Threat Hunting as a Method of Protection Against Cyber Threats

Nataliia Lukova-Chuiko[a], Andriy Fesenko[a], Hanna Papirna[a], Sergiy Gnatyuk[b]

[a] *Taras Shevchenko National University of Kyiv, 24 Bohdan Havrylyshyn St, Kyiv, 04116, Ukraine*
[b] *National Aviation University, 1 Liubomyr Huzar ave., Kyiv, 03058, Ukraine*

Abstract
This paper presents the structuring of a new approach to countering cyber threats – Threat Hunting. This concept is proactive threat search, mainly a manual process with elements of automation, in which the analyst uses his knowledge and skills to check large amounts of information for indicators of compromise, according to a predetermined hypothesis of the presence of a threat. All key elements of Threat Hunting approach were explained as well as functional diagram for a deep understanding and application of this approach in practice by specialists in the field of cybersecurity was proposed in the paper.

Keywords
Threat Hunting, indicators of compromise, proactive cybersecurity, cyber threat.

1. Introduction

To date, most information security threats are known, and can be defended by traditional means of protection such as antivirus, firewalls, and so on. Such threats include spam, denial-of-service attacks, viruses, rootkits, and other classic malware. The remaining minority of threats are unknown and the most dangerous. They are difficult to detect and even more so to protect against them. Examples of such threats are encryption viruses, crypto miners, etc.

In a company with organized information security management processes, the majority of the risk of known threats can be resolved with a traditional risk management approach: avoid, accept (accept possible financial or image losses), reduce (implement the necessary protection) or transfer (for example, to a service provider). Instead, protecting against zero-day vulnerabilities, targeted attacks, phishing, supply chain attacks, and a large number of other attacks is much more difficult. The consequences of these threats will be much more serious than the consequences of spam or viruses, from which modern anti-virus software is quite able to protect.

This situation has led to the development of means of protection against cyber threats in the direction of developing new technology that would be able to counteract the most serious and complex threats.

Proactive threat search or Threat Hunting (hereinafter – TH) is the latest way to counter cyberattacks, which through proactive and iterative search, allows to detect complex threats that traditional means of protection are not even able to notice. It should be noted that TH is not a specific software or hardware product and is not a passive activity. Proactive threat search is, first of all, mainly a manual process with elements of automation, in which the analyst, based on his knowledge and skills, checks large amounts of information for indicators of compromise, according to a predetermined hypothesis of the presence of a threat. Due to the fact that this concept is relatively new in the field of cybersecurity, it is advisable to explain it from the opposite, that is, to describe what this process is not in order to avoid confusion of concepts and technologies.

Thus, TH is not [1]:

- a form of penetration testing, although TH can lead to an understanding of which area of the IT infrastructure to conduct penetration testing;
- a constant search for indicators of compromise, although they are used in the process;
- a security monitoring, although TH results can be used to provide a new detection mechanism followed by security monitoring;
- an incident response, although THs can lead to incident disclosure, thereby initiating an incident response process;
- a request execution in the security tool, although automation and data request are an important part of the TH. In other words, if an information security tool (such as an antivirus, intrusion prevention system, etc.) can run the process autonomously, it is not a proactive threat search. TH professionals should use the tools to support their hypotheses and investigations, but the simple use of the tool is not able to provide this process completely;
- a process that gives a guaranteed result. Not every proactive search is able to detect an attacker or lead to the creation of new detection mechanisms. This does not necessarily mean that there is no malicious influence. For example, the data required for the investigation may be missing, or the compromise indicator under investigation may not have been present during the search. However, the search will always yield some secondary result, such as a deeper understanding of the infrastructure or the identification of missing data;
- a simple process. TH requires a deep knowledge of the information environment and an excellent understanding of the capabilities of the attacker. Compared to traditional security monitoring, TH is a much more challenging task.

The aim of this paper is to study TH as a mean of protection against cyber threats in information systems and networks with a detailed systematization of steps to obtain key factors for its successful implementation. The authors also propose the extended functional diagram of the TH process created by other studies in the field.

2. The essence and purpose of Threat Hunting

Threat Hunting is a preventive, iterative and human-oriented identification of cyber threats that are internal to the IT network and bypass existing security measures [2].

The main purpose of the TH is to reduce the time required to search for traces of attackers who have already compromised the IT environment. By identifying these traces as soon as possible, the impact of violations on the organization can be minimized. Gaps in the detection of violations - an important concept in the context of this goal.

Thus, there will always be a gap between what an organization can detect and the ability of a skilled attacker to avoid detection. An attacker's capabilities will be different for each attacker, and detection capabilities will be different for each organization. Although attackers usually have the opportunity to avoid detection, at some point they may trigger detection mechanisms, either because these detection mechanisms have evolved, or due to human factors.

A good TH program aims to track the behavior of attackers and continuously reduce the gaps in the detection of violations. In particular, the search for threats focuses on actions that may go unnoticed. The TH focuses on events that go beyond traditional detection capabilities, and can detect missed or misinterpreted events that can be used to improve the detection and further training of information security analysts [1].

2.1. Main Threat Hunting characteristics

Based on the above, it is possible to derive the main characteristics of the TH [3].

- Initiative nature. Threat Hunting experts are proactively looking for indicators of harmful activity in the network, instead of waiting for signals from traditional detection mechanisms to start an investigation.
- Finding the unknown. TH mainly focuses on known elements, however, can detect unknown;

- Assuming the possibility of a violation. Proactivity does not make sense if the information security professional believes that prevention and detection mechanisms are sufficient to prevent violations. The TH assumes that the violation has already taken place but has not yet been identified.
- Understanding of the attacker. It is important to understand the motivation and thoughts of the attacker. These are important characteristics that determine how persistent and professional an attacker is.
- Creativity and repeatability of the process. The TH process is a creative process. The quality of the process largely depends on the creativity, experience and knowledge of the specialist conducting the search. It is an iterative process: the search for threats can lead to new discoveries and new investigations; the collection of information may also lead to new assumptions about the current TH process.
- Data driven process. Threats require a lot of data. The best sources of information for TH are considered to be: endpoint information, firewall logs, domain name service (DNS) logs, etc. The higher the quality of the data, the higher the probability of success of the investigation. These data should help the specialist to study the hypotheses of TH, and not complicate the situation by adding information noise.
- Based on hypotheses. Derivation of hypotheses and their proof play a key role in the TH process according to any currently known methodology.
- Requires teamwork. The team of TH specialists uses a common approach and determines what threats to look for. The team will also prioritize hypotheses based on the risk levels associated with the threat. Typically, the skill set for working in such teams is reduced to general knowledge in the field of information security, knowledge of the IT environment, knowledge of analysis methods, knowledge of attacker methods and good communication skills.

2.2. Threat Hunting classification

According to the leading American company in the field of cybersecurity and big data analytics - Sqrrl, there are five types of TH [4].

The first type is TH based on data. A natural starting point for stimulating TH is to create hypotheses based on observational data to review data that is already available. For example, proxy logs, traffic statistics and more. Analysts can use any of the data sources as a basis for generating hypotheses, creating queries or reports that identify abnormal behavior.

The second type is TH based on data mining. Threat data and analytics can provide organizations with ample opportunities for TH. Unfortunately, this approach is one of the most difficult because organizations need to be aware of both the different levels of reliability of information and the usefulness and rare nature of collecting inside information based on things like incident response. Analysts who follow this approach can be assisted by safety graphs to obtain the context of the threat. The graph of security behavior provides critical points of integration, processing and analysis, allowing the analyst to effectively use information about threats to manage and enrich the activities of the TH.

The third type - entity-based TH. This type is characterized by focusing on high-risk / value entities, such as critical intellectual property and network resources. Attackers typically target specific assets or high-risk users in an organization (such as a domain controller or system administrator account). More and more organizations proactively determine what kind of assets it is before the attacker does it for them.

The fourth type is TH based on tactics, techniques and procedures (hereinafter - TTP). This approach focuses on the fact that much more important than just static indicators (domains, IP addresses, hashes) are the methods, tactics and procedures of attackers. These observations are excellent material for TH because they provide contextual starting points that are more suitable for human analysis than for automatic resolution.

The fifth type is a hybrid TH. In fact, any successful TH will combine combinations of the above types. For example, TH can be formed using information about threats around a particular attacker,

which informs the analyst about the types of TTPs that can be used by him, and the critical assets that he can target.

There is also another more simplified approach, which divides TH into two types [1].

A structured TH is a hypothesis-based TH: a hypothesis is created, the scope of the TH is estimated, and then the TH is executed. This approach can be based on both data mining and entity analysis, and TTP.

An unstructured TH is a data-driven TH. Potentially harmful activity can be detected by an analyst who simply looks at the available data for anomalies. This type of threat search does not start with a hypothesis, does not follow a predetermined path and is thus considered unstructured. It should be noted that unstructured TH requires a lot of effort and is less likely to give valuable results.

2.3. Threat Hunting triad

According to a study by the leading analyst and cybersecurity expert Daniel Akacki, a successful TH procedure contains three components [5].

The first component is people, cybersecurity professionals, who have the appropriate skills. In the above study, the following set of skills that a specialist must have to successfully conduct TH is represented:

- analytical thinking as the specialist must be able to make reasonable assumptions and plan a new course;
- ability to analyze the audit log of the system. Service and device logs are just a couple of the most important and underused sources of information for any security department. The ability to analyze logs for anomalies and switch between data sources to see the big picture is a key competency;
- network forensics skills - the ability to read and understand the data of captured network packets and determine the harmful nature of network traffic;
- knowledge of network architecture - understanding of different network devices and how they work in the IT environment;
- understanding the life cycle of the attacker. Understanding the various events that occur at any stage of the attack lifecycle prepares analysts to share and prioritize their results and actions;
- knowledge of working with information security tools is a large area, but at a basic level of understanding how log aggregators retrieve data, as well as the functions of packet capture analysis tools are sufficient for the analyst;
- knowledge of operating system architecture as different operating systems represent different attack vectors;
- understanding of basic methods of attacks. Exploit kits, malware, phishing, and misconfiguration - understanding how an attacker tries to break into a network is key to finding indicators of compromise.

The second component of the TH procedure is processes, because the goal of a mature security process should be to automate much of the threat detection with reliable rules and timely notification to give analysts more time to directly conduct TH.

Processes should be designed taking into account the desire to understand not only what data is already available, but also data sources that are missing or incorrectly configured, because it is impossible to protect what is unknown.

The third and final component of the TH procedure is the technology, which is described in more detail in paragraph 4.2.1.

3. Threat Hunting basics

Closely related to the TH process is cyber threat intelligence - a process of collecting, processing and analyzing information about attackers in cyberspace in order to disseminate effective information about threats by understanding the motives, capabilities and methods of attackers to inform about measures to reduce cybersecurity [2].

For such an analysis, the indicators of compromise are used. To date, the Pyramid of Pain by David Bianco [5] is the most popular and effective, according to leading experts in the field, model that explains the role of indicators of compromise in the process of TH.

This simple pyramid shows the relationship between the types of indicators that can be used to detect an attacker and the degree of "pain" (the probability of failure) that they cause him, if experts can prevent and prohibit the use of these indicators by an attacker.

Indicators by their hierarchy are considered in more detail below [5].

Hash values are at the lowest level of the pyramid. Most hashing algorithms calculate the message digest for the entire input and output a fixed-length hash unique to that input. In other words, if the contents of two files differ even by one bit, the resulting hash values of the two files will be completely different.

On the one hand, hash indicators are the most accurate type of indicator to focus on. The chances that two different files have the same hash value are so low that it is possible to almost completely rule out this possibility. On the other hand, any change to the file, even insignificant, such as throwing a bit in an unused resource or adding a zero to the end, leads to a completely different and unrelated hash value. Thus, the values of hashes are easy to change, in many cases it is not even necessary to track them.

One level above are IP addresses - the most fundamental indicator of the network. Attackers largely need to have some network connection to carry out the attack, and the connection means IP addresses. However, any fairly professional opponent can change IP addresses at any time convenient for him with very little effort. In some cases, if attackers use an anonymous proxy service, they can change IP addresses quite often without even paying attention. That is why blocking attacks based on IP-addresses is ineffective - if you block an attacker at one IP-address, he can usually recover without even interrupting.

Further up in the pyramid are the domain names. They are a little harder to change than IP addresses because they have to be registered, paid for and posted somewhere for them to work. However, there are a large number of DNS providers with non-strict registration standards (many of which are free), so in practice it is not so difficult to change the domain.

In the middle of the pyramid are the artifacts of the network and the host. This is the first level at which the EP specialist can begin to have a negative impact on the opponent. If it is possible to detect indicators at this level and respond to them, the attacker will be forced to change the settings and compilation of their tools.

From a technical point of view, a network artifact can be any byte that passes through the network as a result of an attacker's interaction. However, in practice, this means those parts of the activity that can distinguish harmful activity from the activity of legitimate users.

Host artifacts can be observations caused by the actions of attackers on one or more hosts. These can be registry keys or values that are created by certain malicious programs, files or directories.

Blocking network and / or host artifacts forces attackers to take a few steps back and spend time figuring out how their intelligence tool was discovered and fixing it.

At the penultimate level are the tools. Tools in this context are software used by an attacker to carry out his mission. These are mostly programs that they install themselves, not software or commands that can already be installed on a regular user's computer. Such programs include: utilities designed to create malicious documents for targeted phishing, backdoors used to set password crackers, etc.

New generation antivirus signatures or other systems that can find variants of the same files, even with moderate changes (communication protocols, hash values, etc.), can help detect indicators at such a high level.

At this level, it is possible to deprive the attacker of the opportunity to use one or more specific tools. Thus, attackers will need to spend time researching (finding an existing tool with the same capabilities), developing (creating a new tool if they have the appropriate knowledge and skills) and learning (figuring out how to use the tool and master it). That is why blocking at the level of tools is one of the most effective in counteracting targeted attacks.

Finally, at the top are the TTPs of attackers. When detection and response occur at this level, the action is directed directly at the behavior of the attackers, not against their tools. In terms of efficiency, this level is the most ideal. If a specialist in TH is able to respond quickly enough to

suspicious TTP, it forces attackers to carry out the most time-consuming of the possible actions: retraining and mastering new behavior.

From the above we can conclude that the process of successful TH is to focus on the highest level of the pyramid, the TTPs, to reduce the likelihood of achieving the goal by the attackers with the greatest efficiency. If the attacker uses non-trivial actions or the absence of TTP data in the known frameworks, it makes sense to lock at the lower levels of the pyramid until the TTP is detected.

4. The order of Threat Hunting conduct

According to the approach of the leading American company in the field of cybersecurity and big data analytics - Sqrrl, in general, the whole process of TH can be reduced to four main stages, which are repeated cyclically [6].

The approach is called "Hunting loop" and is designed to avoid potentially inefficient TH processes and create a formalized process. According to the Sqrrl approach, the goal of TH should be to overcome the loop as quickly and efficiently as possible. The more efficiently iterations are performed, the better the ability to automate new processes and move on to finding new threats.

4.1.   Stage 1 - Hypothesis creation

The TH process begins with asking questions about how an attacker can gain access to an organization's network. Then these questions need to be divided into specific and measurable hypotheses that determine what threats may be present in the network and how they can be identified [2].

Hypotheses cannot be generated with the help of tools, instead they should be obtained from the observations of a specialist based on the analysis of cyber threats, situational awareness or knowledge of the subject area [2].

Hypotheses must also be tested. The TH specialists must have the necessary data visibility and tools to find alleged evidence of malicious activity. A large variety of data types allows to explore more methods, and more data sources expand the arena for TH. Hypotheses, as a rule, focus on identifying a specific source of threat, tool or technique [7].

4.2.   Stage 2 - Research using tools and techniques

Once the observations have led to the development of hypotheses, they should be tested using all appropriate tools and methods available to TH specialists. Data visibility should be maximized by increasing the coverage of data collection in a centralized repository, and the types of data collected should include representative logs of organization's major network, infrastructure and security assets.

A fairly effective method at this stage is Linked data - a method of publishing structured data that allows to link them and seek confirmation of hypotheses using semantic queries. Related data analysis is particularly effective in presenting the data needed to solve hypotheses in an understandable form, and is therefore an important component of the TH. Linked data can even help prioritize and direct visualization, making it easier to search large datasets and use more powerful analytics. Methods of analysis of both source and related data should be used to identify patterns in disparate data sets, to detect the actions of attackers [6].

In general, it is possible to identify four types of techniques that can be used by specialists in TH at this stage [2].

- The search is the simplest method of querying the collected data. The search criteria should be specific enough so that the results are not unmanageable, but at the same time general enough not to miss any actions of the attackers. If necessary, it is possible to use wildcards in queries.
- The clustering is a form of statistical analysis that separates groups (clusters) of similar data points from a larger set based on specific characteristics, while grouping determines when several unique data points appear together based on certain criteria, such as several events that occur in a specific time window. The main difference is that the grouping requires an explicit set of data

points as input. However, both methods (clustering and grouping) are useful for detecting anomalies.

- The stack counting or accumulation is a kind of application of frequency analysis to large data sets to detect anomalies.
- The machine learning uses algorithms and statistical models to gradually increase the productivity of a particular task; for TH – it is the detection of abnormal data that may indicate the actions of attackers. In controlled machine learning, a set of training data is entered into the algorithm, and each data point is marked with the desired result, both normal and anomalous data are clearly defined.

4.2.1. Selection of the suitable tools and technologies

Today, there are many information security technologies that can provide assistance in the process of TH. However, the authors of this paper tend to narrow the set of technologies to the next most necessary.

- The class of SIEM (Security Information and Event Management) systems allows to monitor information systems, analyze security events in real time, for example, occurring on workstations, network devices, information security tools and other infrastructure elements. The data collected and analyzed by it helps to identify incidents or anomalies that remain invisible to specialized remedies.
- The class of EDR (Endpoint Detection and Response) systems is an alternative to traditional anti-virus solutions and provides modern protection of endpoints with adaptation to the landscape of complex threats. Such systems include both functionalities to detect complex attacks aimed at endpoints, and are able to respond quickly to detected incidents.
- The class of NTA (Network Traffic Analysis) systems is a new category of network security systems designed to intercept traffic flows and detect signs of complex, often targeted attacks.

It is especially significant that all three types of systems (subject to the selection of relevant solutions) support the possibility of seamless integration and continuous data exchange. That is, in such a scheme, the NTA is responsible for the visibility of information transmitted over the network, EDR delivers relevant information from endpoints, and SIEM aggregates event logs.

As experts in the field of cybersecurity, the authors of this paper note that the above systems are the technical basis for the construction of a modern SOC (Security Operations Center). SOC is a specialized center for monitoring and prompt response to information security incidents. Such a center is a group of information security experts who are responsible for continuous monitoring and analysis of the security of the organization, using a combination of technological solutions and acting within well-structured processes. SOC is designed to monitor activity in networks, servers and workstations, databases, applications, websites and other systems, detecting abnormal and malicious actions that may indicate a security incident or data compromise.

It is important to note that most often TH processes seek to implement organizations that already have their own SOC or use such services through outsourcing. Thus, it can be concluded that the planning and implementation of the TH procedure in the daily process of information security can afford organizations with a high level of maturity of security processes, which already have established procedures and technologies for threat prevention and are ready to move to a higher level - the level of proactive threat response.

4.3.   Stage 3 - Identification of tactics, techniques and procedures

Passing the second stage with the help of tools allows to reveal new harmful patterns of behavior and TTPs. This stage is one of the most critical in the whole cycle [6].

The gap in the detection of violations arises from the ability of attackers to evade the mechanisms of detection. As detection capabilities continue to evolve and expand, cybercriminals will find new

ways to evade these measures. Thus, over time, TTPs of attackers will evolve to ensure that they can evade detection and act unnoticed in the IT environment.

There are various models of attacker methodologies from which it is possible to begin to lay the foundation for a TH strategy. The most famous collections of such methodologies are: The Lockheed Martin Cyber Kill Chain, The Mandiant Attack Lifecycle and The MITRE ATT&CK Framework.

MITRE ATT&CK is a structure that describes the methodologies used by attackers during cyberattacks. It is presented in the form of a matrix consisting of eleven tactics, each of which contains a list of related techniques [8].

1. Initial access tactics are the vectors that attackers use to anchor a network.
2. Execution tactics are methods that lead to the execution of code controlled by an attacker in a local or remote system.
3. Persistence tactics are any change in access, action or configuration of a system that ensures that an attacker is constantly present in that system.
4. Privilege-enhancing tactics are the result of actions that allow an attacker to obtain a higher level of permissions on a system or network.
5. Evasion tactics consist of techniques that the attacker can use to evade detection or evade other defenses (firewalls, antivirus, etc.).
6. Credential access tactics include methods that provide access to or control over the credentials of a system, domain, or service used in an enterprise environment.
7. Detection tactics consist of methods that allow an attacker to obtain information about the system and internal network.
8. Lateral (horizontal) tactics consist of methods that allow an attacker to access and manage remote systems on the network, as well as run tools in remote systems.
9. Collection tactics consist of methods used to identify and collect information, such as confidential files, from the target network prior to exfiltration.
10. Exfiltration tactics include methods and attributes that lead or help an attacker to delete files and information from the target network.
11. Command and control tactics show how attackers interact with systems under their control in the target network.

In addition, MITRE ATT&CK can be useful in assessing the visibility of data - to identify all currently described methods requires a sufficient number of data sources [9]. The tables in this framework list the data sources and the number of methods that each source helps to identify (it should be noted that most methods require multiple data sources). This can help determine the priority of the organization's efforts to collect data [2].

4.4.   Stage 4 - Expansion of analytics

The fourth stage of the cycle forms the basis for informing and enriching automated analytics. Under no circumstances should threats be missed, it is important to automate them with the help of analytics so that the TH team can continue to focus on the following procedures. This can be done in a variety of ways, including developing a default search for regular execution, creating new analytics using tools such as Sqrrl, Apache Spark, R or Python, or even providing feedback to a controlled machine learning algorithm that confirms that the identified pattern of behavior is abnormal and harmful [6]. The extension of analytics can take a simpler form - to provide a new compromise indicator for comparison or to write a new SIEM rule for reactive detection. The faster automation of TH, the less repetition will be required of specialists and the faster their skills can be used to test new hypotheses [2]. Care should also be taken to ensure that any automated TH processes are reliable and continue to be beneficial. After automation, each analytical process must be tested for accuracy, which can be performed using metrics.

4.5.   Metrics

One of the main mistakes of organizations initiating the TH process to their overall information security strategy is that they do not define metrics for assessing the TH either because of the difficulty

of defining such indicators or because they believe that because threat detection must be a flexible process, indicators cannot be identified [11].

However, there are useful metrics that can measure the performance of the TH process to help improve it, as well as help build a business case for further investment (financial and time) in staff training and tools. The following is an approximate set of metrics that can be used [2].

- Graph of trends and / or comparisons: number of incidents detected proactively (compared to reactively).
- Percentage: data coverage (data types and asset coverage).
- Pie chart: number of hypotheses on MITRE ATT&CK tactics.
- Service level: the percentage of successful THs that led to a new analytical conclusion or detection rule; sensitivity and specificity of analytics or rules obtained as a result of TH (number of true and false positive results) [12].

Thus, we can conclude that the TH cycle is a simple but effective process that can radically improve the level of security of the organization. This procedure is most effective when used in conjunction with traditional security systems, complementing the measures and tools to detect cyber threats that already exist in most organizations. The ultimate goal of TH should always be to go through the four-stage loop as efficiently as possible.

5. Developing Threat Hunting process diagram

The "Hunting loop" by Sqrrl [6] provides a working and stable cycle of actions for experts. However, in the context of organizations where there are information security departments wishing to implement TH process, the cycle needs to be detailed and supplemented with initial data, as well as a connection with the classic incident management process [15]. The authors of the paper propose the improvement of the original cycle, described on the **Error! Reference source not found.** and its integration in the whole information security management system of organizations.
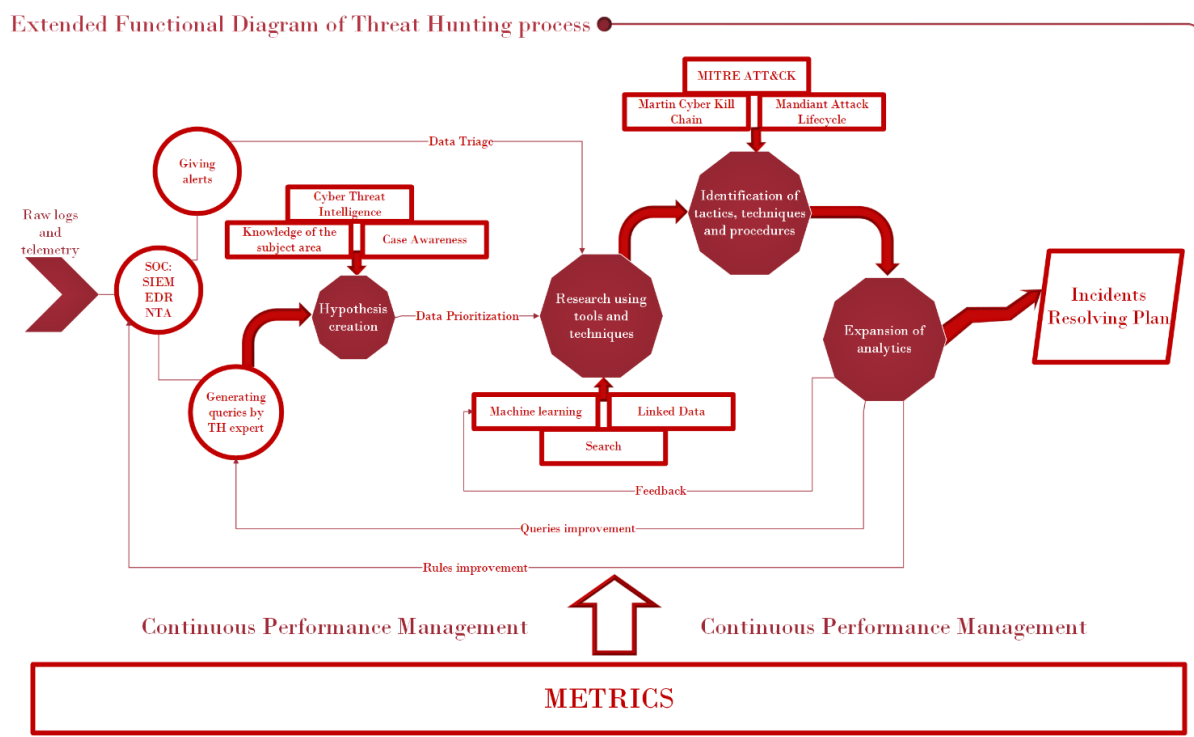

Figure 1: Extended Functional Diagram of Threat Hunting process

The proposed extended functional diagram of the TH process includes components listed below:
- raw logs and telemetry from network and infrastructure assets of the organization;

- means of the initial analytics and information processing – SOC: SIEM, EDR, NTA, which are described in more detail in paragraph 4.2.1;
- alerts and queries conducted by TH expert using initial analytics means;
- stage of the hypothesis creation enriched by the knowledge of the subject area, case awareness (an individual situation description, identification of key issues of a case, analyzing a case using relevant theoretical concepts from TH, learning problem-solving case studies) and cyber threat intelligence as discussed in paragraph 4.1;
- preliminary data triage – an important stage in the analytic process, which involves the tasks of ruling out the noise in the raw data, and identifying and grouping the data indicating the suspicious events worth of further investigation [17], and prioritization before conducting the research stage;
- stage of the research using tools and techniques enriched by the classic search, machine learning and linked data method as discussed in paragraph 4.2;
- stage of the identification of the TTPs enriched by the trusted frameworks in the field, as described in paragraph 4.3;
- detailed outputs of the expansion of analytics stage, such as: a default search for regular execution, tuning a controlled machine learning algorithm, a new compromise indicator, a new SIEM rule for reactive detection;
- incidents resolving plan as an ultimate goal of the incident management process, which is an integral part of the organization's overall information security management process;
- metrics or Key Performance Indicators (KPIs), such as: accuracy, timeliness, completeness and those described in more detail in paragraph 4.5 for the continuous performance management of the TH process.

6. Conclusions

In this paper the structuring of a new approach to countering cyber threats – Threat Hunting, was presented.

This study shows that TH is an effective method of modern cyber threat countering for organizations at a high level of maturity, that already have their own SOC or use such services through outsourcing and are ready to move to the level of proactive threat response.

The main purpose of the paper is achieved – a functional diagram of Threat Hunting approach is proposed and the following objectives are solved:
- enrichment of the existing TH model with SOC analytics tools, queries improvement and continuous performance measurement is proposed;
- a triad of the most important SOC tools is proposed for the effective implementation of the analytics stage;
- the process of converting raw logs and telemetry from network and infrastructure assets of the organization into structured queries for conducting TH is proposed and clarified;
- the place of cyber threat intelligence at the stage of hypothesis creation is determined as a small part of the whole process of TH, which helps to disseminate effective information about threats by understanding the motives, capabilities and methods of attackers to inform about measures to reduce cybersecurity;
- the examples of effective metrics or KPIs to evaluate TH are proposed;
- the incidents resolving plan is proposed as an ultimate goal of the TH, which helps to fit it into the existing process of incident management process, which is an integral part of the organization's overall information security management process.

7. Acknowledgements

8. References

[1]   R. van Os, M. Bakker, R. Bouman, "TaHiTI Threat Hunting Methodology", FI-ISAC NL Publication, version 17.12.2018, 2018. URL: https://www.betaalvereniging.nl/wp-content/uploads/DEF-TaHiTI-Threat-Hunting-Methodology.pdf.

[2]   "Detecting the Unknown: A Guide to Threat Hunting, Home Office Digital, Data and Technology, version 2.0, 2019. URL: https://hodigital.blog.gov.uk/wp-content/uploads/sites/161/2020/03/Detecting-the-Unknown-A-Guide-to-Threat-Hunting-v2.0.pdf.

[3]   A. Chuvakin, "How to Hunt for Security Threats", Gartner, Inc., 2017. URL: https://www.gartner.com/en/documents/3664330.

[4]   D. Akacki, "5 types of Threat Hunting", Sqrrl Data URL: https://sqrrl.com/5-types-threat-hunting/

[5]   D. Akacki, D. Bianco, R. Bejtlich, "Huntpedia: Your Threat Hunting Knowledge Compendium", Sqrrl Data, 2018. URL: https://www.threathunting.net/files/huntpedia.pdf.

[6]   *White paper: A Framework for Cyber Threat Hunting, Sqrrl Data,* 2018. URL: https://www.threathunting.net/files/framework-for-threat-hunting-whitepaper.pdf.

[7]   R. Lee, D. Bianco, "Generating Hypotheses for Successful Threat Hunting", SANS Institute, 2016. URL: https://www.sans.org/reading-room/whitepapers/threats/generating-hypotheses-successful-threat-hunting-37172.

[8]   *MITRE ATT&CK Framework,* 2020. URL: https://attack.mitre.org/tactics/enterprise/

[9]   S. Schmitt, F. I. Kandah and D. Brownell, "Intelligent Threat Hunting in Software-Defined Networking", *2019 IEEE International Conference on Consumer Electronics (ICCE),* Las Vegas, NV, USA, 2019, pp. 1-5, doi: 10.1109/ICCE.2019.8661952.

[10] M. Iavich, S. Gnatyuk, E. Jintcharadze, Y. Polishchuk, A. Fesenko and A. Abisheva, "Comparison and Hybrid Implementation of Blowfish, Twofish and RSA Cryptosystems", *Proceedings of 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON),* Lviv, Ukraine, 2019, pp. 970-974.

[11] M. N. S. Miazi, M. M. A. Pritom, M. Shehab, B. Chu and J. Wei, "The Design of Cyber Threat Hunting Games: A Case Study", *2017 26th International Conference on Computer Communication and Networks (ICCCN),* Vancouver, BC, 2017, pp. 1-6, doi: 10.1109/ICCCN.2017.8038527.

[12] K. Wafula and Y. Wang, "CARVE: A Scientific Method-Based Threat Hunting Hypothesis Development Model", *2019 IEEE International Conference on Electro Information Technology (EIT),* Brookings, SD, USA, 2019, pp. 1-6, doi: 10.1109/EIT.2019.8833792.

[13] S. Gnatyuk, V. Sydorenko, A. Polozhentsev, A. Fesenko, N. Akatayev, G. Zhilkishbayeva, "Method of cybersecurity level determining for the critical information infrastructure of the state", *CEUR Workshop Proceedings,* vol. 2616, pp. 332-341, 2020.

[14] H. Haddadpajouh, A. Mohtadi, A. Dehghantanaha, H. Karimipour, X. Lin and K. -K. R. Choo, "A Multi-Kernel and Meta-heuristic Feature Selection Approach for IoT Malware Threat Hunting in the Edge Layer", in *IEEE Internet of Things Journal,* doi: 10.1109/JIOT.2020.3026660.

[15] O. Oksiiuk, V. Chaikovska, A. Fesenko, "Security technique for authentication process in the cloud environment", *Proceedings of 2019 IEEE International Scientific-Practical Conference: Problems of Infocommunications Science and Technology, PIC S and T 2019,* October 2019, 9061248, pp. 379-382.

[16] A. N. Jahromi, S. Hashemi, A. Dehghantanha, R. M. Parizi and K. -K. R. Choo, "An Enhanced Stacked LSTM Method With No Random Initialization for Malware Threat Hunting in Safety and Time-Critical Systems", in *IEEE Transactions on Emerging Topics in Computational Intelligence,* vol. 4, no. 5, pp. 630-640, Oct. 2020, doi: 10.1109/TETCI.2019.2910243.

[17] C. Zhong, T. Lin, P. Liu, J. Yen, K. Chen, "A cyber security data triage operation retrieval system", Computers & Security, Volume 76, 2018, 12-31, ISSN 0167-4048, DOI: 10.1016/j.cose.2018.02.011.