PROJECT

Multi-tenant cybersecurity operations centre designed as an open cloud service with machine learning elements

V17 Test results and evaluation

KPB5 – Implementation of the solution

type – report

month of submission – M24

Funded by the European Union
NextGenerationEU

RECOVERY AND RESILIENCE PLAN

GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

Contents

1. Machine learning and artificial intelligence in cybersecurity 3
1.1. Detection of DoS/DDoS attacks 3
1.1.1. DoS (Denial of Service) 3
1.1.2. DDoS (Distributed Denial of Service) 4
1.1.3. Analysis of statistical parameters on selected DDoS attacks 5
1.1.4. Variability, skewness and kurtosis 7
1.1.5. Detection values for autocorrelation and regression 10
1.1.6. Response of divergence to a DDoS attack 12
1.1.7. Use of the 3σ-Tunnel for recognizing DDOS attacks 12
1.1.8. Detection functions for machine recognition of DDoS attacks 14
1.1.9. Evaluation of tests 16
1.2. Detection of unwanted e-mail messages (Spam) 17
1.3. Detection of malicious code (Malware) 17
1.4. Phising detection 17
1.5. Detection of advanced persistent threats (APT) 18

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

1. MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

With the advent of the information age, cybersecurity has become a pressing issue in today's society and a skill that is in high demand in industry. Businesses, governments and individual users are all exposed to the risk of security attacks and breaches. The fundamental goal of cybersecurity is to protect users and their data. Cybersecurity is a multifaceted problem, ranging from highly technical areas (cryptography and network attacks) to user-facing areas (detection of hate speech or fraudulent credit card transactions). It helps prevent damage to sensitive information, prevent financial fraud and losses, and protect users and their devices from malicious actors. A large part of cybersecurity analytical processes, investigation and detection is currently driven by machine learning ML (Machine Learning) and "intelligent" systems.

In the following text, we describe the test results for the following defined security threats:

- DoS/DDoS attacks
- spam
- malware
- phishing
- advanced persistent threats APT

1.1. Detection of DoS/DDoS attacks

"Denial-of-service" (DoS) and "distributed denial-of-service" (DDoS) attacks aim to disrupt the normal operation of a target server, service or network by overwhelming it with a flood of internet traffic. DoS attacks usually originate from a single source, while DDoS attacks are launched from multiple compromised devices, often dispersed globally. This multitude of sources makes DDoS attacks harder to stop, and they can have a more significant impact.

These attacks work by overloading the target with more requests than it can handle, causing the system to slow down significantly or crash completely and denying service to legitimate users. Common methods include exploiting server vulnerabilities or using a network of bots, known as a botnet, to flood the target with traffic. The motivations for these attacks can range from simple vandalism and personal grudges to extortion and political activism.

1.1.1. DoS (Denial of Service)

The abbreviation DoS translates as an attack denying a service. This type of attack is used to overwhelm a server. In this case, the attacker uses a device with which they attempt to disrupt the normal operation of the target server, service or network. They very quickly became a cybersecurity threat. They are simple, but their impact can be devastating for individuals as well as companies.

DoS attacks are often used to shut down devices and networks, and there are many different ways of using the above DoS attacks. These include, for example:

- Buffer overflow: A buffer is temporary storage for storing data. If a program or system process places more data than was predetermined, the additional data then overflow and some of them may then leak into other buffers. This then

results in the corruption or overwriting of other data. Based on this principle, an attacker can corrupt files and alter or disclose various information.

- ICMP attack (ICMP attack or Ping of Death): This type of attack uses the ICMP protocol, employing ICMP echo requests, so-called pings. In general, echo requests and echo replies are used to ping a network device in order to verify the connectivity between the device and the sender. If an attacker flooded the target with request packets, the network must respond with the same number of packets, which makes the device unavailable for other incoming requests.
- UDP attack (UDP attack): The attack is based on sending UDP packets with a source IP address that the attacker can spoof, the aim being to flood random ports on the target server. The advantage of the UDP protocol is that it does not need to establish a connection, as is the case with the TCP protocol. If the target server receives a UDP packet on a port that is not used by any application, it informs the sender about this. If the attacker sends a sufficient number of packets to such a port, it may become overloaded.
- SYN attack (SYN flood): This is an attack that arises when a TCP connection is established between 2 devices. TCP uses a three-way handshake. Establishing a connection begins with sending a SYN message. If the server is active, a SYN/ACK message is sent back to the client that wanted to establish the connection. The server tries to respond until the time for establishing communication expires, while keeping the data in memory. If the client is an attacker with a spoofed and unreachable source address, sending a large number of such packets can exhaust the memory allocated for TCP connections, and all further communication attempts will then be ignored.
- Teardrop attack (Teardrop attack): During this attack, the attacker tries to send fragmented packets, and the network then tries to compile these data back into their original state. The attacker can exploit a vulnerability of an outdated operating system. It is common for many companies to still use older versions of systems in order to keep running an old application, and since some versions of operating systems, such as Linux or Windows, contain a bug in reassembling a fragmented packet, the server may be attacked precisely because of this vulnerability.

1.1.2. DDoS (Distributed Denial of Service)

Unlike a simple DoS attack, a DDoS (Distributed Denial of Service) attack gains strength because the attacker uses multiple systems at once, thereby increasing the severity of the targeted attacks. The attacker may use malware that they send over the network to devices with the intention of infecting them. They thus become part of a so-called botnet and, following the attacker's commands via a command-and-control server, these bots then attack the targeted server in unison. DDoS attacks are very difficult to detect because the attacking bots may be located in different places around the world. The main and key difference between a DoS and a DDoS attack is that with DDoS, using bots, the attacker can send huge volumes of traffic.

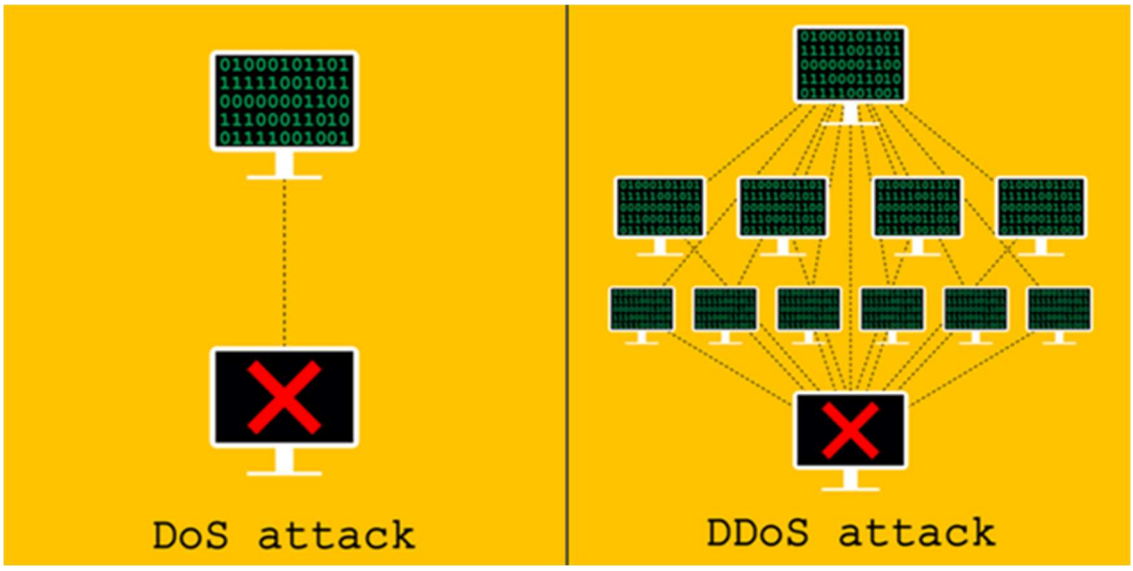


Fig. 1: DoS and DDos attacks

DDoS attacks are divided into 3 main categories:

1. Volumetric attack: This type exploits vulnerable services such as NTP, DNS or SSDP. The attacker tries to flood the target with a large volume of packets to which the target tries to respond, until the targeted network infrastructure collapses. It is usually a broadband attack that can exceed 100 GB or even several TB per second and is immediately obvious to connectivity providers. Network monitoring, flow analysis and flow telemetry analysis have proved effective for detecting volumetric attacks.
2. Protocol attack: This type of attack tries to exhaust the resources of the server or its systems and relies on vulnerabilities of internet communication protocols. Protocol attacks include, for example, the SYN attack, UDP attack and ICMP attack. These attacks are considered sufficiently crippling for the targeted infrastructure owing to the frequency of their duration. In 2018, attackers used the BGP protocol to redirect traffic intended for the “MyEtherWallet” service, which, however, was merely a cover for the theft of cryptocurrency wallets. It lasted approximately 2 hours.
3. Application layer attack: This attack targets a specific application and its vulnerabilities so that it is unable to process, send or receive data. The attacks are weak and slow and concentrate on GET and POST requests. The aim is then to cripple the web server, and they are measured in requests per second (Rps). They are low and weak so that they comply with the protocol.

There are several symptoms that can help us detect whether our network, application or the entire infrastructure is under a DDoS attack. This may be a sudden increase in web activity on the server, reduced network performance or a sudden inability to communicate with the application.

1.1.3. Analysis of statistical parameters on selected DDoS attacks

We will present the response of selected statistical parameters on various records of DDoS attacks. We have omitted a number of experiments that we carried out on various simulated scenarios. On these scenarios, the selected statistical parameters worked “as in a textbook”, and we gained an initial idea of the effectiveness of using the individual parameters to recognize a change and an increase in the simulated traffic. When deployed to recognize

a real attack, the responses of the statistical parameters were no longer so unambiguous. For the presentation, we selected eight different records of DDoS attacks.

The first four records represent standard DDoS attacks (N – normal), in which the parameters behave similarly to the simulated data. The next two records also represent standard attacks at first sight, but the parameters behave in a substantially different, indeterminate manner in them (S – special). The seventh record is a representative of so-called Low Rate (LR) DDoS attacks, i.e. attacks with a low average flow intensity. The last record is an attack which, in our opinion, cannot be detected by the above statistical parameters (P – as in problem). For the records, together with the number of packet increments a(t), we also displayed their moving average m(t) with a compute window cw = 1024 ts.

The records are sampled into given time windows (ts – time slot), mostly ts = 1ms. For the step-by-step calculation of the parameters, we used a shift of the so-called compute window (cw) by one ts. When conducting the experiments, we used various powers of two (256,512,1024,2048,4096) for the compute window sizes because of the estimation of the Hurst exponent. After a subjective evaluation of the course of the values of the parameters examined, we decided to focus further mainly on a compute window of size cw = 1024 ts. With cw = 512 ts, significant “undulation” occurred in some parameters; with cw = 2048, the time needed to calculate the parameters, especially the Hurst exponent, increased significantly.

In the graphs of the DDoS attack records, we have also displayed the course of the moving average and shown the time interval in which the moving average m(t) reached its local maximum from the start of the attack. We will use this interval in the subjective evaluation of the other statistical coefficients. In all records of DDoS attacks, we can observe a linear growth of the moving average. We are looking for coefficients whose values start to grow faster than linearly at the start of the attack.

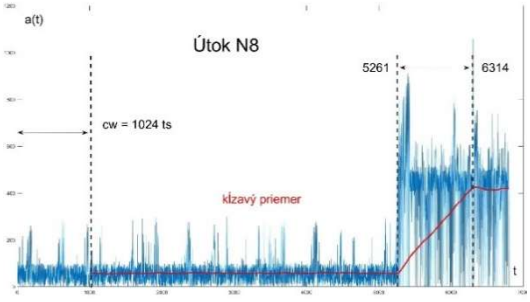


Fig. 2: Attack N8 and moving average

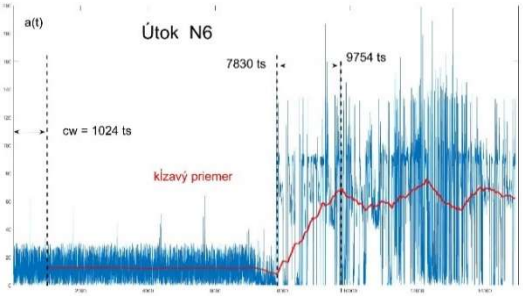


Fig. 3: Attack N6 and moving average

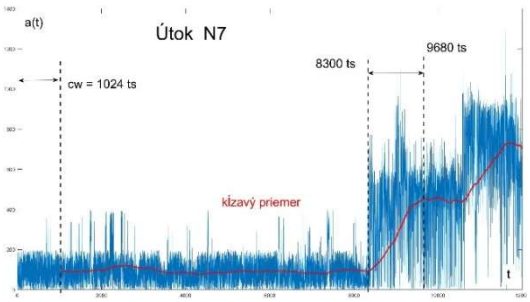


Fig. 4: Attack N7 and moving average

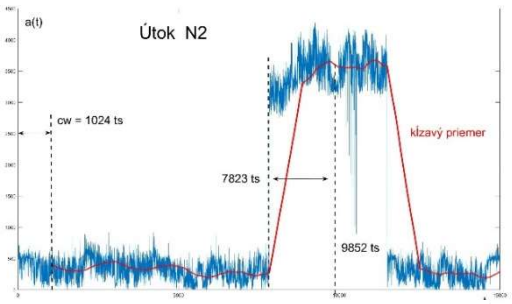


Fig. 5: Attack N2 and moving average

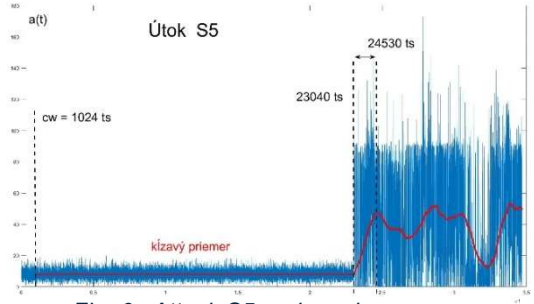


Fig. 6: Attack S5 and moving average

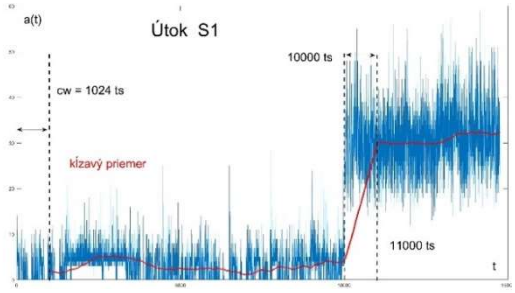


Fig. 7: Attack S1 and moving average

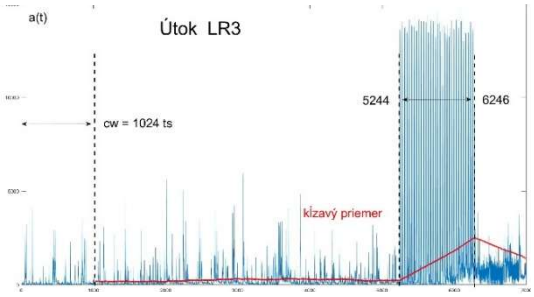


Fig. 8: Attack LR3 and moving average

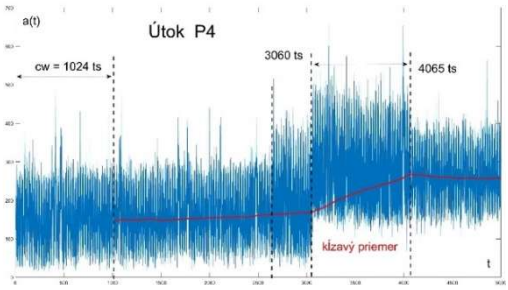


Fig. 9: Attack P4 and moving average

1.1.4. Variability, skewness and kurtosis

First, we will deal with the basic statistical moments, or rather the coefficients of variation V(t), skewness K(t) and kurtosis Skw(t). The coefficient of variation is the ratio of the standard deviation to the mean value of the increments in the given time window cw. Variability expresses the degree of dispersion of the increments a(i) around the mean value. In the first graph we show the course of the moving average and the standard deviation; the second graph shows the values of the coefficient of variation:

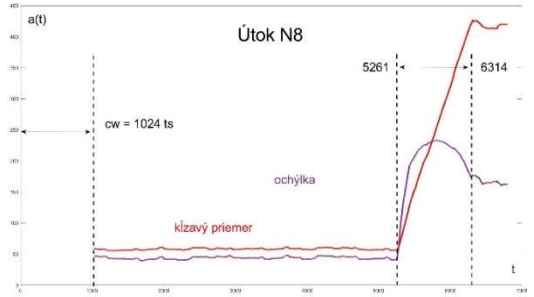


Fig. 10: Attack N8, mean and deviation

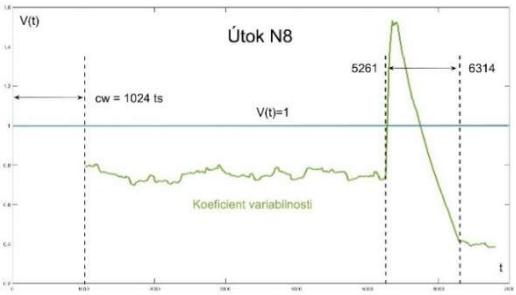


Fig. 11: Attack N8, coefficient of variation

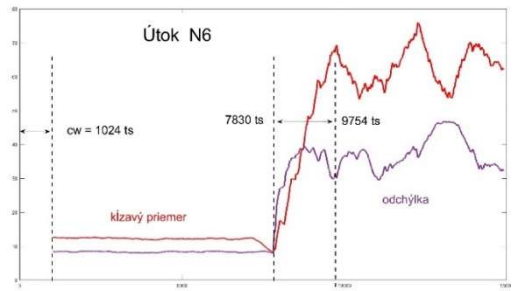


Fig. 12: Attack N6, mean and deviation

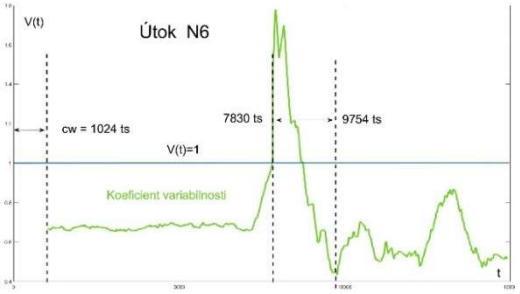


Fig. 13: Attack N6, coefficient of variation

In the case of standard DDoS attacks (N2,N6,N7,N8), the coefficient of variation responded in the same way. Standard traffic represented a stationary flow with a deviation smaller than the mean flow intensity. At the moment the attack began, the deviation responded with a faster increase in values than the linear growth of the moving average, so the ratio of the deviation to the mean reached the value V(tA) = 1 at the moment of the attack. This fact could be used for fast and simple detection of this type of attack. However, this is not a rule; in Fig. 14, for attack S1, we can see that during standard traffic stationarity is disrupted and the coefficient exceeded the value V(tA) = 1 several times before the attack.

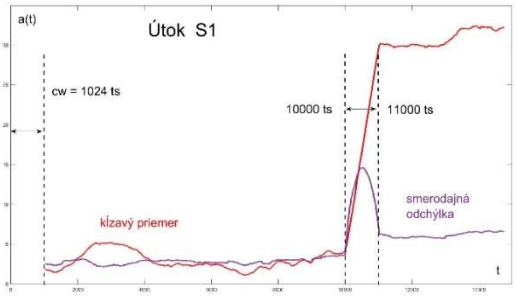


Fig. 14: Attack S1, mean and deviation



Fig. 15: Attack S1, coefficient of variation

Figs. 16 and 17 show graphs of the mean, deviation and coefficient of variation for the so-called Low Rate DDoS attack LR3.

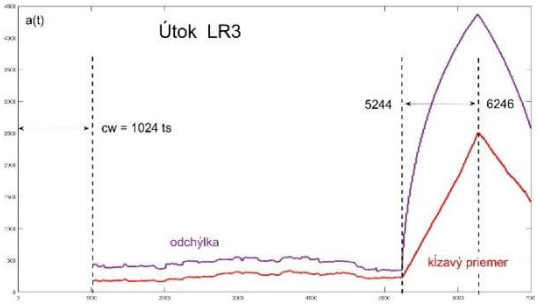


Fig. 16: Attack LR3, mean and deviation

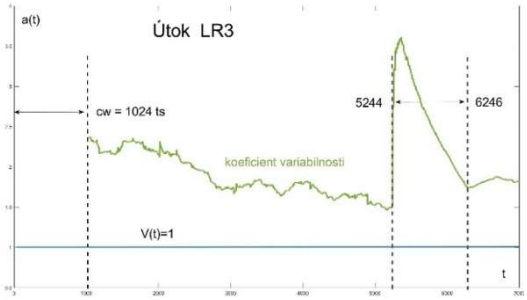


Fig. 17: Attack LR3, coefficient of variation

In the case of LR3, the deviation during normal traffic is greater than the mean flow intensity, so the coefficient of variation is V(t)>1 throughout. The value 1 cannot now be used for fast detection. However, at the moment of the attack, the deviation values again increased faster than the linear growth of the mean intensity, which manifested itself in a marked step increase in the coefficient of variation. It should be possible to recognize this change using prediction methods.

Other simple statistical parameters are the coefficient of skewness and the coefficient of kurtosis. These are in fact the third and fourth central moments scaled by the standard deviation.

$$\mu_3 = \frac{1}{\sigma^3} E[X - EX]^3 \qquad \mu_4 = \frac{1}{\sigma^4} E[X - EX]^4 \qquad \sigma^2 = E[X - EX]^2$$

In our case, the realizations of the variable X are the values of the increments a(t), and the coefficients describe the properties of the probability distribution of the increments in the given compute window cw. As the attack traffic was progressively loaded into the evaluated

window cw, we assumed a marked change in the probability distribution and consequently also a change in the coefficients. This assumption was confirmed for almost all the flows analysed. Both statistical coefficients responded with a marked increase in their values right at the start of the DDoS attack.

First, we present graphs of the coefficients applied to two selected standard attacks, followed by some problematic situations.

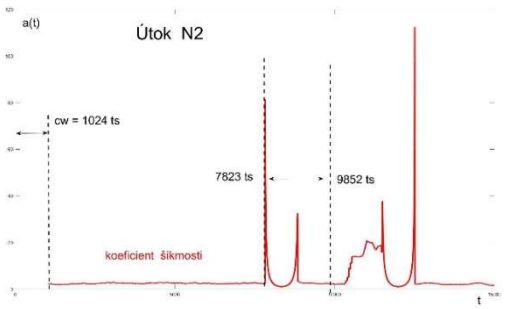


Fig. 18: Attack N2, coefficient of skewnes

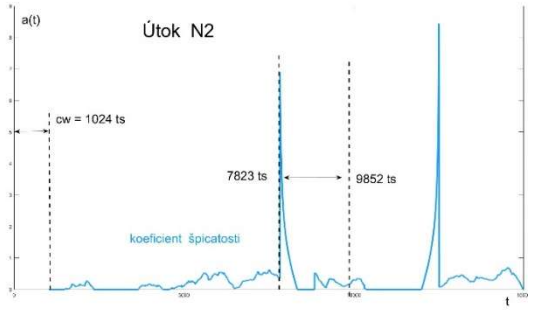


Fig. 19: Attack N2, coefficient of kurtosis

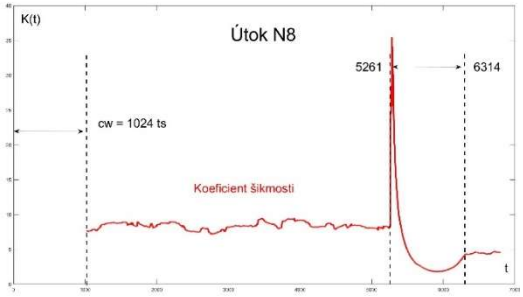


Fig. 20: Attack N8, coefficient of skewne

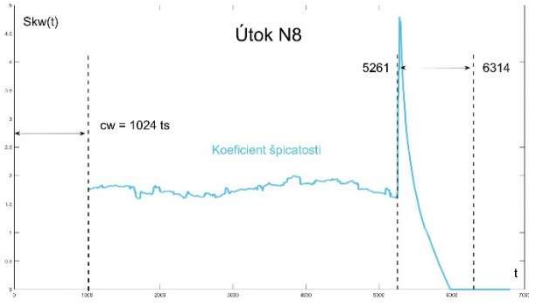


Fig. 21: Attack N8, coefficient of kurtosis

In standard type N attacks, the coefficients of skewness and kurtosis fluctuate only slightly during calculation on standard traffic owing to stationarity, and at the moment of the onset of the DDoS attack their values increase markedly. We assume that we will be able to recognize this marked jump automatically using suitable prediction methods.

A considerably less favourable situation arose with non-standard attacks, e.g. S1. Owing to the non-stationarity of normal traffic and the frequent occurrence of high peaks, both coefficients oscillate strongly, as a result of which the jump at the onset of the attack is lost in the overall course of the values of both coefficients. In such a case, the onset of the DDoS attack cannot be recognized by machine.

To visualize the response of the coefficients to a change in the character of normal traffic and to the occurrence of peaks, we displayed the values of the coefficients in a common graph together with the increments of flow S1.

Funded by the European Union
NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

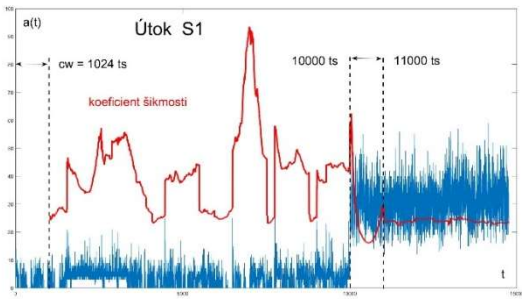


Fig. 22: Attack S1, coefficient of skewness

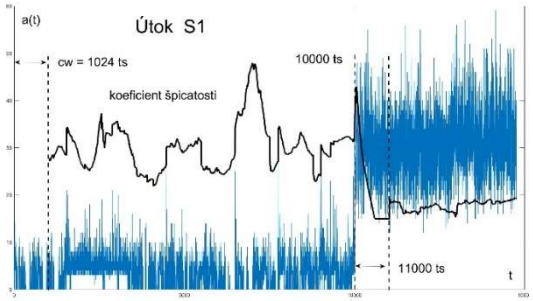


Fig. 23: Attack S1, coefficient of kurtosis

The response of the coefficients to the attack with a low mean intensity LR3 turned out as positively as for the standard N-type attacks.

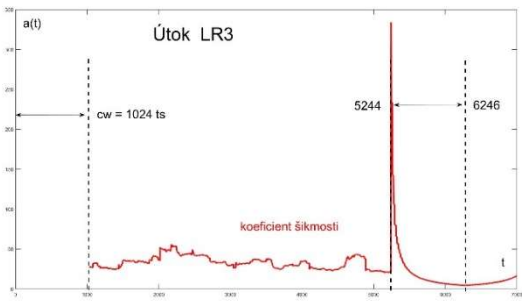


Fig. 24: Attack LR3, coefficient of skewness

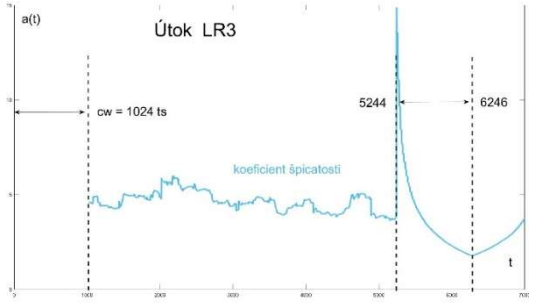


Fig. 25: Attack LR3, coefficient of kurtosis

Again, we can see the possibility of automatically distinguishing the marked increase in the values of the coefficients at the start of the attack from the values describing normal traffic.

1.1.5. Detection values for autocorrelation and regression

The next two coefficients describe the autoregressive and correlation dependence between two consecutive cw windows, or rather between the random variables X(t) and X(t-1), whose realizations are the flow increments a(t) in the windows considered.

$$X(t) = c(t).X(t-1) + \varepsilon(t), \qquad \rho(t) = \frac{E[(X(t) - EX(t))(X(t-1) - EX(t-1))]}{\sigma(t)\sigma(t-1)}$$

A great advantage of these coefficients is that their values range within the interval (-1,1). At the moment the attack was loaded into the evaluated cw windows, for most attacks their values began to rise rapidly towards the upper limit of 1. For this reason, it is possible to set a certain limit value (threshold) for both coefficients, the exceeding of which would signal the beginning of a DDoS attack. The lower this value, the earlier the attack is signalled; on the other hand, attacks that are too low may cause false alarms. Based on the experiments carried out, and in an effort to identify as many DDoS attacks as possible without false alarms, we set the limit for the autoregressive coefficient to c(t) = 0.9 and for the correlation coefficient to ρ(t) = 0.75. In the following graphs, we again show the evolution of the coefficients for attacks N8, N2, S1 and LR3. Attack N8 represents the behaviour of the coefficients in most standard attacks; in N2 we see a standard attack in which, however, the autocorrelation coefficient has all values greater than 0.9 and the detection cannot be used. In the non-standard attack S1,

both coefficients exceeded the detection levels during the attack. For the low-intensity attack LR3, this method is unusable.

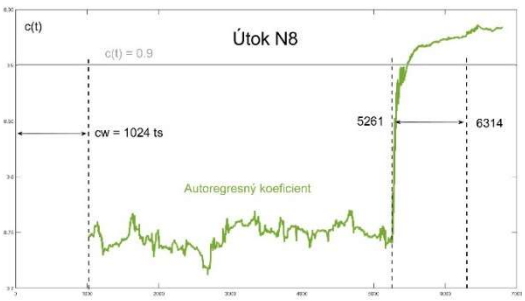


Fig. 26: Attack N8, Autoregressive coefficien

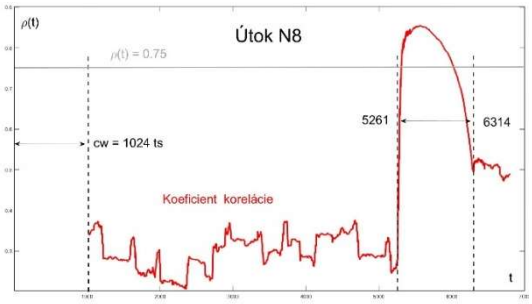


Fig. 27: Attack N8, Correlation coefficient

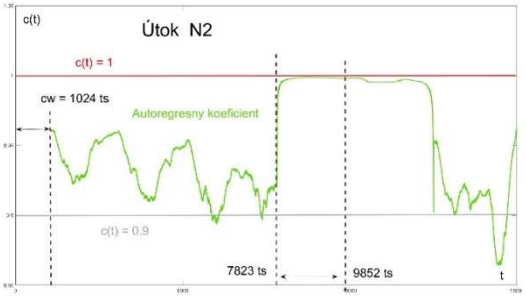


Fig. 28: Attack N2, Autoregressive coefficier



Fig. 29: Attack N2, Correlation coefficient

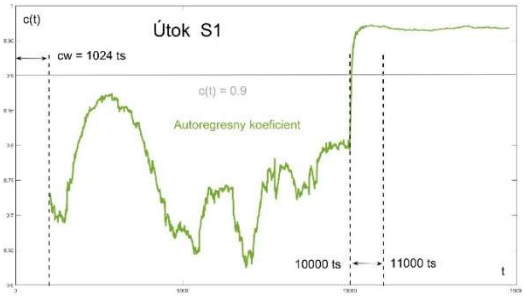


Fig. 30: Attack S1, Autoregressive coefficien

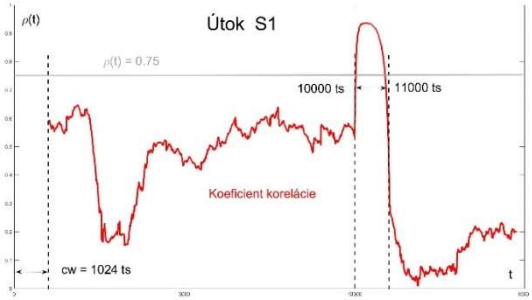


Fig. 31: Attack S1, Correlation coefficient



Fig. 32: Attack LR3, Autoregressive coefficie

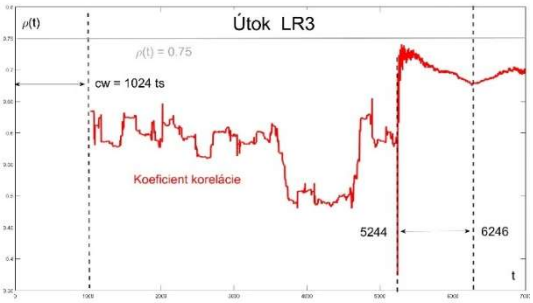


Fig. 33: Attack LR3, Correlation coefficient

For most attacks, it can be assumed that exceeding the detection values for autoregression and correlation can be used, in combination with other coefficients, for early and simple detection of a flood attack.

1.1.6. Response of divergence to a DDoS attack

The last analysed parameter, the Kullback-Leibler divergence, in a certain way compares the probability distribution P(.) of the increments a(t) in successive overlapping computation windows cw.

$$dvg(t) = \sum_{k=1}^{n} P(t-1) ln(\mathrm{P(t-1)/P(t)})$$

The advantage of divergence is that, on the samples monitored, in most cases it responded immediately to the onset of a DDoS attack with a high impulse (peak). The disadvantage is that it tended to respond in this way to peaks in normal traffic as well, which could cause a large number of false alerts.

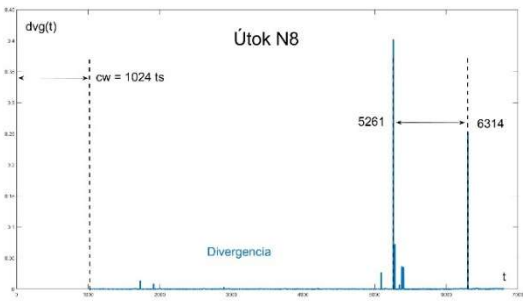


Fig. 34: Attack N8, Divergence

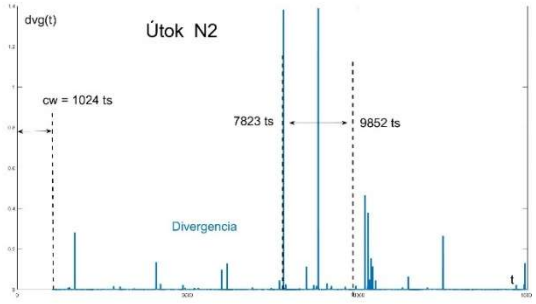


Fig. 35: Attack N2, Divergence

A further disadvantage of using divergence is the fact that the high impulse at the onset of an attack lasts a very short time compared with the other parameters; we therefore consider its application in combination with the other parameters to be problematic. We have therefore postponed the use of divergence to the future.

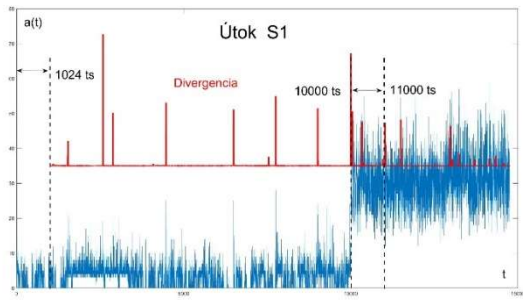


Fig. 36 Attack S1, Divergence

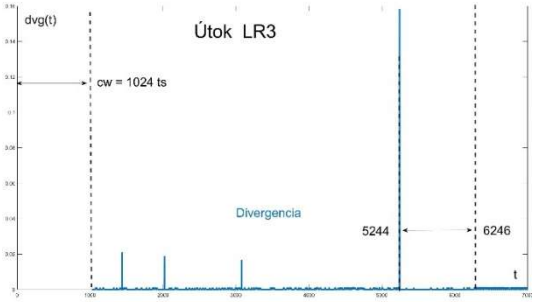


Fig. 37: Attack LR3, Divergence

1.1.7. Use of the 3σ tunnel for recognizing DDoS attacks

Based on the experiments, we chose 1000 previous values for calculating the mean and deviation, and we chose the interval width (EX-3σ, EX+3σ) (with a smaller interval size, frequent false alarms occurred). We called this method the predictive 3σ-tunnel. A natural question arises as to why not apply the predictive tunnel directly to the IP flow increments a(t) and not bother with calculating statistical coefficients. After carrying out experiments, we found that although exceeding the upper limit of the tunnel detects the onset of a DDoS attack

relatively early, false alarms are also frequently reported. When the multiple of the standard deviation was increased, the number of false alarms decreased, but at the same time the ability to detect the attack decreased. In the following figures, we will show only the upper limit of the 3σ-Tunnel for selected attacks.

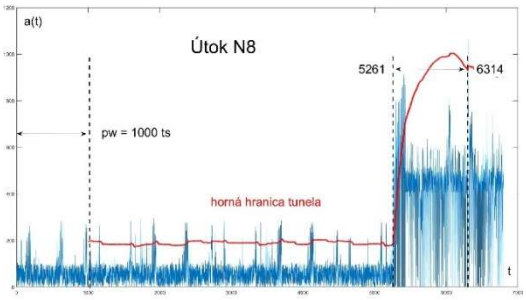


Fig. 38: Attack N8, upper limit of the 3σ tunn

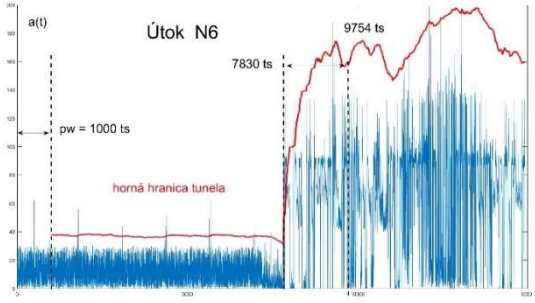


Fig. 39: Attack N6, upper limit of the 3σ tunnel

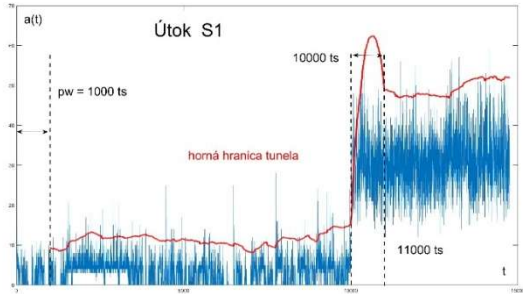


Fig. 40: Attack S1, upper limit of the 3σ tunn

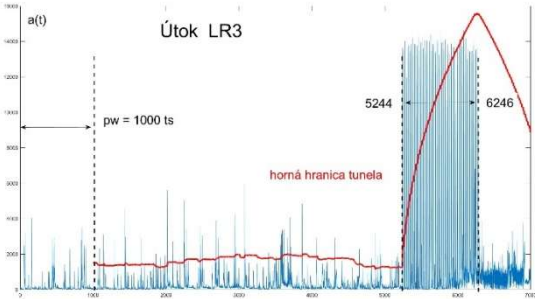


Fig. 41: Attack LR3, upper limit of the 3σ tunnel

In the following graphs, we will demonstrate the behaviour of the 3σ tunnel on the parameter with the greatest variability among the statistical coefficients considered, namely the Hurst exponent (estimated by RS statistics with removal of AR1 autoregression).

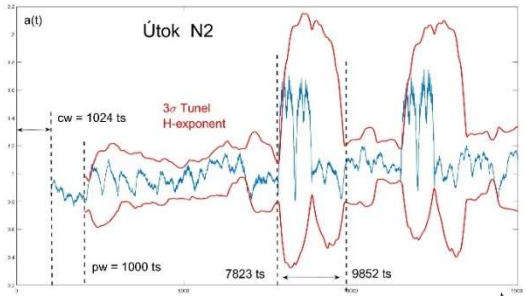


Fig. 42: N2, 3σ tunnel for the H-exponent

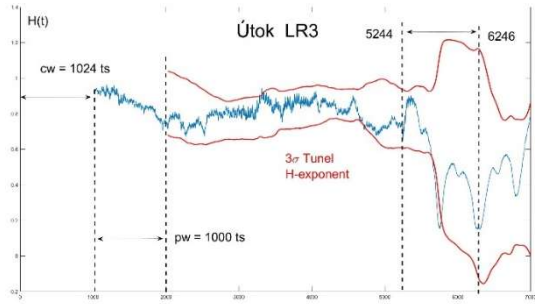


Fig. 43: LR3, 3σ tunnel for the H-exponent

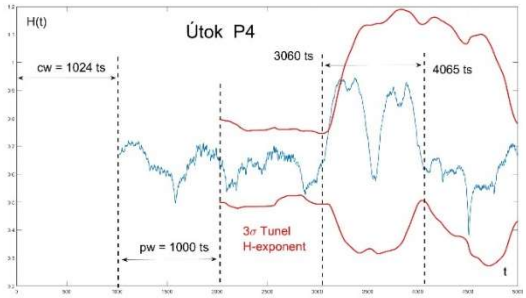


Fig. 44: P4, 3σ tunnel for the H-exponent

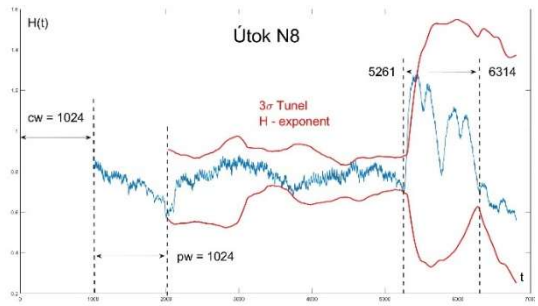


Fig. 45: N8, 3σ tunnel for the H-exponent

In most cases (except for attack LR3), the 3σ tunnel recognized the start of the DDoS attack early, with a minimal, sometimes even zero, number of false alerts.

As we have already stated, on the basis of experiments we set the size of the prediction window to 1000 ts and the interval width for the 1001st value to three times the standard deviation. With this setting, in most experiments (various coefficients and various attacks) we obtained early recognition of the start of the attack with a minimal number of false alerts. In the preceding figures, this is 3 cases out of 4. A tunnel set up in this way has two useful properties:

- It has a sufficiently “long memory” to be able to cover the local variability of the statistical coefficient and not cause a large number of false alerts
- It has a sufficiently "long memory" so as not to be able to respond to a relatively large increase in the values of the investigated coefficient at the start of a DDoS attack (large relative to the relative variability)

Of course, in order to be able to draw more relevant conclusions and recommendations, it is necessary to carry out a large number of experiments on records of various DDoS attacks.

1.1.8. Detection functions for automatic recognition of DDoS attacks

In the preceding chapters, we presented two ways of recognizing DDoS attacks using statistical coefficients. The first way was to determine detection values for suitable parameters: 1.00 for variability V(t) and the H-exponent, 0.90 for the autoregressive coefficient c(t) and 0.75 for the correlation coefficient ρ(t). The second way, using the predictive tunnel, was analysed in the preceding section.

When using the limit values, we introduced a two-valued logical function Y[.], which evaluates the exceeding of the limit value for a given statistical coefficient. In the following graphs, we will show the variability V(t), the autoregressive coefficient c(t) and the correlation coefficient ρ(t) and the course of their function Y[.] for attacks S1a N2.

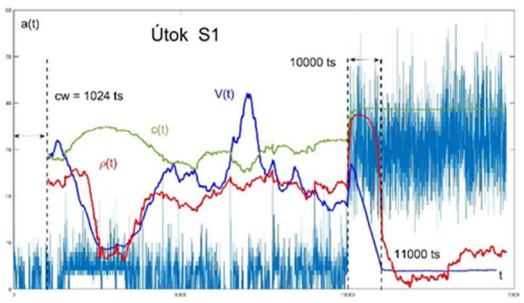


Fig. 46: S1, coefficients V(t), c(t), ρ(t)

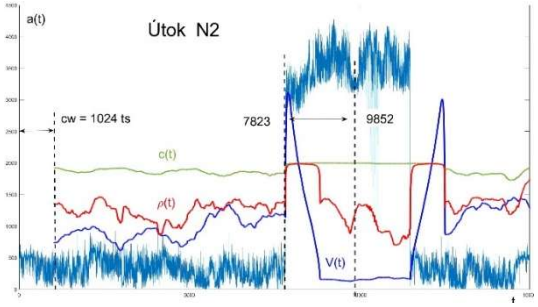


Fig. 47: N2, coefficients V(t), c(t), ρ(t)

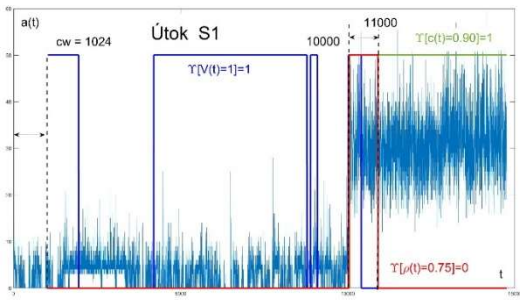


Fig. 48: S1, course of the functions Y[.]

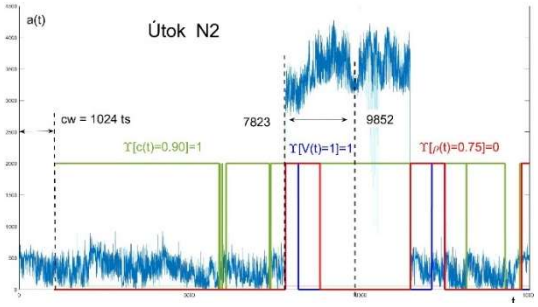


Fig. 49: N2, course of the functions Y[.]

Funded by the European Union
NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

The sets of values on which the coefficients exceed their detection limits will be called detection intervals. In Figs 48 and 49 we can see how these intervals overlap one another. In order to eliminate false alerts and at the same time achieve timely recognition of attacks, we will introduce a new detection function D(t) for the coefficients monitored, where, for simplicity, we denote e.g. for the coefficient V(t) the function Y[V(t)>1,00] as YV(t):

$$D(t) = YV(t).YC(t) + YV(t).Y\rho(t) + YC(t).Y\rho(t)$$

The following figures show the course of the detection function for attacks S1 and N2.

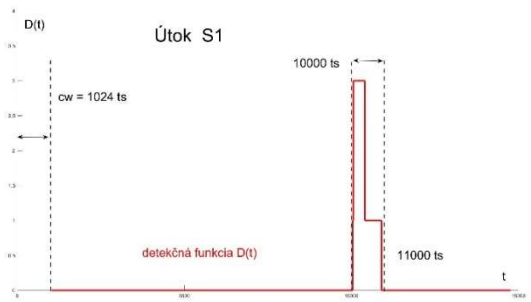


Fig. 49: S1, course of the detection functi

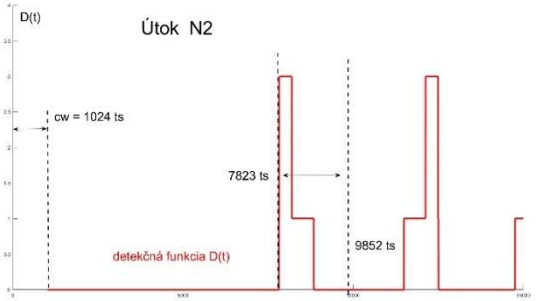


Fig. 50: N2, course of the detection function D(t)

Owing to the shape of the function D(t), we eliminated possible false alarms and, at the same time, on the basis of the values 1 or 3 we can recognize at what time two or all 3 statistical parameters detected the attack. Table 1 presents the evaluation of the use of the detection function D(t) for the analysed attacks. The first value represents the number of false alarms, the second determines the time at which at least two coefficients detected the attack, and the third value indicates the time at which all three statistical coefficients V(t), c(t) and ρ(t) detected the attack.

attack	S1	N2	LR3	S5	N6	N7	N8
D(t)	0 – 34 - 52	0 - 5 - 5	x	x	0 – 75 - 119	0 - 228 - 480	0 - 47 - 210

Tab. 1: Evaluation of DDoS attack recognition using the detection function D(t)

Recognition of the attack based on the limit value was effective mainly for standard attacks, while in several cases two coefficients exceeded their limit values substantially earlier than all three. For non-standard attacks such as LR3, P4 and S5, no two detection intervals, if they arose at all, had a common intersection, and therefore the attack was not recognized. We will also apply the same methodology to attack recognition using the predictive tunnel. In the previous chapter, taking into account the attack recognition time and the number of false alarms, we determined the three most effective statistical coefficients: skewness K(t), kurtosis Skw(t) and variability V(t). Since we selected coefficients whose values increase significantly at the onset of a DDoS attack, we will deal only with the upper limit of the predictive tunnel. We will evaluate the time of exceeding the upper limit as a detection interval and describe it with a two-valued function Γ[.], where, e.g. for the coefficient of kurtosis, we denote ΓK(t) = Γ[K(t)>Hr(t)], where Hr(.) is the upper limit of the prediction interval for the coefficient K(.).

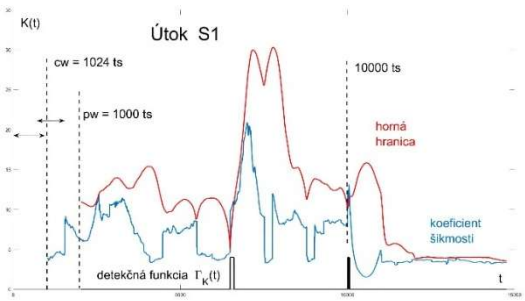


Fig. 51: S1, course of the function Γ(t)

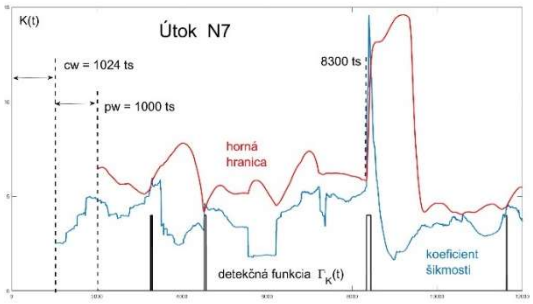


Fig. 52: N7, course of the function Γ(t)

Similarly to the previous method, recognition by means of a threshold value, we now also introduce a detection function that will be a combination of the identifiers Γ(t):

$$DP(t) = ΓK(t).ΓSkw(t) + ΓK(t).ΓV(t) + ΓSkw(t).ΓV(t)$$

Table tab.2 presents the evaluation of the use of the detection function DP(t).

attack	S1	N2	LR3	S5	N6	N7	N8
$D_P(t)$	1 – 6 - 28	0 - 1 - 1	1 - 1 - 1	1 - 18 - 41	1 - 12 - 12	1 - 39 - 49	0 - 1 - 1

Tab. 2: Evaluation of DDoS attack recognition using the detection function D(t)

1.1.9. Evaluation of tests

On the basis of the experiments carried out so far, we can see that the use of selected statistical coefficients has considerably greater potential for fast recognition of DDoS attacks (FAD – Fast Attack Detection) than standard monitoring of average and peak intensity. In a DDoS attack, there is a strong assumption that these intensities will increase significantly. However, the average intensity grows only linearly depending on the size of the sliding window, which leads to late detection of the attack, and recognizing an attack by setting some threshold for the peak rate can lead to frequent false alerts owing to the occurrence of peaks in the IP flow.

When standard IP traffic turns into a DDoS attack, not only do the average and peak intensities change significantly, but the overall probabilistic structure of the flood flow also changes. To capture this change, we used the statistical coefficients presented. These can be divided into two types: coefficients that exceeded some detection value at the onset of the attack, and coefficients that responded to the attack with a marked change in their values or with an impulse. Some coefficients even belong to both groups. On the basis of this division, we created:

- The method of using detection limits (MDH)
- The method of using prediction tunnels (MPT)

By comparing Tables 1 and 2, we can see that MPT recognized every type of attack considered, and did so considerably earlier than MDH; however, for most attacks it signalled one false alarm. We assume that this problem could be eliminated in several ways, for example by adding another suitable detection coefficient, or by some type of smoothing of the coefficient values before applying the prediction tunnel, e.g. exponential or Holt smoothing.

1.2. Detection of unwanted e-mail messages (Spam)

Spam e-mails (unwanted messages) account for approximately 60% of global e-mail traffic. Although spam detection software has improved considerably since the first spam message in 1978, anyone who has an e-mail account knows that spam remains a time-consuming and costly problem.

From the wider range of commonly used methods and algorithms for detecting and distinguishing between harmless/solicited e-mail messages (ham) and unwanted/unsolicited messages (spam), we selected and tested the following methods:

- Decision Tree
- SVM

1.3. Detection of malicious code (Malware)

Malware, short for malicious software, comprises various types of malicious software, including viruses, worms, Trojan horses and ransomware. Each type has unique characteristics and methods of infection and damage. Viruses, for example, attach themselves to clean files and spread uncontrollably, damaging file systems. Worms, on the other hand, replicate themselves without human intervention and often exploit network vulnerabilities. Trojan horses deceive users by disguising themselves as legitimate software, thereby gaining unauthorized access to systems.

Ransomware represents a particularly dangerous form of malware. It encrypts the victim's data and demands a ransom for the decryption key. This type of malware can cause considerable operational disruption and financial losses. The impact of malware is not limited to individual users; it also extends to large organizations and can lead to massive data breaches, financial losses and reputational damage.

Preventing and mitigating malware attacks involves a combination of robust anti-virus software, user education and adherence to cybersecurity best practices, such as regular software updates and caution when downloading files or applications from the internet. Understanding the various forms of malicious software is crucial for effective cybersecurity measures, as it enables individuals and organizations to recognize potential threats and respond accordingly.

1.4. Phising detection

Phishing is a deceptive cyber tactic used by attackers to obtain sensitive information from unsuspecting victims. This method usually involves sending fraudulent e-mails or messages that imitate legitimate sources, such as banks, government agencies or well-known companies. The aim is to deceive recipients and make them disclose personal data, such as login credentials, credit card numbers or other sensitive information.

These attacks can be highly sophisticated and use social engineering techniques to create a sense of urgency or fear, which forces victims to react quickly without questioning the authenticity of the request. Phishing messages often contain links to fake websites that closely resemble legitimate ones, where victims unknowingly enter their sensitive information.

1.5. Detection of advanced persistent threats APT

Advanced persistent threats (APT) represent a category of cyberattacks in which an unauthorized user gains access to a network and remains undetected for an extended period. These attacks are usually aimed at stealing data or monitoring network activity rather than causing immediate damage. APTs are often sponsored by nation states or large criminal organizations, which makes them highly sophisticated and difficult to detect and eliminate.

APTs involve a lengthy, stealthy and complex infiltration process. Attackers use various methods to gain entry, including spear-phishing, exploitation of vulnerabilities or the use of malicious software. Once inside, they establish a foothold, often create backdoors and use techniques to remain undetected. The ultimate goal of an APT is not immediate disruption, but long-term espionage, data theft or the acquisition of strategic advantages.

Combating APTs requires a multifaceted approach, including the implementation of multi-layered security defences, continuous monitoring of network activity, regular security audits and training of employees to recognize potential threats. Early detection and rapid response are essential to mitigate the impact of APTs. Organizations must adopt a posture of constant vigilance and readiness to respond to these advanced threats.