were mainly divided between the DNS and Skype classes. Consider how the classification performance indicators presented in Figure 2 have changed.

As it can see, the presence of background traffic practically did not affect the completeness value, but significantly worsened the classification accuracy, since the number of false positive instances caused by the presence of background traffic belonging to classes that did not participate in training increased.

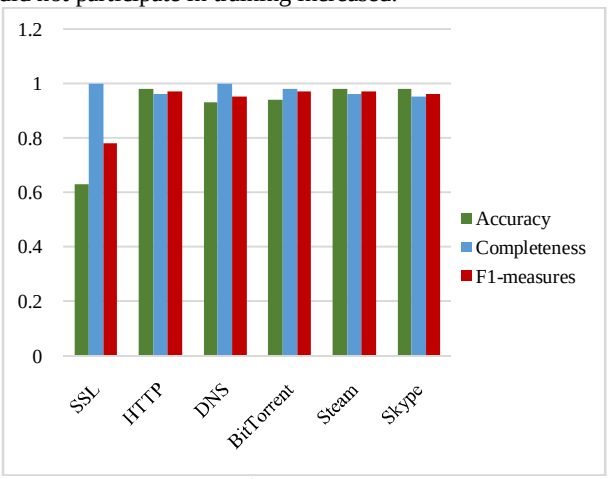

Figure 2: Diagramofaccuracy, completenessandF1-measuresinthepresenceofbackgroundtraffic

3.3. Attribute selection method based on network traffic filtering

A typical filtering-based attribute selection method is Sequential Forward Floating Selection (SFFS), which finds the best approximation for the number of functions selected. SFFS starts with an empty attribute pool and, using local optimal selection of attributes in two stages, increases the pool, including the inclusion stage and the conditional exclusion stage. The heuristic basis of the SFFS algorithm is the assumption that the selection criterion is monotonic with a change in size and information set. SFFS approximates the optimal solution at an affordable computing cost.

InfoGain Algorithm

The feature selection algorithm based on the information gain InfoGain is one of the simplest and fastest feature extraction algorithms. The algorithm is often used in solving the problem of text categorization, where the dimensionality of the data does not allow the use of more complex methods for selecting features [8-9-10]. The work of the method is based on calculating the entropy of the

class in question before and after applying the attribute.

So if $\boldsymbol{A}$ −this is a sign, and$\boldsymbol{C}$ −the class in question, then the entropy before observing the sign is estimated by the expression:

$$\boldsymbol{H(C) = -\sum_{c\in C} p(c) log_2 p(c)} \quad \textbf{(1)}$$

and after observation

$$\boldsymbol{H(C|A) = -\sum_{c\in C} p(a) \sum_{c\in C} p\,(c|a) log_2 p(c|a)} \quad \textbf{(2)}$$

The change in entropy due to the use of the attribute characterizes the information gain. To each attribute a from the set of attributes$\boldsymbol{A}$an assessment is assigned based on the informational gain between himself and the class:

$$\boldsymbol{IG_i = H(C) - H(C|A_i) = H(A_i) - H(A_i|C) = H(A_i) + H(C) - H(A_i, C)} \quad \textbf{(3)}$$

The result of the InfoGain algorithm is the ranking of signs according to their importance.

Using a correlation measure allows you to optimize the selection of features. As a result, the method focuses on two problems: criteria for a correlation measure and an algorithm for selecting features. As a criterion for the correlation measure, the Pearson correlation coefficient, the mutual information criterion, and other relevant criteria can be used.A typical feature selection algorithm using a correlation measure is the CFS (Correlation-based Feature Selection) algorithm, a feature selection algorithm based on correlation.The CFS algorithm is one of the first that evaluates many features, rather than each feature individually. The algorithm is based on an assessment of a set of attributes, taking into account the utility of each independent attribute in the definition of a class, and the correlation between them:

$$\boldsymbol{Merit_s = \frac{\overline{kr}_{cf}}{\sqrt{k + K(k-1)\overline{r}_{ff}}},} \quad \textbf{(4)}$$

where

$\boldsymbol{Merit_s}$ −subset quality assessment$\boldsymbol{S}$containing$\boldsymbol{k}$signs;

$\overline{\boldsymbol{r}_{cf}}$ −average correlation «sign-class»;

$\overline{\boldsymbol{r}_{ff}}$ −average correlation between features of a given subset.

The numerator of expression (4) is the quality metric of a given subset of features, and the denominator is how much redundant information it contains.As a result, “bad” or valueless traits will be discarded due to poor quality assessment in the given subset, and redundant traits due to high correlation with one or more traits in the subset.

To apply estimate (5), correlation calculations, or dependencies between attributes, should be performed:

$$\boldsymbol{SU = 2.0 \times \left[\frac{H(X) + H(Y) + H(X,Y)}{H(X + H(Y))}\right]} \quad \textbf{(5)}$$

After calculating the correlation matrix, CFS uses a heuristic search to find a good subset of features.

3.4. Method for filtering HTTP packets based on subsequent analysis of requests to a web resource

The proposed filtering method consists in the fact that the processes of query execution and its verification by the filter occur in parallel. Figure 3 shows the time diagram of the passage of the user's request through the filtering device with the subsequent analysis of the request to the resource.

The request successfully passes through the filter (a) or is blocked by the filter by breaking the TCP connection (b). In a filter that uses the subsequent analysis of HTTP requests, all packets arriving at the device’s input, including those containing a user’s request, are always passed to the device’s output without delay and modification, and a copy of the missing HTTP protocol packets is created to analyze the request. In filtering devices that use preliminary analysis of HTTP requests, the request received at the input of the filter is checked before it is sent to the Internet. The request, which is recognized as allowed during the scan, is passed to the filter output and then to the web server with the requested resource. The filter, which uses post-analysis of HTTP requests, checks a copy of the received request after it is sent to the Internet [11]. The check takes place while the request on the communication lines reaches the web server on which the requested resource is located, a response is generated from the server, which returns to the filter, i.e. in the mode of subsequent analysis of the past request. Based on the test results, the response received from the web server is either skipped to the user (Fig.3, a) or blocked (Fig.3, b).

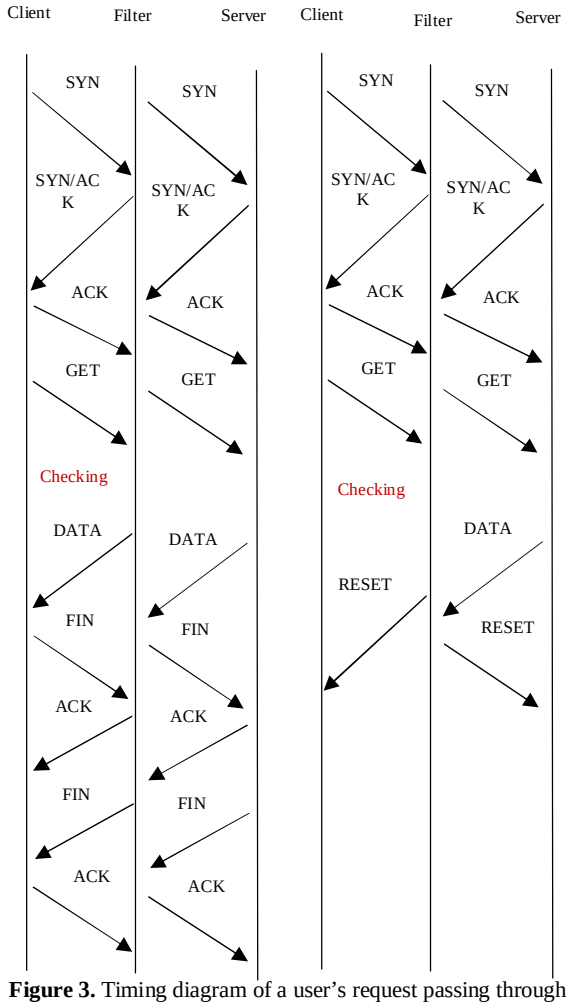

Figure 3. Timing diagram of a user’s request passing through a filtering device with subsequent analysis of a resource request

Device model and operation algorithm

The presented filtering method can be illustrated by the example of a simplified model of a packet filter. Figure 4 shows a model of a filtering device that implements a method for subsequent analysis of requests to a web resource.

The filter model consists of Network user interface (NUI), Internet Interface (II), two selectors (S1 and S2), corrector analyzer (CA), storage of the current monitored TCP session states (MSS) and Forbidden Resource Identifier stores (FRI).

Selectors extract HTTP protocol packets from the general traffic, with S1 passing all traffic coming from the NUI into filter II immediately and without changes, and sending copies of HTTP packets to the CA. S2 passes all traffic coming from II to NUI, with the exception of HTTP packets that are sent to the CA for inspection. CA, using information from FRI, checks requests for access to the requested resource and, if necessary, blocks the user from receiving a response from a web server with prohibited content.

A session formation and analysis unit (SAU) is built into the CA, which from network packets of the HTTP protocol forms TCP sessions, in which users request certain web resources, stores information about these sessions in MSS and, upon request, provides the request status: is it permitted or not. The model operates as follows. The packet flow with the user's request, reaching the filtering device on the way to the web server, enters the NUI user network interface, and from it into the first selector.

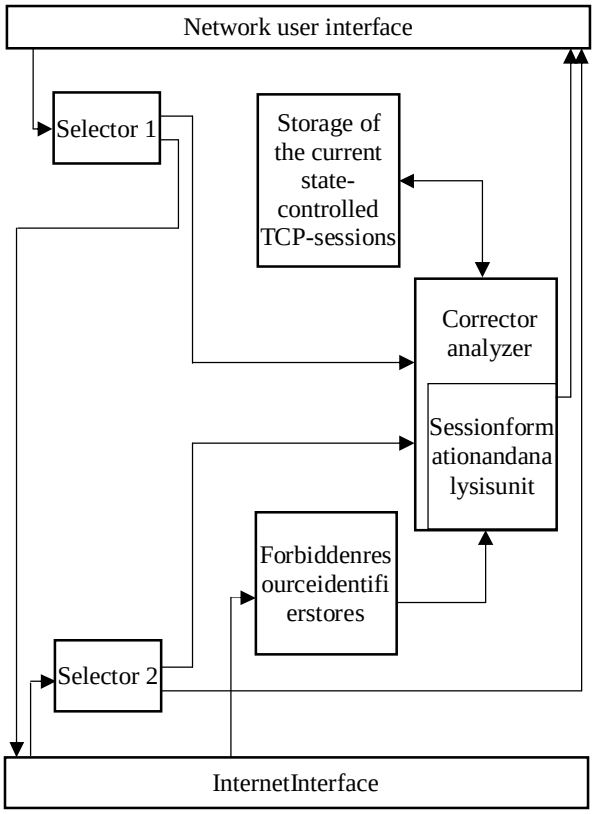

Figure 4: A model of a filtering device that implements a method for subsequent analysis of requests to a web resource

S1 sends the packets constituting the request to the Internet II interface, and copies of these packets to the corrector analyzer. A TCP session is formed in it, the URL resource identifier is extracted from the request, and access rights to this resource are checked. Thus, verification of a copy of a user's request occurs in parallel with the transportation of the original user request from the filtering device to the web server and the response from the web server to the user to the filtering device [12-13-14]. The response from the web server, reaching the filtering device, enters the user network interface and then to the second selector. S2 separates packets of TCP sessions of the HTTP protocol and passes them to the corrector analyzer for subsequent processing. SAU, having requested MSS, determines for the packet the corresponding TCP session from the list of monitored TCP sessions. It then checks if the current request is allowed for this TCP session. If the request is allowed, the packet is sent to the user network interface without changes. Otherwise, actions related to a specific response blocking algorithm are performed.

The gain in the time the user request travels through the filtering device when using subsequent analysis compared to the preliminary analysis is the time spent on determining the TCP session for each packet, generating the user request from the packets, extracting the identifier of the requested resource URL and checking the request for access to the requested a resource using internal forbidden URL lists.

3.5. Method for filtering the contents of a network packet using and training a neural network during DDOS attacks

The purpose of the method is to develop an algorithm for protecting the availability of information by filtering the contents of a network packet based on the use and training of a neural network. The developed neural system is shown in Figure 5.

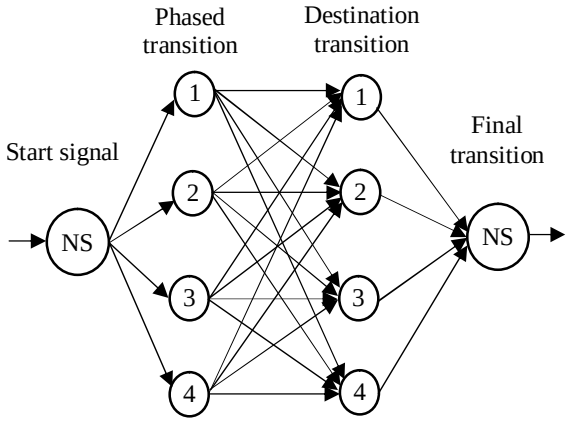

Figure 5: Neural network for protection against DDOS attacks

Under the initial signal is supplied, the algorithm is activated: analysis of network traffic, followed by sending data to the MySQL database. This allows you to take further measures to protect data availability, based on the obtained values for the volume of incoming traffic.Next,

there is a “phased transition" to the connection of the firewall to send filtering rules. By “destination transition” is meant the distribution of network traffic across physical and logical cores to reduce the load on server resources, with the subsequent application of filtering rules. This allows you to increase the performance of the physical server, with the subsequent reflection of DDOS attacks.The general scheme of the algorithm for protecting the availability of web server information is presented in Figure 6.

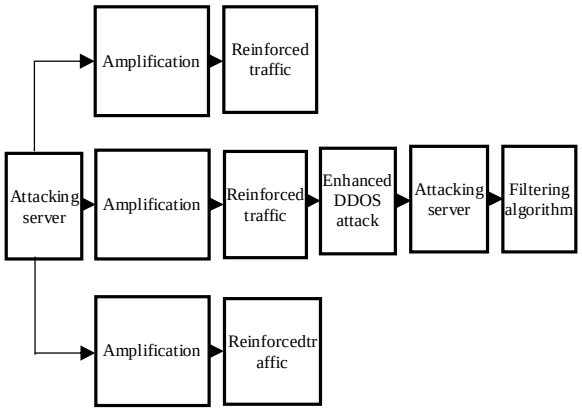

Figure 6. The scheme of operation of the algorithm for protecting the availability of information

Testing of the developed algorithm is presented in Table 5. In the table: designations 1.00/2.00 - activated algorithm/without algorithm.

Table 5: Testing the developed algorithm

Day	Attack, Gb/s	Load onCPU, %	ConsumptionRAM, %	Load onSSD, %
1.	1,00	1,00/2,00	0,12/0,24	0,10/0,20
2.	2,00	2,00/4,00	0,13/0,26	0,20/0,40
3.	3,00	3,00/6,00	0,14/0,28	0,30/0,60
4.	4,00	4,00/8,00	0,15/0,30	0,40/0,80
5.	5,00	5,00/10,00	0,16/0,32	0,50/1,00

According to the above results (Table 5), it can notice an increase in performance with the activated algorithm. The load on the central processor decreased by 2 times. The use of RAM is halved. SSD utilization decreased by 50%. This is due to the impossibility of filtering DDOS attacks on the physical server of layer 4-7 levels using standard methods [15-16-17].Thus, the developed algorithm for protecting the availability of information through the use and training of a neural network can improve server performance, and also contributes to the stable operation of the server during an attack by malicious traffic.

Content filtering network traffic are evaluated according to the criteria of speed and accuracy.

1. The classification speed reflects how quickly the classification module determines the category of the analyzed document.
2. The decision-making speed reflects how long it takes on average to decide whether to allow or deny access to a resource.
3. The knowledge base speed - the time to save to the knowledge base.
4. The accuracy of the method - is carried out according to the value of the steady-state error in the implementation of the method in the form of an algorithm or software.

Table 6 shows the evaluates of methods content filtering network traffic according to the above criteria.

Table 6. Evaluation of methods content filtering network trafficby the criteria of speed and accuracy

№	Name of the method/ Criteria	The classification speed	The decision-making speed	The knowledge base speed	The accuracy of the method
1.	Method for intercepting outgoing and incoming packets	High	Average	Average	Average
2.	Method for filtering unwanted Internet traffic applications using the Random Forest classification algorithm	Average	High	High	High
3.	Attribute selection method based on network traffic filtering	High	High	High	Average
4.	Method for filtering HTTP packets based on subsequent analysis of requests to a web resource	Average	High	Average	High
5.	Method for filtering the contents of a network packet using and training a neural network during DDOS attacks	High	Average	Average	High

Here: $\boldsymbol{n}-$ number of rows; $\boldsymbol{b_i}-$ a general argument of evaluation. The following arguments were accepted for evaluation:

$\boldsymbol{b_1} = \mathbf{3}$ (low); $\boldsymbol{b_2} = \mathbf{4}$ (average); $\boldsymbol{b_3} = \mathbf{5}$ (high).

Based on Table 6, the following $\boldsymbol{B}$ matrix is formed:

$$\boldsymbol{B} = \begin{pmatrix} \mathbf{5} & \mathbf{4} & \mathbf{4} & \mathbf{4} \\ \mathbf{4} & \mathbf{5} & \mathbf{5} & \mathbf{5} \\ \mathbf{5} & \mathbf{5} & \mathbf{5} & \mathbf{4} \\ \mathbf{4} & \mathbf{5} & \mathbf{4} & \mathbf{5} \\ \mathbf{5} & \mathbf{4} & \mathbf{4} & \mathbf{5} \end{pmatrix}$$

Above the following results can be obtained from the $\boldsymbol{B}$ matrix:

1. Total evaluation sum of method for intercepting outgoing and incoming packets:

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{4} b_i = 17$$

2. Total evaluation sum of method for filtering unwanted Internet traffic applications using the Random Forest classification algorithm:

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{4} b_i = 19$$

3. Total evaluation sum of attribute selection method based on network traffic filtering:

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{4} b_i = 19$$

4. Total evaluation sum of method for filtering HTTP packets based on subsequent analysis of requests to a web resource:

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{4} b_i = 18$$

5. Total evaluation sum of method for filtering the contents of a network packet using and training a neural network during DDOS attacks:

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{4} b_i = 18$$

In Figure 7 is shown the diagram of evaluation of methods content filtering network traffic by the criteria of speed and accuracy.

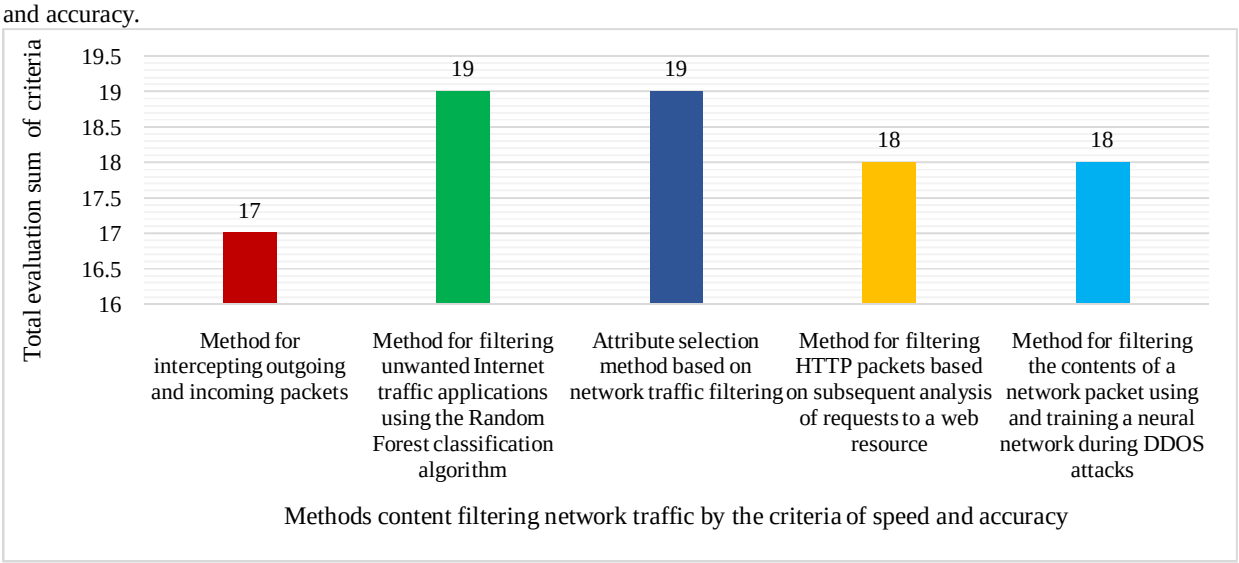

Figure 7: The diagram of evaluation of methods content filtering network traffic by the criteria of speed and accuracy

4. CONCLUSION

In conclusion, under analysis of methods content filtering network traffic were revealed, Method for filtering unwanted Internet traffic applications using the Random Forest classification algorithm and Attribute selection method based on network traffic filtering are more stability and reliability in line with all criterion. Established that method for filtering unwanted Internet traffic applications using the Random Forest classification algorithm is suitable for filtering network traffic in real time due to the time complexity of processing, estimated by the ratio. And attribute selection method based on network traffic filtering is used to solve the problem of text categorization, where the dimensionality of information does not allow the use of complex methods for distinguishing features.

REFERENCES

1. ZhivoyZhurnal. **Internet Tsenzor** [Elektronnyyresurs]. - Rezhimdostupa: http://icensor.livejournal.com/ (data obrashcheniya: 22.05.2018).
2. Chemodurov A.S., KarputinaA.Yu. **Internet gateway protection and filtering network traffic corporate network**. Nauchno-metodicheskijelektronnyjzhurnal «Kontsept» [Scientific and methodological electronic journal «Concept»], 2015, no. 1, pp. 96-100. (in Russian).
3. Rosen R. **Linux Kernel networking: Implementation and theory**. N.Y.: Apress, 2014. 612 p. https://doi.org/10.1007/978-1-4302-6197-1
4. Corbet J., Rubini A., Kroah-Hartman G. **Linux devices drivers**. 3rd ed. Sebastopol: O'Reilly Media Inc., 2005. 615 p.
5. Sheluhin O.I., Simonyan A.G., Vanyushina A.V. (2017). **Benchmark data formation and software analysis for classification of traffic applications using machine learning methods**. // T-Comm, vol. 11, no.1, pp. 67-72.
6. En-Najjary T, Urvoy-Keller G., Pietrzyk M., and Costeux J.-L.**Application-based feature selection for internet traffic classification**. In Teletraffic Congress (ITC), 2010 22nd International, pages 1 - 8, 2010. https://doi.org/10.1109/ITC.2010.5608734
7. Pietrzyk M., En-Najjary T., Urvoy-Keller G., and Costeux J.-L. **Hybrid traffic identify cation**. Technical Report EURECOM+3075, InstitutEurecom, France, 04 2010.
8. Andrew Moore, Denis Zuev, and Michael Crogan.

(2005). **Discriminators for Use in Flow-Based Classification**. Technical Report RR-05-13, Department of Computer Science, Queen Mary, University of London.

9. Thuy T.T. Nguyen and Grenville Armitage. (2008). **A Survey of Techniques for Internet Traffic Classification using Machine Learning**. IEEE Communications Surveys & Tutorials, vol. 10, no. 4, pp. 56-76, IEEE Press, Piscataway, New Jersey, USA.
10. Szabo G., Orincsay D., Malomsoky S., Szabo I. (2008). **On the validation of traffic classification algorithms**. Proceedings of the 9th International Passive and Active Measurement conference, April 29-30, 2008, pp. 72-81. https://doi.org/10.1007/978-3-540-79232-1_8
11. Jai Balasubrahmaniyan, KuntalDaftary, Venkateswara Rao Yarlagadda, Krishna Kumar. **System and method for URL filtering in a firewall** // Patent US 20060064469A1, Int. Cl.G06F 15/16 (2006.01), Publ. Date: Mar. 23, 2006.
12. HegazyZaher, H. A. Khalifa, Abeer Ahmed. **Fuzzy Max Plus Algebra Algorithm for Traffic Problems**. International Journal of Emerging Trends in Engineering Research. Volume 7, No. 11 November 2019 https://doi.org/10.30534/ijeter/2019/217112019
13. Bloch E., Mohan Sh., Pagaku R. R. et al. **Apparatus for monitoring network traffic** // Patent US 7849502 B1, Int Cl G06F 15/16 (2006.01), G06F 11/00 (2006.01). Publ. Date: Dec. 7, 2010
14. M.V.D Prasad, Syed Inthiyaz, M.Teja Kiran Kumar, K.H.S.Sharma, M. Gopi Manohar, RupaKumari, SkHasaneAhammad. **Human activity recognition using Deep Learning**. International Journal of Emerging Trends in Engineering Research. Volume 7, No. 11 November 2019 https://doi.org/10.30534/ijeter/2019/227112019
15. Peng Liu. **Denial of Service Attacks**. School of Information Sciences and Technology. University Park, 2004.
16. Zhang, M. **Study on modeling and simulation of DDoS active defense** / M. Zhang, X. Liu, J. Tang, H. Kong // Xitongfangzhenxuebao. Т. 26. №11. Изд-во: «ZhongguoXitongFangzhenXuehui», 2014. – С. 2698-2703.
17. O.S. Eluyode, DipoTheophilusAkomolafe, "**Comparative Study of Biological and Artificial Neural Networks**", European Journal of Applied Engineering and Scientific Research, 2013, 2(1):36-46