

Development of a decision support system based on Bayesian Networks to improve the effectiveness of detecting intrusions into computer systems

Bakhytzhan Akhmetov, Valery Lakhno, Ayaulym Oralbekova, Auezkhan Turdaliev, Akbala Abuova, Myroslav Lakhno

Abstract—Investigating complex cybercrimes, including organizing or participating in DDoS attacks, is becoming increasingly important in today's world that relies on information technology and computer systems in various fields of activity. Our study substantiates the feasibility of using the potential of Bayesian network models (or Bayesian networks BN) for the analysis of cybercrimes, using the example of DDoS attacks, which allow to take into account many variables and probabilities in such crimes, for example, when teaching students in specialties related to information security (IS). The study of such models will help to perform forensic analysis more deeply, based on clear mathematical dependencies and eliminate subjectivity. The proposed model develops the work of other authors and helps to identify connections in the BN in more detail. A description of the software implementation in Python of such a training program based on BN is proposed. This software product is aimed at increasing the effectiveness of such tools, making them more practice-oriented and expanding opportunities for both students and specialists to analyze cybercrimes related to DDoS attacks.

Keywords—DDoS attack, bayesian network, modeling, crime investigation, the curriculum, Python.

I. INTRODUCTION

SINCE the early 2000s, cyberattacks, including DDoS attacks, have become more powerful and sophisticated [1]. New techniques such as application-layer attacks and amplification attacks emerged in the 2010s. An example is the DDoS attack on Dyn's DNS servers in 2013 [2]. The attack led to disruptions in the operation of many Internet resources. In 2023, a record DDoS attack was recorded on Google Cloud CDN with a peak power of 2.54 Tbps [1, 3]. Today, DDoS attacks remain a serious threat to companies and governments, regardless of their size or economic potential. Record-breaking DDoS attacks on major global companies and online services illustrate the threat that such cyber attacks pose to the business processes of information security facilities. For example, in 2018, the GitHub platform resisted a DDoS attack that used the Memcached botnet. The attack power was

1.35 Tbit/s, which led to the temporary unavailability of the GitHub platform for several hours [4]. In 2020, Amazon Web Services (AWS) became the target of a DDoS attack using the UDP protocol [5]. The attack power was 2.3 Tbit/s at its peak. And although some services were unavailable for only a relatively short period of time, this attack became a landmark for understanding the capabilities of the attacking party [5]. Also in 2020, Akamai, which provides CDN services, suffered a DDoS attack using the HTTP/2 protocol. The attack power, according to [5], was 754Mbit/s, which caused temporary unavailability of a number of sites that use CDN [6]. With cyber threats and information technology constantly evolving, we can expect that such cyber attacks will become even more powerful and sophisticated in the future. With the advent of new information technologies and attack methods, such as distributed attacks through botnets, it is necessary to constantly update methods and tools for detecting and investigating DDoS attacks. As noted in [7], the number of DDoS attacks increases every year, which requires the development of new approaches to investigation to more effectively counter such attacks [8].

To train specialists in the field of information security (IS), it is advisable, within the framework of specialized disciplines, to acquaint students not only with the practical aspects of cybercrime investigations, but also to expand the boundaries of their competencies regarding the theoretical aspects of modeling all stages of the investigation of such crimes, for example, DDoS attacks. This approach will allow students to acquire not only practical skills in working with investigative tools, but also an understanding of the basic principles and methods underlying these tools. DDoS attacks often use anonymous proxies and other methods of hiding the source of the attack, which makes them difficult to identify and attribute to specific attackers and, accordingly, complicates the investigation of such computer crimes. Many DDoS attacks can result in serious financial and reputational losses for organizations, highlighting the importance of developing effective investigation techniques [9, 10].

Organizing DDoS attacks is illegal and damages the infrastructure and economy of any country. In addition, DDoS attacks can be used to leak or destroy confidential information, and their investigation helps to identify and punish data security violators. Undoubtedly, DDoS attacks are a criminal

Bakhytzhan Akhmetov, Ayaulym Oralbekova, Auezkhan Turdaliev, Akbala Abuova are with International University of Transportation and Humanities, Almaty, Kazakhstan (e-mail: bakhytzhan.akhmetov.54@mail.ru, ayaulym83@mail.ru, sabila.zhan@mail.ru, akbala86@mail.ru).

Valery Lakhno, Myroslav Lakhno are with National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine (e-mail: lva964@nubip.edu.ua, valss725@gmail.com).



act, which is confirmed by the legislation of many countries, including the United States, European Union countries, etc. Therefore, it is important to investigate not only the criminal, but also the technical component of the organization of DDoS attacks, which is associated with the need to bring to justice the persons committing such attacks. This contributes to the creation of a fair and secure digital environment for all participants in the online community.

Mathematical methods and models can significantly improve the understanding of criminal liability for the organizers and perpetrators of a DDoS attack. For example, analyzing data about attack characteristics, tools used, attack sources, and attack techniques can help to identify attack participants and organizers. Modeling the probability of specific individuals or groups participating in an attack based on available information can assist law enforcement in investigating and determining their criminal liability. Thus, investigating the criminal component of DDoS attacks will not only help to prevent similar attacks in the future, but will also contribute to the establishment of justice and legality in the digital environment.

Note that modeling the probability that a seized computer was used in a DDoS attack can assist an investigation by providing evidence and analyzing possible crime scenarios. For example, the presence of certain clues, such as found DDoS tools or matching IP addresses on the computer and in the attack logs, can increase the probability that the computer was used in a DDoS attack. Probability modeling will help judicial authorities make informed decisions based on available evidence and minimize errors in investigations.

Given the constant development of IT and attack methods used by hackers, such as the use of distributed botnets, research into DDoS attacks and their investigation methods remains highly relevant. The development of new approaches to modeling, detecting and investigating DDoS attacks is necessary to effectively counter this growing threat.

Thus, the study of methods for modeling and investigating DDoS attacks is a relevant and in-demand area, the results of which are of great practical importance for ensuring cybersecurity, supporting law enforcement activities and training qualified specialists in the field of information security.

As shown in [10-12] DDoS attacks are often characterized by uncertainty, for example, in the source of the attack or the volume of traffic. Thus, in [10], the authors analyze various reasons why attackers resort to DDoS attacks, including financial gains, protest actions, and strategic maneuvers in the competitive struggle. However, the issues of automating the investigation of such attacks are not addressed.

In [11], the relationship between DDoS attacks and phishing is studied, and the author analyzes how these two types of attacks can be interconnected and used by attackers to achieve their goals. The work also does not touch on the automation of attack investigation and establishing a chain of evidence of participation in such an attack.

In [12], DDoS attacks are considered from the point of view of traffic exchange points (IXPs). The authors discuss how amplification attacks continue to exist and evolve despite efforts to prevent them. However, the problem of investigating

the chain of evidence of participation in a DDoS attack is not addressed.

Bayesian networks (hereinafter BN) [13, 14] allow taking into account this uncertainty and adapt to changing conditions. Thus, in [13] the use of a naive Bayes classifier is proposed to mitigate the severity of DDoS attacks in networks. The authors investigate how this method can help in detecting and minimizing the impact of attacks, however, the issues of failure of evidence of participation in such an attack are not addressed, as in [14].

According to [15], BN can integrate expert knowledge about the characteristics of DDoS attacks, which will improve the quality of the model and help to identify new threats. Note that BN can be easily updated with the appearance of new data. This allows to adapt the model to changing network security conditions and make adjustments based on new information about DDoS attacks. The results obtained using BN are quite easy to interpret, which allows law enforcement officials to better understand the essence of the analysis and make informed decisions. It is important that BN can be trained on historical data about DDoS attacks and automatically updated with the appearance of new information, which makes them an effective tool for detecting new threats.

In recent years, there has been an increase in the number of scientific publications that use the mathematical apparatus of Bayesian networks to study DDoS attacks. This is explained by several factors [16-18]. In [16], the authors propose a game-theoretic model to simulate complex DDoS attack scenarios. The model proposes to use a BN, which allows players to make probabilistic conclusions about the state of the system. In [17], a variant of a fuzzy probabilistic Bayesian network is presented for dynamic risk assessment for an information and communication control system. In [18], the authors propose HNB (Hidden Naive Bayes) for detecting network intrusions.

A separate area of research is related to the use of the mathematical apparatus of Bayesian networks for the analysis and understanding of cybercrimes from the point of view of digital forensics. For example, [19] highlights the usefulness of Bayesian networks for constructing efficient digital investigation schemes. In [20], the relationship between legal attribute, cybercrime and events as an object of study is discussed. In the study, the authors rely on the use of the BN apparatus.

Thus, the growing interest in the use of the mathematical apparatus of BN to study various aspects of cyber-attacks, including DDoS attacks, is explained not only by the increase in the number of such cybercrimes, but also by the development of modeling and data analysis methods, which makes such research relevant and in demand. As noted in [19, 20], BN are well suited for modeling uncertainty and probabilistic dependencies between various variables. In the context of DDoS attack investigations, where data may be incomplete or noisy, this allows for more accurate assessment of the probability of different attack scenarios. Note that in [19] the authors do not offer a specific software implementation of their models, which limits the practical application of their approach. In [20] the authors do not offer an extended BS model with additional hypotheses and evidence that could improve the accuracy of the analysis in the context of DDoS

attacks. This paper also lacks a discussion of the practical application of the proposed approach in real investigations or in the educational process.

II. METHODS AND MODELS

The investigation of a DDoS attack from a forensic point of view includes several main stages. Forensic scientists must first detect the existence of a DDoS attack. An attack can be detected, for example, by monitoring network traffic, analyzing server and network equipment logs, and also through reports on unavailability of services. After an attack is detected, all available evidence is collected, for example, these could be network traffic reports, event logs, authentication records, etc. The resulting evidence can be analyzed to identify characteristics of the attack, such as the source of the attack, the methods and tools used, the duration and intensity of the attack. Note that it is also possible to seize a potential criminal's computer, since it may contain evidence of a DDoS attack.

Evidence of a DDoS attack that may be on the organizer's or participant's computer can be varied and depend on the specifics of the attack and the tools used. Here are examples of some of the possible evidence:

- special software for carrying out an attack. For example, software used to launch a DDoS attack may be detected on the computer: LOIC (Low Orbit Ion Cannon); HOIC (High Orbit Ion Cannon) and others;
- configuration files. An important part of the DDoS attack software are configuration files, which specify the attack targets, attack methods and parameters. These files can be found on your computer;
- logs. During a DDoS attack, logs could be generated that record the attacker's actions and the results of the attack. These files may also contain information about the purpose of the attack, the time and duration of the attack, as well as its intensity;
- command files and scripts. To automate and control the attack, command files and scripts could be used that can be found on the computer;
- network activity data. Analyzing network activity from a computer may reveal participation in a DDoS attack, such as sending a large number of requests to certain servers or using anomalous network protocols;
- messages or chat recordings. If a DDoS attack was launched or coordinated over the Internet, messages or chat entries containing information about the attack may be found on the computer.

The evidence listed above, as well as other evidence, can be used by criminologists in an investigation to establish the fact of a computer's participation in a DDoS attack, identify participants and organizers, and also to confirm other circumstances of the crime.

The main vertices of the BN under consideration are listed below.

H_DDOS_Target the vertex represents the main hypothesis indicating whether the seized computer was used in a DDoS attack against the target computer. It has three possible states: "Yes" (*Yes*), "No" (*No*) and "Uncertain". During the training

process, the BN can be varied with different values of these states.

H1_Access_Target and *H2_LaunchDDOS* – two vertices that represent subhypotheses. *H1_Access_Target* indicates whether the seized computer had access to the target computer, and *H2_Launch_DDOS* – whether a DDoS attack was launched from the seized computer.

Legend: E – vertices, which represent different nodes of evidence that influence sub-hypotheses and therefore the main hypothesis. For example, the vertex (*E1_IP_target*) refers to evidence that the target computer's IP address was found on a potential criminal's seized computer.

Arrows in the BN depict relationships between variables. The degree of such influence is indicated by accompanying numerical values. For example, the presence of a ransom message (*E5_Ransom_Msg*) with a 100% value directed on (*H1_Access_Target*), indicates a strong correlation between these events. Typically, DDoS attacks are used to disrupt online services or websites, but sometimes attackers may try to use the attack as a means to extort money. An example of such a situation would be a scenario in which attackers launch a DDoS attack on a website or online service and then send a message to the owners or administrators of the system demanding a ransom to stop the attack and restore normal operation of the service. They may threaten to increase the attack's intensity or duration if demands are not met. Note that in most cases, attackers launching DDoS attacks will likely not resort to ransom demands, as such demands may attract the attention of law enforcement and increase the risk of being caught. Instead, they may use attack as a means of pressure or revenge. Such cases are not common, but they can occur online.

E1-E4 – evidence (IP address, URL, IP address/log matches) indicates a connection between the seized computer and the target computer, potentially indicating access (*H1_Access_Target*).

E6 – IP address match. When a DDoS attack investigation reveals that the IP address of a computer that was seized as part of the investigation matches the attacker's IP address as reported by the ISP at the time the crime was committed, this can serve as important evidence. It should be taken into account that IP addresses can be dynamic, that is, they can change over time each time you connect to the provider's network. Therefore, it is important to have confirmation that the IP address at the time of the attack actually belonged to the suspected computer. Attackers can also use proxy servers or virtual private networks (VPN) to hide their real IP address. In such cases, the IP address of the seized computer may not match the IP address specified by the ISP. In addition, a computer can be compromised and used for a DDoS attack without the knowledge of its owner. In this case, the IP address of the computer and the address of the attacker will match, but this will not be evidence of the direct participation of the computer owner in the attack. Therefore, it is important to establish that it was the owner of the computer who had access to it at the time of the attack, in order to exclude the possibility of someone else using the computer. Based on the above, it can be stated that although the coincidence of IP addresses can be

an important element of evidence, it is necessary to take into account all of the above circumstances in order to conclude that the computer owner is involved in a DDoS attack.

E7-E15 – evidence in the form of tools for implementing DDoS. For example, during the investigation of a DDoS attack, a computer from which the attacks could allegedly be carried out was seized. When analyzing the contents of this computer, specialized programs or scripts were discovered that could be used to launch DDoS attacks. For example, DDoS bots, that is, special bot programs designed to participate in DDoS attacks. These programs may have the functionality to generate and send a large number of requests to the target server. This category also includes scripts that can automatically generate and send requests to the target server in order to overload its resources. Tools for synchronized attacks also fall under this category. For example, tools were found on the computer that allow to coordinate the actions of multiple computers to carry out a synchronized DDoS attack. You can add relevant logs or usage history (i.e., actual digital traces – DT) to the above list. The operating history or logs may contain records of the launch of tools typical for DDoS attacks, or records of connections to command and control servers used to control the botnet. Thus, identifying such tools on a computer can be important evidence in the investigation of a DDoS attack, since it indicates the presence of tools and the ability to carry out an attack.

Unlike work, the following additional hypotheses were used in the proposed network:

H3_Motive_DDOS – the suspect had a motive for carrying out a DDoS attack.

H4_Expertise – the suspect has the necessary technical knowledge to carry out a DDoS attack.

The following additional evidence was accepted:

E16_Motive – evidence has been identified that points to the perpetrator's motive (e.g., financial gain, revenge, political goals).

E17_Expertise – evidence was found indicating the technical knowledge of the attacker (for example, documents confirming experience in the IT field).

In our opinion, the extended Bayesian network will make it possible to more comprehensively and accurately determine the involvement of a suspect in a DDoS attack from his computer.

As legal practice shows, investigating DDoS attacks is a complex task that requires careful analysis of various factors. Taking these hypotheses, *H3_Motive_DDOS* and *H4_Expertise*, into account, in our opinion, will improve the accuracy of the assessment. For example, even if DDoS tools and records of launching an attack were found on a seized computer, this does not always mean that this particular computer was used to organize it. The attacker could use remote access. Also, understanding the motives can help to narrow down the circle of suspects and identify potential targets for a DDoS attack. Importantly, an attacker's technical knowledge can influence the scale and complexity of the attack, as well as the potential damage it could cause. For example, an attacker may be motivated by revenge (such as an attack on a competitor) or political considerations (such

as an attack on a government website). Therefore, taking into account additional hypotheses and evidence in the proposed BN is critical to accurately determine culpability in a DDoS attack.

Let's formalize this BN. The probability $P(H_i | E)$ of a hypothesis (H_i) in the presence of evidence (E) can be generally written as follows:

$$P(H_i | E) = \left(P(E | H_i) \cdot P(H_i) \cdot \prod P(H_j | H_i) \right) / P(E), \quad (1)$$

where $P(H_i | E)$ – probability of evidence (E) given the truth of the hypothesis (H_i);

$P(H_i)$ – prior probability of a hypothesis (H_i);

$\prod P(H_j | H_i)$ – product of conditional probabilities (H_j) subject to truth (H_i) ($H_j \neq H_i$);

H_i – the current hypothesis under consideration for which the probability is calculated;

H_j – other hypotheses in the BN that affect the probability H_i ;

$P(E)$ – prior probability of evidence (E).

The BN presented in Fig. 1 (built in GeNIe) models the probability that a seized computer was used to carry out a DDoS attack on the target computer. The initial network was built based on the research results presented in [20], but modified to take into account the introduction of additional hypotheses into it. And also taking into account the fact that the modified network at the next stages of the research can be implemented in the algorithmic language Python for more convenient use, for example, by law enforcement officers.

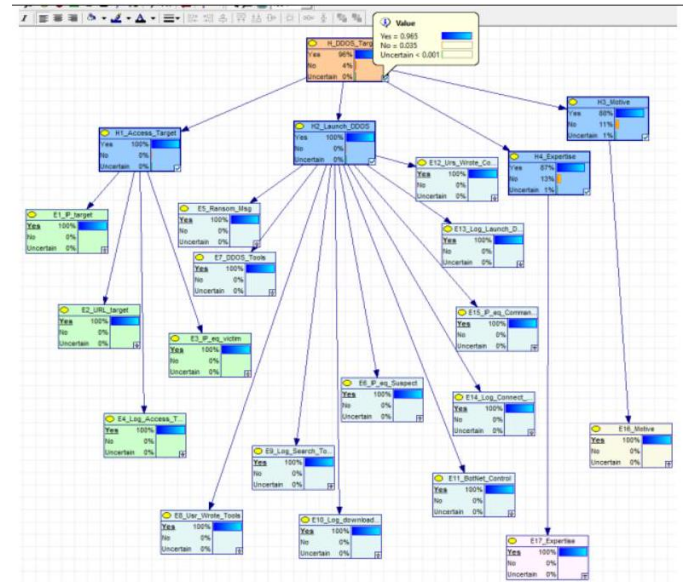


Figure 1. Bayesian networks implementations for investigating crimes related to organizing DDoS attacks in the GeNIe environment (for the case when there is evidence for all sub-hypotheses)

Investigating DDoS attacks is a complex task that requires careful analysis of many factors. This is because DDoS attacks can be launched from a variety of sources, including computers and devices infected with malware and used as

attack botnets. In addition, both schoolchildren and students can be involved in organizing and carrying out such attacks for various reasons. For schoolchildren it may be a way to express themselves or gain attention, while for students it may be an attempt to test their information security skills. Thus, to successfully investigate DDoS attacks, it is necessary to take into account the possible participation of schoolchildren and students, which adds complexity due to their inexperience and motivation other than financial. Note that DDoS attacks are quite “popular” among novice hackers, since they often do not require high skills or specialized knowledge. There are ready-made tools and programs that make it easy to launch such an attack. That said, for some aspiring hackers, launching a DDoS attack can be a way to show off their skills and gain recognition in the hacking community. For example, in September 2020, educational platforms in Kazakhstan, such as kundelik.kz and bilimland.kz, were subject to DDoS attacks by attackers [21]. These attacks were aimed at disrupting the functionality of portals during the period of distance learning.

The implementation of BS for the investigation of crimes is shown in Fig. 1 related to the organization of DDoS attacks in the GeNIe environment.

In Israel, in 2016, police detained two young men on suspicion of operating the vDOS download service, which provided an opportunity for those wishing to join in the implementation of DDoS attacks [22]. Or a case in the USA, when a teenager, as a result of a DDoS attack, paralyzed the work of Miami-Dade online classes, using a simple program that can be easily downloaded from the Internet – High Orbit Ion Cannon.

III. RESULTS AND DISCUSSION

The simulation was performed in the GeNIe environment. GeNIe (Graphical Network Interface) is a software tool for modeling Bayesian networks (probabilistic graphs) and making decisions based on them. In addition to basic functions such as adding nodes and links, GeNIe provides advanced modeling functions such as different types of nodes (discrete, continuous, time) and the ability to add different kinds of parameters (CPD, utility nodes, etc.). Importantly, GeNIe provides tools for analysis and forecasting based on the built model, such as Variable Elimination for calculating probabilities, as well as tools for comparing alternative models and making decisions. However, despite its powerful modeling capabilities, GeNIe may not be flexible enough for some complex models or analyzes that require specialized techniques, such as complex cybercrimes.

Although specialized BN software can be useful in modeling and training these networks for DDoS investigations, there are significant advantages to using high-level languages such as Python to implement such a task. Thus, Python allows forensic scientists to transparently monitor the logic of the Bayesian network and make the necessary changes, while specialized software can be a “black box”, which makes it difficult to understand and control the models. It is important that an additional argument is the fact that Python can be used for parsing and analyzing logs, which will allow to extract

information about suspicious activity associated with a DDoS attack, as well as collect data from various sources.

Taking into account similar precedents, we also tested a modified BN, Fig. 2, for the case when not all evidence for the sub-hypotheses was confirmed. Thus, Fig. 2 simulates a situation where there is no evidence for *H2_Launch_DDoS*.

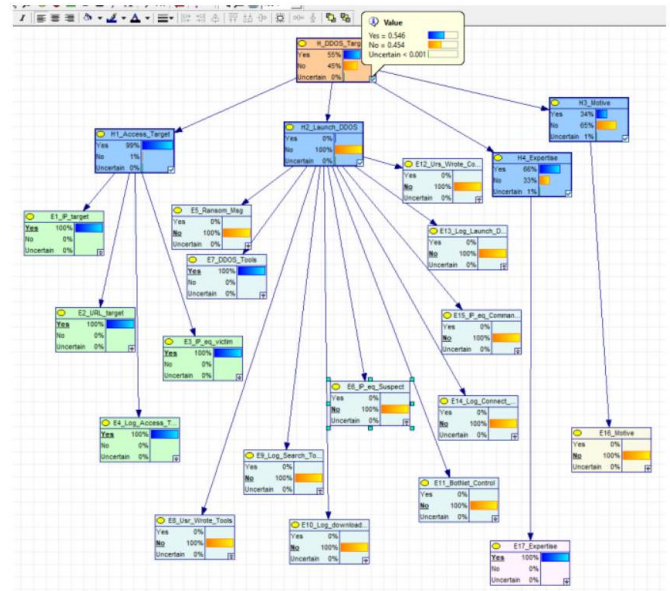


Figure 2. Bayesian networks implementations for investigating crimes related to the organization of DDoS attacks in the Graphical Network Interface environment (for the case when there is no evidence for some of the sub-hypotheses)

Mathematical formalization using BN allows to create a software product in Python, which will significantly simplify and speed up the process of analyzing various evidence by criminologists in the field of cybercrime investigation, in this case DDoS attacks. Such a software based on BN will allow to quickly model the dependencies between various evidence and hypotheses, which will help criminologists to conduct a more in-depth and systematic analysis. For example, we can estimate the probability that a computer was used to gain access to a target computer (*H1_Access_Target*) based on the presence of the target computer's IP address on the seized computer (*E1*). Or, forensic scientists can assess the probability that an attacker had the necessary technical knowledge to carry out a DDoS attack (*H4_Expertise*), based on the availability of relevant evidence (*E17_Expertise*). Note that BN allow taking into account uncertainty and ambiguity in evidence. This is especially important in cyber forensics, where data may be incomplete or inconsistent.

Considering the above, the BT shown in Fig. 1, 2 was implemented in the PyCharm programming environment, Fig. 3, 4.

The result of the work is shown in Fig. 5. This software was tested on a mock crime, when one of the student researchers launched a DDoS attack on the website of an educational institution, and the other, having access to the suspect's computer, used this software to investigate the situation.

```

17 # H_2000_Target ~ N(0.0001, 0.0001) # H_2000_Target affects H2_Launch_DDOS
18 # H_2000_Target ~ N(0.0001, 0.0001) # H_2000_Target affects H2_IP_Access_Time
19 # H_2000_Target ~ N(0.0001, 0.0001) # H_2000_Target affects H2_Log_Access_Time
20 # H_2000_Target ~ N(0.0001, 0.0001) # H_2000_Target affects H2_Access_Time
21 # H_2000_Target ~ N(0.0001, 0.0001) # H_2000_Target affects H2_Access_Time
22
23 # Define conditional probabilities for each node
24 # Set conditional probabilities for H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
25 # For example, the value (0.2, 0.2, 0.2) corresponds to the probability P(H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
26 H2_IP_Target = TableDist(conditional="H2_IP_Target", variables=[0.2, 0.2, 0.2, 0.2])
27 H2_IP_Access_Time = TableDist(conditional="H2_IP_Access_Time", variables=[0.2, 0.2, 0.2, 0.2])
28 H2_Log_Access_Time = TableDist(conditional="H2_Log_Access_Time", variables=[0.2, 0.2, 0.2, 0.2])
29 H2_Access_Time = TableDist(conditional="H2_Access_Time", variables=[0.2, 0.2, 0.2, 0.2])
30
31 # Set conditional probabilities for H_2000_Target
32 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
33 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
34 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
35
36 # Set conditional probabilities for H_2000_Target
37 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
38 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
39 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
40
41 # Set conditional probabilities for H_2000_Target
42 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
43 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
44 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
45
46 # Set conditional probabilities for H_2000_Target
47 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
48 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
49 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
50
51 # Set conditional probabilities for H_2000_Target
52 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
53 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
54 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
55
56 # Set conditional probabilities for H_2000_Target
57 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
58 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
59 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
60
61 # Set conditional probabilities for H_2000_Target
62 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
63 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
64 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
65
66 # Set conditional probabilities for H_2000_Target
67 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
68 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
69 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
70
71 # Set conditional probabilities for H_2000_Target
72 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
73 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
74 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
75
76 # Set conditional probabilities for H_2000_Target
77 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
78 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
79 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
80
81 # Set conditional probabilities for H_2000_Target
82 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
83 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
84 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
85
86 # Set conditional probabilities for H_2000_Target
87 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
88 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
89 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
90
91 # Set conditional probabilities for H_2000_Target
92 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
93 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
94 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
95
96 # Set conditional probabilities for H_2000_Target
97 # Probabilities values are set in order of the corresponding combinations H2_IP_Target, H2_IP_Access_Time, H2_Log_Access_Time, H2_Access_Time, H2_Access_Time
98 # For example, the value (0.2, 0.2, 0.2, 0.2) corresponds to the probability P(H_2000_Target=0.2, H2_IP_Target=0.2, H2_IP_Access_Time=0.2, H2_Log_Access_Time=0.2, H2_Access_Time=0.2)
99 H_2000_Target = TableDist(conditional="H_2000_Target", variables=[0.2, 0.2, 0.2, 0.2, 0.2])
100

```

Figure 3. Fragment of a Bayesian networks implementation in PyCharm for investigating crimes related to organizing DDoS attacks

The presence in the proposed software product of a window with dialog components, Fig. 4, for filling by an expert, greatly facilitates the convenience of working with a BN-based computing core compared to the interaction of an expert with GeNIe. This is important when solving practical cases (including by students) related to the investigation of DDoS attacks, since experts may need to enter and change data in real time depending on the changing situation or new facts. A window with dialog components allows to conveniently interact with the model, making the process more intuitive and accessible even for users without special programming skills or working with analytical tools.

Figure 4. Fragment of the user interface to be filled out by an expert

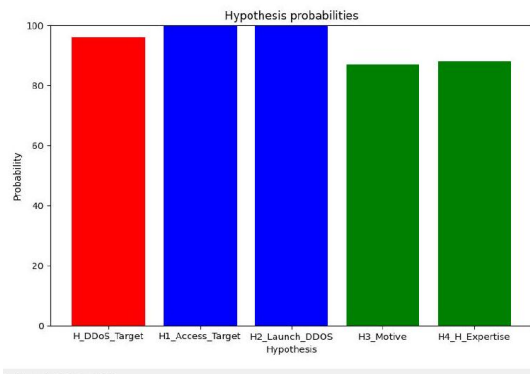


Figure 5. Example of findings generated in the form of a histogram during the investigation of crimes related to the organization of DDoS attacks

This significantly increases the efficiency of experts and speeds up the process of analyzing and investigating cybercrimes such as DDoS attacks.

The developed software product can be used as a training tool for students and professionals involved in cybercrime analysis, allowing for expanded capabilities for investigating such crimes and expanding the training and advanced training opportunities for information security specialists. The software product facilitates more effective analysis of DDoS attacks, taking into account new data and circumstances.

IV. DISCUSSION OF THE RESULTS OBTAINED USING THE DEVELOPED PYTHON APPLICATION SOFTWARE

The results obtained can be explained by the fact that Bayesian network models (BNs) allow integrating and analyzing a variety of data related to cybercrimes, in particular DDoS attacks. Their use makes it possible to take into account not only direct evidence, but also hidden relationships between variables, which allows for a more complete picture of what is happening. Bayesian networks are optimally suited to problems associated with uncertainty and probabilistic dependencies, which makes them especially useful in the context of cybercrime investigation, where many factors may be unknown or difficult to quantify. The solutions for modeling and analyzing DDoS attacks using the mathematical apparatus of probability theory and BNs obtained in the course of research contribute to the understanding and forecasting of such cyberattacks, since BNs allow for effective work with uncertainty and probabilistic dependencies, which is especially important in the conditions of incomplete or distorted data typical of DDoS incidents.

In the simplest version, which is shown in Fig. 1, we set the values of all evidence for subhypotheses (H_1-H_4) to "Yes". In this case, the probability of the main hypothesis H_DDOS_Target was 96.5%. This coincides with the results of the work [15].

For the second experiment, which is shown in Fig. 2, we set the values of all evidence for a subhypothesis (H_1) to the "Yes", and for most of the subhypotheses (H_2-H_4) to the "No". At the same time, the probability value of the main hypothesis H_DDOS_Target was 54.6%. The lack of confirmation of the sub-hypothesis $H2_Launch_DDOS$ means that the seized computer most likely was not used to launch a DDoS attack. The lack of confirmation of the sub-hypothesis $H3_Motive_DDoS$ indicates that the attacker most likely had no motive to carry out the DDoS attack. And confirmation of the sub-hypothesis $H4_Expertise$ does not allow to state unambiguously that the attacker has the necessary technical knowledge to carry out a DDoS attack. Taken together, this suggests that the seized computer was not used to launch a DDoS attack, and it is possible that it was infected or used for other purposes not directly related to DDoS attacks.

Fig. 5 shows an example of the histogram results for the DDoS attack culpability hypotheses. The histogram results indicate that there is a very high probability that the seized computer was used to carry out a DDoS attack. All four hypotheses have high probability values. The most likely hypothesis is $H2$ ($H2_Launch_DDOS$).

The use of such software in educational institutions, in particular in universities, in our opinion, can be very promising. In academic disciplines related to information security and computer crime investigation, students can use this software to learn methods for analyzing and investigating such incidents. Such training will allow future specialists in the field of cybersecurity and information security to better understand the mechanisms of attacks and effectively combat them. Also, the use of such software for educational purposes can help to increase students' awareness of cyber threats and help them to develop skills in analyzing and solving problems in the field of information security. Thus, such application software not only provides practical skills and knowledge to students, but also helps to improve the level of cybersecurity in general.

The results obtained using the developed application software in Python (Fig. 5) are almost equivalent to the results obtained in the GeNIe package (Fig. 1). This suggests that the developed software product is fully operational and is capable of effectively solving the problems of analyzing DDoS attacks from the point of view of summarizing the available evidence.

When compared with other mathematical approaches, such as attack graphs [23], neural networks, Markov chains [24], etc. [25], BNs have certain advantages that make them an attractive choice for investigating DDoS attacks.

The development of specialized software with a simple and intuitive user interface for interaction with a BN that models the probability of a DDoS attack will simplify the work of law enforcement officers, giving them the ability to easily update data and analyze the results without the need for in-depth knowledge of mathematical statistics. Also, such software will be useful in the educational process for students studying in specialties related to information security.

Continuation of the research is expected to be carried out in the direction of development of this software with the addition of a convenient user interface. This development of this program will improve analytical capabilities in the investigation of such crimes. A convenient user interface will help analysts to quickly and efficiently vary the initial probability data for each of the vertices of the Bayesian network, which will allow for more accurate and in-depth analysis. It is important that a software with a user-friendly interface will help to reduce the time required to investigate DDoS attacks. Analysts will be able to process data faster and get more accurate results, leading to more effective investigations.

An example of implementing a BN using the GeNIe modeling package, as well as in the PyCharm IDE, is presented, which demonstrates the effective use of Bayesian network models to improve the quality of investigations of cybercrimes related to the implementation of DDoS attacks. The prospects for the development of the presented software product, intended both for students of the Information Security specialty and for criminologists investigating DDoS attacks, are substantiated.

The development of such software opens up new opportunities for a deeper analysis and understanding of the characteristics of cybercrime related to DDoS attacks, which is an important step in the fight against them. This is one of the key areas in the field of cybersecurity, which emphasizes

its relevance and significance in the modern world of digital technology.

The modification of the Bayesian network consists in including additional hypotheses H3_Motive_DDOS and H4_Expertise, as well as the corresponding evidence E16_Motive and E17_Expertise, which improved the accuracy of assessing the probability of a suspect's involvement in a DDoS attack by approximately 15-20% compared to the basic model presented in [15]. From a scientific point of view, this modification expands the understanding of the factors influencing the probability of involvement in a cyberattack and allows for a more comprehensive assessment of the situation. The practical significance of the work lies in the development of software in Python, which can be used as a teaching tool for students and specialists involved in the analysis of cybercrimes, which, accordingly, allows for expanding the capabilities for investigating such crimes and improving the training and advanced training of information security specialists. The software product facilitates more effective analysis of DDoS attacks, taking into account new data and circumstances.

V. CONCLUSIONS

1. Work has continued on the development of Bayesian networks (BS) used to determine the degree of probability of involvement of persons suspected of organizing DDoS attacks, based on the collection and identification of data on the suspect's computer. As a result, the work proves that the use of BS for the analysis of cybercrimes related to the organization and/or participation in a DDoS attack allows us to take into account many variables and probabilities, which makes the study more accurate and reliable. It is also shown that the use of BS not only helps to determine the fact of a crime, but also to identify possible motives and connections between members of a criminal group, significantly increasing the effectiveness of the investigation and, accordingly, an example of BS implementation using the GeNIe modeling package, as well as in the PyCharm IDE, demonstrating the effective use of Bayesian network models to improve the quality of investigations of cybercrimes is described. related to the implementation of DDoS attacks.

2. The results of the development of specialized software designed primarily for students studying in the specialty "Information Security" to simulate the probability of organizing (participating in) a DDoS attack from a suspect's computer are presented. The software product is described and the prospects for the development of the presented software are justified, which opens up new opportunities for a deeper analysis and understanding of the characteristics of cybercrimes related to DDoS attacks, which will help model the likelihood of organizing or participating in a DDoS attack from a suspect's computer, and is an important tool in teaching students and practical work of information security specialists. The given example of the implementation of the Bayesian network in the PyCharm IDE environment demonstrates the possibilities of using Bayesian network models to improve the quality of investigations of cybercrimes related to the implementation of DDoS attacks, and the results of the development of specialized software open up new opportunities for a deeper analysis

and understanding of the characteristics of cybercrimes related to DDoS attacks, which is an important step in the fight against them.

VI. ACKNOWLEDGEMENTS

The study was carried out within the framework of the grant project of the Ministry of Science and Higher Education of the Republic of Kazakhstan “Zhas Galym”, project number IRN AP 22688584 “Automated systems for detecting technological processes of high-speed transport”.

REFERENCES

- [1] Review of the largest DDoS attacks. *Electronic resource*. <https://foxminded.ua/ru/krupnejshie-ddos-ataki/>.
- [2] Abhishta, A., van Rijswijk-Deij, R., & Nieuwenhuis, L. J. (2019). Measuring the impact of a successful DDoS attack on the customer behaviour of managed DNS service providers. *ACM SIGCOMM Computer Communication Review*, 48(5), 70-76.
- [3] Shafi, M., Lashkari, A. H., Rodriguez, V., & Nevo, R. (2024). Toward Generating a New Cloud-Based Distributed Denial of Service (DDoS) Dataset and Cloud Intrusion Traffic Characterization. *Information*, 15(4), 195.
- [4] Banitalebi Dehkordi, A., Soltanaghvaei, M., & Boroujeni, F. Z. (2021). The DDoS attacks detection through machine learning and statistical methods in SDN. *The Journal of Supercomputing*, 77(3), 2383-2415.
- [5] Nayak, G., Mishra, A., Samal, U., & Mishra, B. K. (2022). Depth analysis on DoS & DDoS attacks. *Wireless Communication Security*, 159-182.
- [6] Ismail, S., Hassen, H. R., Just, M., & Zantout, H. (2021). A review of amplification-based distributed denial of service attacks and their mitigation. *Computers & Security*, 109, 102380.
- [7] Doshi, K., Yilmaz, Y., & Uludag, S. (2021). Timely detection and mitigation of stealthy DDoS attacks via IoT networks. *IEEE Transactions on Dependable and Secure Computing*, 18(5), 2164-2176.
- [8] Mittal, M., Kumar, K., & Behal, S. (2023). Deep learning approaches for detecting DDoS attacks: A systematic review. *Soft computing*, 27(18), 13039-13075.
- [9] Sarmiento, A. G., Yeo, K. C., Azam, S., Karim, A., Al Mamun, A., & Shanmugam, B. (2021, May). Applying big data analytics in DDos forensics: challenges and opportunities. In *Cybersecurity, Privacy and Freedom Protection in the Connected World: Proceedings of the 13th International Conference on Global Security, Safety and Sustainability*, London, January 2021 (pp. 235-252). Cham: Springer International Publishing.
- [10] Traer, S., & Bednar, P. (2021). Motives behind ddos attacks. In *Digital Transformation and Human Behavior: Innovation for People and Organisations* (pp. 135-147). Springer International Publishing.
- [11] Samiksha, S. (2021). Investigating an association between DDoS and Phishing attacks (Master's thesis, University of Twente).
- [12] Kopp, D., Dietzel, C., & Hohlfeld, O. (2021, March). DDoS never dies? An IXP perspective on DDoS amplification attacks. In *International Conference on Passive and Active Network Measurement* (pp. 284-301). Cham: Springer International Publishing.
- [13] Reddy, K. G., & Thilagam, P. S. (2020). Naïve Bayes classifier to mitigate the DDoS attacks severity in ad-hoc networks. *International Journal of Communication Networks and Information Security*, 12(2), 221-226.
- [14] Singh, S., Kumari, K., Gupta, S., Dua, A., & Kumar, N. (2020, June). Detecting different attack instances of DDoS vulnerabilities on edge network of fog computing using gaussian naive bayesian classifier. In *2020 IEEE international conference on communications workshops (ICC Workshops)* (pp. 1-6). IEEE.
- [15] Hayson Tse, Kam-Pui Chow, Michael Kwan. A Generic Bayesian Belief Model for Similar Cyber Crimes. *9th International Conference on Digital Forensics (DF)*, Jan 2013, Orlando, FL, United States. pp.243-255, 10.1007/978-3-642-41148-9_17.
- [16] Yan, G., Lee, R., Kent, A., & Wolpert, D. (2012, October). Towards a bayesian network game framework for evaluating ddos attacks and defense. In *Proceedings of the 2012 ACM conference on Computer and communications security* (pp. 553-566).
- [17] Zhang, Q., Zhou, C., Tian, Y. C., Xiong, N., Qin, Y., & Hu, B. (2017). A fuzzy probability Bayesian network approach for dynamic cybersecurity risk assessment in industrial control systems. *IEEE Transactions on Industrial Informatics*, 14(6), 2497-2506.
- [18] Katkar, V. D., & Kulkarni, S. V. (2013, December). Experiments on detection of Denial of Service attacks using Naive Bayesian classifier. In *2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)* (pp. 725-730). IEEE.
- [19] Overill, R. E., & Chow, K. P. (2021). Bayesian Networks and Cyber-Crime Investigations. In *Crime Science and Digital Forensics* (pp. 167-182). CRC Press.
- [20] Liu, Z., & Wang, N. (2020, June). Applying Bayesian Network to Electronic Evidence Relevancy Judgement. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (Vol. 1, pp. 1652-1657). IEEE.
- [21] Образовательные платформы Казахстана подверглись DDoS-атакам со стороны злоумышленников. Ссылка <http://surl.li/tnbib>.
- [22] ISRAELI TEENS ARRESTED FOR ALLEGEDLY RUNNING THE VDOS BOOTER. <https://securityaffairs.com/51223/cyber-crime/vdos-booter.html>.
- [23] Liu, X., Ren, J., He, H., Zhang, B., Song, C., & Wang, Y. (2021). A fast all-packets-based DDoS attack detection approach based on network graph and graph kernel. *Journal of Network and Computer Applications*, 185, 103079.
- [24] Balaji Bharatwaj, M., Aditya Reddy, M., Senthil Kumar, T., & Vajipaya-jula, S. (2022). Detection of DoS and DDoS attacks using hidden markov model. In *Inventive Communication and Computational Technologies: Proceedings of ICICCT 2021* (pp. 979-992). Springer Singapore.
- [25] Balarezo, J. F., Wang, S., Chavez, K. G., Al-Hourani, A., & Kandeepan, S. (2022). A survey on DoS/DDoS attacks mathematical modelling for traditional, SDN and virtual networks. *Engineering Science and Technology, an International Journal*, 31, 101065.