# MATCHING EXPECTATIONS AND REALITY IN AI SYSTEMS - CYBERSECURITY USE CASE

by

Tala Defo Aymar

A Project Paper submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

**Master of Information Technology Security**

Faculty of Business and Information Technology

University of Ontario Institute of Technology (Ontario Tech University)

Oshawa, Ontario, Canada

April 2023

## PROJECT PAPER REVIEW INFORMATION

Submitted by: **Tala Defo Aymar**

**Master of Information Technology Security**

| Project/Major Paper title: |
| --- |
| MATCHING EXPECTATIONS AND REALITY IN AI SYSTEMS<br>- CYBERSECURITY USE CASE |

The Project Paper was approved on April 6, 2023 by the following review committee:

**Review Committee:**

| | |
| --- | --- |
| Research Supervisor | Dr. Peter Lewis |
| Second Reader | Dr. Khalil El-Khatib |

The above review committee determined that the Project Paper is acceptable in form and content and that a satisfactory knowledge of the field was covered by the work submitted. A copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

## ABSTRACT

Artificial intelligence (AI) is a growing field in computer science which develops intelligent systems capable of performing things that a human mind can do. The manufacturers of security systems integrate AI capabilities into their systems for threat hunting, and market them with an emphasis on AI used to provide security features. This study evaluates the expectations of marketed AI features with reality in a use case of a cybersecurity system. To this end, we evaluated a system in a real-live environment with huge amount of data sent to it for analysis. Our evaluation demonstrates that, first, the virtual security analyst feature provided by the system cannot replace a human security analyst as it can only perform 3 amongst the 8 tasks of a human security analyst. Secondly, marketing claims exaggerate regarding the features provided by AI in the system.

**Keywords:** AI; claims; threat; security; analyst

## AUTHOR'S DECLARATION

I hereby declare that this project paper consists of original work of which I have authored. This is a true copy of the work, including any required final revisions, as accepted by my committee.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this work to other institutions or individuals for the purpose of scholarly research. I further authorize the University of Ontario Institute of Technology (Ontario Tech University) to reproduce this work by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my work may be made electronically available to the public.

TALA DEFO Aymar

**TO MY FAMILY**

**ACKNOWLEDGEMENTS**

I would like to extend my first gratitude to my research supervisor, Dr. Peter Lewis, who accompanied me in this research project. His knowledge of artificial intelligence and his critical thinking pushed me out of my comfort zone and were essential to the success of this project.

This project has also been possible due to the exchanges and advice of Professor Stephen Marsh.

Special thanks to the contribution of Dr. Khalil El-Khatib for accepting to evaluate my work.

I would like to thank the IT Director of Collège Boréal, Mr. Benoit Bonin, who encouraged me for this master's degree studies.

I am also grateful to my professional supervisor, Mr. Pierre Trudeau, who allowed me to accommodate my studies and my work.

Special thanks to the Fortinet sales team, Mr. Lee Pecori, and Mr. Dave Hogan, who provided me with the FortiNDR licences evaluation and the support for the installation and configuration of the system.

I would like to recognize the contribution of the professors and teaching assistants who supported me during this master's degree.

I would like to acknowledge the contribution and feedback of the members of the Trustworthy AI Lab at the Ontario Tech University during the meetings.

This achievement would not have been possible without the support of my daughters, Orlanne, Cheryl, Mercedes, and Oprah for their patience, understanding, and love for me. For the time I had to devote to this project instead of spending a moment with them.

Obviously, I am deeply indebted to Pascaline, my wife, my friend, for your patience and your encouragement which gave me bursts of energy to get to the end of this master's degree.

Thanks to all those who have contributed directly or indirectly to the production of this report.

Finally, I am grateful for the health and blessings that God granted me to complete my master's degree.

## TABLE OF CONTENTS

## LIST OF TABLES

### CHAPTER 3



### CHAPTER 4



## LIST OF FIGURES

### CHAPTER 3



### CHAPTER 4

# Chapter 1. Introduction

There is a clear distinction between what Artificial Intelligence (AI) is advertised as, what it is, could be, and even what it ought to be (Lewis et al., 2021, p. 8). We see many marketing spots, journals, or documents stating what AI or their AI system can do. For example, about the recent chatbot ChatGPT which can answer any question regarding any subject, Bass and Bloomberg (2023) indicate the chatbot is so prone to mistakes that their creator has made a public commitment to address the technology by reducing biases in the system. In the fields of cybersecurity, we observe similar statements about what the cybersecurity system built with AI is capable of doing in threat hunting. IBM (n.d.) on its website Artificial Intelligence for cybersecurity, states that "As cyberattacks grow in volume and complexity, Artificial Intelligence (AI) is helping under-resourced security operations analysts stay ahead of threats. Curating threat intelligence from millions of research papers, blogs and news stories, AI technologies like machine learning and natural language processing provide rapid insights to cut through the noise of daily alerts, drastically reducing response times" and "Cognitive security combines the strengths of AI and human intelligence. Cognitive computing with Watson for Cyber Security offers an advanced type of artificial intelligence, leveraging various forms of AI, including machine-learning algorithms and deep-learning networks, that get stronger and smarter over time". We can collect millions of statements like these regarding the capabilities of AI. These announcements are global and generic, and they do not take into consideration the peculiarity of each environment where the AI system will be used. The purpose of our study is to evaluate a cybersecurity system built with AI and compare the marketed features with practical results from a real-live environment.

We begin this report with a literature review in which we provide some background information about AI, the claims made on AI systems about what they can do. Additionally, we present what the researchers say about the discrepancies between the claims and what AI can really do, and the cause of these discrepancies. Finally, we discuss the use of AI in cybersecurity and how it relates to humans. Chapter 4 presents the cybersecurity AI system that we choose to evaluate, and the data collection methodologies used. We then study how AI technologies are built into the system and the features they provide. We continue by

presenting the results and discussing them in chapter 5. This chapter also provides comparison between a virtual security analyst and a “human” security analyst; and presents the gaps between AI marketing claims and our findings. Finally, in chapter 6, we conclude with a synthesis of our ideas and findings.

## Chapter 2. Review of Literature

Since the first half of the 20th century, AI has been identified as the technology of the future to replace human minds. Boden (2018) defines AI as technology that ”seeks to make computers do the sort of things that minds can do”. She continued by saying that minds can do things which are intelligent like reasoning and other things like vision that AI cannot do (p. 1). In addition, there have been many predictions and claims about AI and what it will do in our daily life and how it will replace human beings. Marcus & Davis (2020) presents some claims about AI capabilities from famous CEOs and founders : Eric Schmidt from Google declared that climate change, poverty, war, and cancer will be fixed by AI; Peter Diamandis, XPRIZE founder, also claimed in his book Abundance that, AI will certainly propel us into the high level of abundance; finally, Sundar Pichai from Google, claimed that AI is one of the most relevant things mankind is creating, deeper than electricity or fire (p. 5). Some examples from Marcus & Davis (2020, p. 5) of AI implementations that have not been achieved as expected include for example, the fully autonomous car which was supposed to drive without human intervention and still needs human as a safety backup; The IBM Watson AI system used in health care which was supposed to provide solutions for diagnosis and treatment of some known rare diseases such as cancer did not have the expected outcome; Another example is the M chatbot from Facebook which was created to use AI technology to help people to make reservations and order items on the Internet which failed and needed more human intervention to accomplish the task. We can have many of such examples of AI implementations that did not accomplish as expected. If we want computers to do everything our minds are capable of, to what level do we trust AI to do thinks like human minds can do? We see that there is a difference between these marketing assertions from famous CEOs and founders and the reality in the implementations of AI. Lewis et al. (2021) indicate that there is a clear difference between AI as artificial minds and AI as marketing speech which is more a perception of what AI can do from the marketing angle. Therefore, trusting AI is not only based on marketing claims. Pieters (2011) defines trust as a form of self-assurance, and it involves relying on something else and believing that this thing will not miss to meet certain expectations. He continues by adding that risks and alternatives must be evaluated to provide self-assurance and a decision must be made by the person using AI to rely on it or not. Based on that,

people trust AI depending on the achieved expectations and then decide to trust it or not. To go further, Malle et al. (2020) present trust as an assurance that the product (AI in our case) is predictable, authentic, qualified, honest, or ethical. AI is expected to be trustworthy, whatever the solutions or implementations, even in cybersecurity for threat intelligence and detection which is the subject of our study.

Despite the widespread use of AI technologies today, they encounter some limitations. One limitation as identified by Marcus and Davis (2020) is the difficulty or the unlikelihood of AI to adapt in an open-ended world in which possibilities are infinite, data generated will never be sufficient, the world is constantly changing, and no data can exactly mirror these changes (pp. 15-16). And since AI is based on the quality of data input, we cannot have all cases in advance with the AI. It may be difficult for AI to identify and provide logical connections in some complex situations in an open-ended world, and consequently made it unable to reuse the knowledge acquired in previous contexts. This is because AI still have many limitations in the area of evaluating different situations, foreseeing the likely future, and choosing dynamically as situations change in an open-ended world (Marcus & Davis, 2020, p. 113). This limitation to adjust when situations changed is a result of AI learning only based on how the programming is done. We do not know if AI technology will ever reach the maturity level to collect and process the infinite situations in an open-ended world like the human minds, but industries and researchers are working towards that maturity.

This limitation may also occur when using AI for threat intelligence, as it may happen that some patterns related to new unknown malicious activities may not be identified by the AI and not recognized as a cyberthreat. Another limitation as per Marcus & Davis (2020) is related to the moral and ethic of AI (pp. 193-194). For example, Asimov's laws are sufficient for robots (which are AI systems) in several simple ethical decisions in day-to-day situations. But how AI could apply ethical values to do the right things in ordinary circumstances when situations become complex as in an open-ended world?

We can see in the first part of our review that there are many claims regarding AI, now we want to understand how those claims are verifiable. Marcus and Davis (2020) remind us that IBM claimed its AI system Watson will transform healthcare with the latest

improvement in cognitive computing by accomplishing more than what has ever been done (pp. 4-5). This assertion prepares us for something revolutionary, but it does not indicate what exactly, and people can fill it with whatever possible. Another claim presented by Marcus and Davis (2020) and collected from the economic journal Quartz about one of the revolutionary projects of Google, Google Talk to Books, by Ray Kurzweil, futurist and inventor and director of engineering at Google, saying that "Google's astounding new search tool [that] will answer any question by reading thousands of books" (pp. 68-69). We realize that this claim cannot be taken exactly as it is because AI is not at this level of maturity of reading books. Another marketing claim related by The Economic Times (2021) regarding the use of AI by stockbroker on their online platform was refuted by the CEO of the company by indicating that it was a perfect marketing stratagem. We can conclude that there is a discrepancy between the claims and the reality. We are going to investigate the causes of these discrepancies; whether it is the wrong input data provided to the AI system, or simply the claims are beyond the limit of what is possible.

We can begin to have an answer to the causes of these discrepancies from Azimi and Pahl (2021) who did a study on the effect of the completeness and correctness of input data on a Machine Learning model in an IoT (Internet of Things) environment. Their results show that the quality of input data, that is correctness and completeness have an impact on the model construction and data quality of the ML model, and thus on the outputs provided by the AI system. In addition, Grosse (2018) in one of his lectures on ML indicates that during the training phase of a ML model, it is important to generalize the model to data beyond the training data. Then generalization must be measured by dividing the data into three batches: a training batch which is a batch of training examples on which the network is trained; a validation batch which is used to adjust hyperparameters such as the number of masked units, or the learning rate; and a test batch which is used to evaluate the generalization performance. Finally, Redman (2018) from the Harvard Business Review declares that machine learning needs the right input data. First, during the training of a predictive model, it must be accurate, labeled perfectly, and secondly during the implementation phase.

Now we come to ask ourselves if the claims are beyond the limit of what is possible. For Marcus and Davis (2020), the claims are beyond the limit of what is possible, as they declare that "One reason that people often overestimate what AI can actually do is that media reports often overstate AI's abilities, as if every modest advance represents a paradigm shift" (p. 6). Presently, AI is limited because they are programmed for specific tasks. They continue by saying that Narrow AI alone is not sufficient and to bring AI to the next level, we need a broad intelligence approach with AI, that is an intelligence that is capable of readjusting in an open-ended world.

One of the various implementations of AI in cybersecurity is in threat intelligence which refers to "the set of data collected, assessed and applied regarding security threats, threat actors, exploits, malware, vulnerabilities and compromise indicators" (Conti et al., 2018, p. 2). It assists security professionals in identifying the indicators of violation, taking out detailed information about the violation strategy and reacting to the violation in a correct and appropriate way. In addition, it collects data from various sources and uses analytical techniques to mine, translate and extract security facts. Kaloudi and Li (2020) indicate that AI is effective in cybersecurity as it gathers massive amounts of data and then rapidly filtrates them to uncover malicious patterns and abnormal behaviors. Accordingly, AI in threat intelligence is used for detection, response, and mitigation of the following threat's examples: malicious activities and behaviours, advanced persistent threats (APT), and zero-day attacks. AI in cyber threat intelligence is facing the same limitations as described above. For the Information Systems Audit and Control Association (ISACA) (2021), in its journal Emerging Technology, the overuse of AI terms is amongst the significant challenges of AI in cybersecurity business (p. 11). ISACA continues by indicating that cybersecurity vendors and service providers development have been obstructed in some respects and their true innovation in the field has been disoriented by the marketing claims, and it becomes difficult for business leaders to choose the right cybersecurity solutions that add values to their business because of these marketing noises (p. 12). To continue with the fact that AI is used to assist security professional, Stolte (2018) declares that AI is a perfect addition to security investigators as it can serve in Security Operations Center (SOC) to analyze bunch of data faster than human can do and present the results to the investigators in prioritized order even depending on risks. Amongst different research

relating AI and human in cybersecurity, Nguyen and Choo (2021) come up with the conceptual framework called 'Human-in-the-Loop Explainable-AI-Enabled Vulnerability Detection, Investigation, and Mitigation' (HXAI-VDIM), in which human security expert and AI work together to create a process for vulnerabilities detection, investigation and mitigation, and AI supports human intelligence to rapidly refine important components that are reused by the AI to have a successful and adequate process. In the Pillsbury (2021) report about AI and Cybersecurity, AJ Abdallat, CEO of Beyond Limits says that "we champion a hybrid approach of AI to gain trust of users and executives as it is very important to be able to have explainable answers" (Pillsbury, 2021, p. 18); Using AI alone will bring answers but will not demonstrate how these answers were found or produced. They continue by saying that we need to combine both human and AI, as humans can assess and adapt the technology based on real things happening (p. 18).

In summary, the literature review indicates that the level of trust in AI depends on what it really does and not on marketing assertions. In addition, the review shows that one limitation of AI is the hardness to adapt in a situation where there are innumerable possibilities and all types or sets of input data will never be sufficient to train the AI system. Lastly, the review specifies that there is a divergence between AI marketing allegations and the reality. This divergence comes first from the quality of input data on the ML model and secondly on the real capability of AI which is far less than the marketed capabilities. These findings from the literature review also apply to AI implementations in cyber security, specifically in threat intelligence. AI is used in threat intelligence to identify malwares of all sorts. It assists human intelligence by collecting huge amounts of data from different sources, aggregates and screens them to detect, investigate, and mitigate malicious or abnormal activities.

# Chapter 3. Cybersecurity AI solution and Data Collection Methods

This chapter presents the cybersecurity solution we choose for the evaluation and the AI technology underneath, followed by the description of the methods and tools used in our study. The selected cybersecurity system used AI technologies to detect malware and provides a virtual security analyst (VSA) feature that we are going to study in detail. Finally, the methods and tools also described the demo environment we built for our study with the data collection approaches used.

## 3.1 The Cybersecurity AI System evaluated

The cybersecurity AI system evaluated is FortiNDR (Network Detection and Response) from Fortinet. The description from the datasheet indicates that ”FortiNDR represents the future of AI-driven breach protection technology, designed for short-staffed Security Operation Center (SOC) teams to defend against various threats including advanced persistent threats through a trained VSA that helps you identify, classify, and respond to threats including those well camouflaged” (Fortinet, 2022a, p. 1). FortiNDR was formerly FortiAI and it “employs patent-pending Deep Neural Networks (DNN) based on Advanced AI and Artificial Neural Network (ANN) to provide sub-second investigation by harnessing deep learning technologies that assist you in an automated response to remediate different breeds of attacks” (Fortinet, 2022a, p. 1). In addition to that Fortinet (2023b) declares that FortiNDR uses FortiGuard security services like Artificial Neural Network updates and other NDR related updates in conjunction with local machine learning and cloud machine learning to identify attacks. Bringing security to the application content, web, device, user, and cloud, FortiGuard security services is a group of industry-leading, AI-enabled security capabilities. It continuously evaluates the threats and automatically modifies the Fortinet security fabric and ecosystem for protection. It offers coordinated and reliable real-time protection against the most recent assaults on network endpoints and clouds.

### *3.1.1 FortiAI - Deep Neural Networks and Artificial Neural Network used by FortiNDR*

FortiAI uses Deep Neural Networks (DNN), an approach of Artificial Neural Network (ANN) for malware detection named “Malware Identification Using Multiple Artificial Neural Networks” developed by Yang Xu from Fortinet and under a patent pending serial

No. 16/053479. In the solution brief document of FortiNDR, Fortinet (2022b, p. 1) defines ANN as a popular ML technique which uses hardware and software to build configurations that mimic how neurons in the human brain work through ML training. The model is continuously fed with a large amount of information, and the system analyzes this information to adjust the algorithm based on new tactics and capabilities employed by malware or attack vectors. They continue by saying that DNN is a ML approach in which multiple ANN with two or more layers between the input and output layers are used to create sophisticated and non-linear connections.

FortiAI uses the workflow presented in figure 1 below to detect malware. The inventor Yang Xu (2020) from Fortinet in the patent document indicates that a supervised learning process is carried out on each training set (including malware samples and benign samples), as follows:

(i) Produce a plurality of code blocks of assembly language instructions by decomposing the machine language instructions contained in the training sets. When FortiNDR receives files or input data, machine language instructions contained in the file are disassembled or de-obfuscated and converted into code blocks or a plurality of code blocks of assembly language instructions. The patent defines a code block as a group of disassembled instructions that can be found, for instance, by analyzing an executable file for certain code block delimiters like jump instructions, conditional jump instructions, and call instructions. A code block can just be a series of instructions with a predetermined and/or customizable number of instructions in it.

(ii) Separate dynamic features corresponding to each code block by running each code block in a virtual environment. A single file can contain an average of 3000 code blocks.

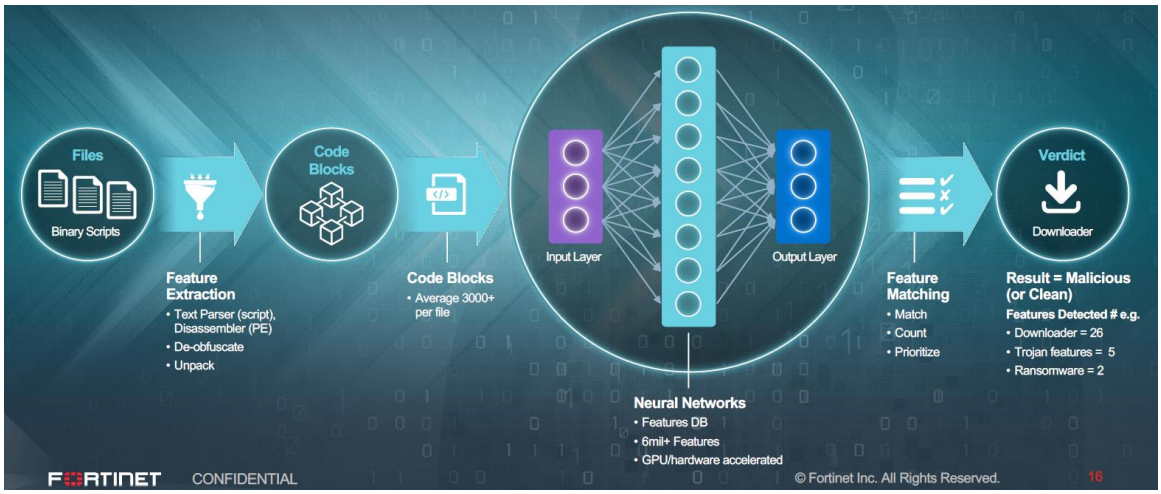
(iii) Provide as input to the ML, each block of codes into a first neural network and the associated dynamic features into a second neural network. Fortinet (n.d., p. 7) defines a feature as a specific quantitative quality or trait of an observed phenomena. The neural networks contain a Features DB with more than 6 million

features, including clean and malicious to be identified on a file. These DBs are downloaded on a regular basis from FortiGuard to FortiNDR.

(iv) Update the neural network weights and biases based on whether the training set were infected or not by malware.

(v) After processing a given or configurable number of training sets, the neural networks criticize each other, unify the various weights and biases by swapping and adapting them appropriately. The first neural network and the second neural network update their weights and biases based on whether the training sample was malicious or benign, and after processing a predetermined or configurable number of the plurality of training samples, the first neural network and the second neural network criticize each other and combine their respective weights and biases by exchanging and adapting their respective weights and biases appropriately.

(vi) The output from the ML is either the file is matching one or more features, the counts, and the priority level (critical, high, medium, low), which results in a verdict whether the file is clean or contains a malware.




Source: FortiAI Next Generation Network Security - Artificial Intelligence in Cyber Security Document

*Figure 1. FortiAI malware detection workflow*

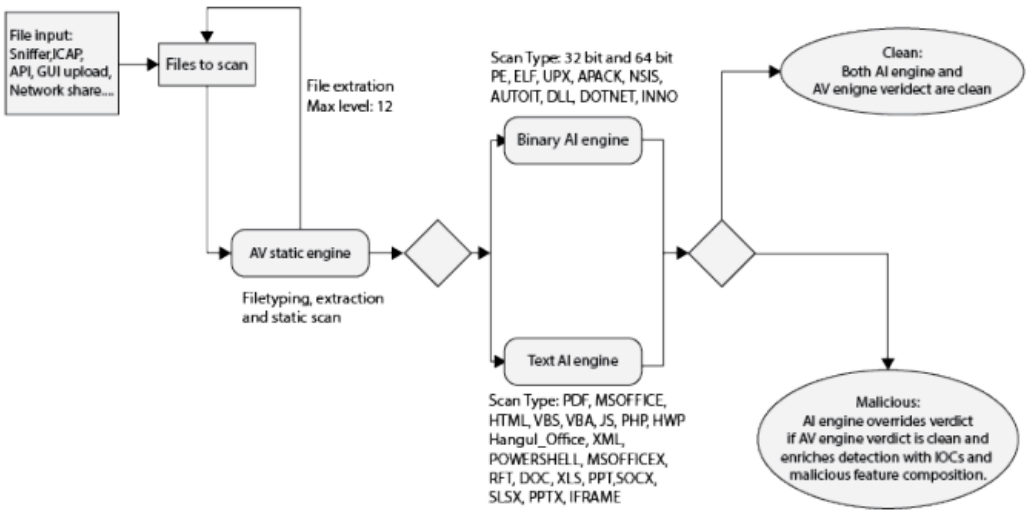
FortiNDR is pretrained in FortiGuard labs with about 20 million of files containing clean and infected files; and when used by the customer, it continues to receive ANN update from FortiGuard to maintain with updated features and latest threats. The aim of this

training is to provide the highest detection rate or lowest false positive rate when it comes to detecting malwares.

### *3.1.2 FortiNDR file scan flow*

Fortinet (2023a, pp. 17-18) in the administration guide of FortiNDR, indicates that the scan follows a two-stage flow. At the first stage as indicated in figure 2, the scan is done by the Antivirus static engine (i.e., AV static engine). The file types are recognized by the AV engine which gives a verdict simultaneously. The archive files like ZIP or TAR are decompressed (up to 12 layers of decompression) and returned to the antivirus static engine for scanning.

Stage 2 scan occurs if the files are supported by the ANN as indicated in figure 2, they are sent to the Binary and Text AI engine for scanning despite the verdict in stage 1. The AI engine will reverse the verdict only in case the file is clean in stage 1 and virulent in stage 2. The types of files that are supported by ANN (i.e., the ML) are scanned by both the ANN engines and the AV engines. The other supported file types are scanned by the AV engine only.




Source: FortiNDR 7.1.0 administration guide

*Figure 2. FortiNDR file scan flow*

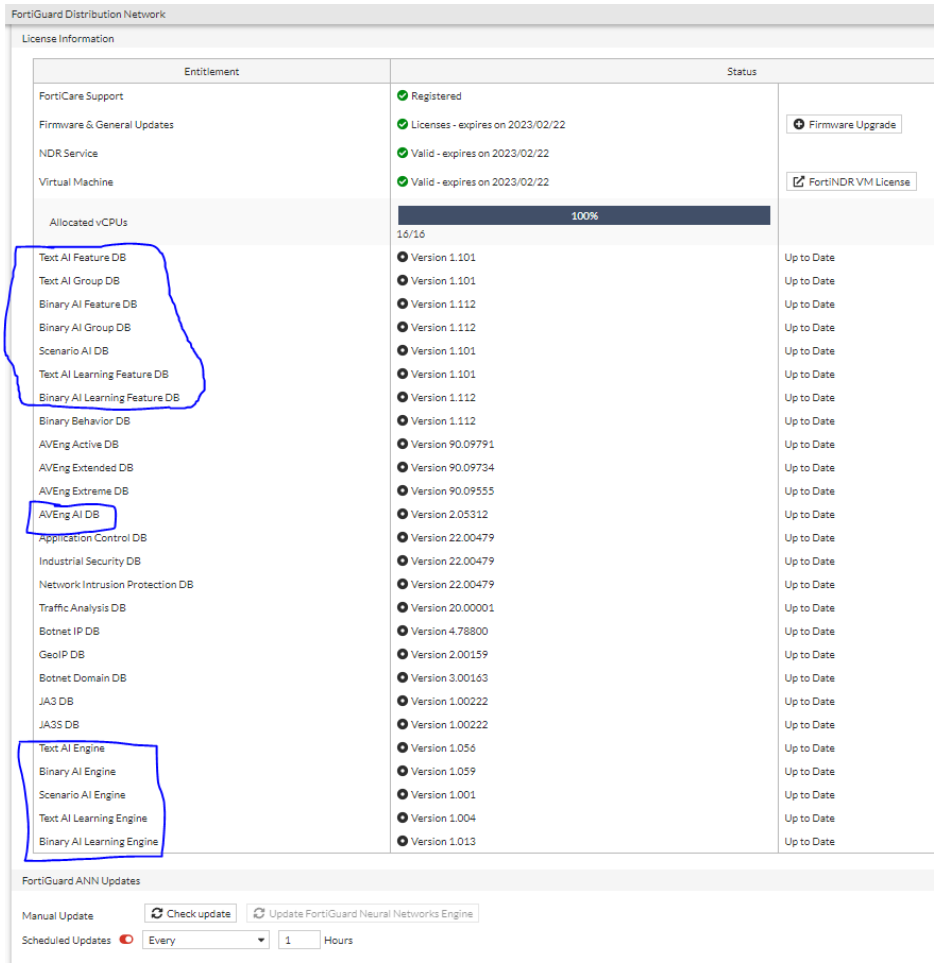



*Figure 3. Engines and databases in the FortiNDR*

Figure 3 shows an extract of the engines and databases in the FortiNDR that are used to detect malwares. We see the various features DB and AI engines that are updated on a regular basis from FortiGuard. These components of FortiNDR, the 7 features DB and the 5 engines are based on AI. We did not find any information related to each of these AI related components from Fortinet resources, but from our own research and analysis, we were able to provide the description contained in the table below:

| **Component name** | **Description** |
|---|---|
| Text AI Feature DB | Database of features related to the following text file types: PDF, MSOFFICE, HTML, VBS, VBA, JS, PHP, HWP, Hangul_Office, XML, Powershell, MSOFFICEX, RFT, DOC, XLS, PPT, SOCX, SLSX, PPTX, IFRAME |
| Text AI Group DB | Database of groups of features or feature combinations related to text file types |

| | |
|---|---|
| Binary AI Feature DB | Database of features related to the following binary file types: 32-bit and 64-bit PE, ELF, UPX, APACK, NSIS, AUTOIT, DLL, DOTNET, INNO |
| Binary AI Group DB | Database of groups of features or feature combinations related to binary file types. |
| Scenario AI DB | Database of attack scenario that are used to identify, build, and feed the attack scenarios menu in the FortiNDR. |
| Text AI Learning Feature DB | Database of features used by Text AI Learning Engine to detect and match anomalies in the text file types. |
| Binary AI Learning feature DB | Database of features used by Binary AI Learning Engine to detect and match anomalies in the Binary file types. |
| AVEng AI DB | Database used by the Antivirus engine |
| Text AI Engine | AI engine used to scan the text file types |
| Binary AI Engine | AI engine used to scan the binary file types |
| Scenario AI Engine | AI engine used to detect and build attack scenario |
| Text AI Learning Engine | AI engine used by ML to profile data and detect anomalies in text type files. |
| Binary AI Learning Engine | AI engine used by ML to profile data and detect anomalies in binary type files. |

Table 1. *Description of each AI related component in FortiNDR*

As indicated in the administration guide, this is the list of text file types: PDF, MSOFFICE, HTML, VBS, VBA, JS, PHP, HWP, Hangul_Office, XML, Powershell, MSOFFICEX, RFT, DOC, XLS, PPT, SOCX, SLSX, PPTX, IFRAME; and binary file types are: 32-bit and 64-bit PE, ELF, UPX, APACK, NSIS, AUTOIT, DLL, DOTNET, INNO

### 3.2 Demo environment and Data collection

Another important part of our work is to build an environment in which we can test and evaluate the claims' features from the marketing document (i.e., the datasheet). In this part, we first described the demo environment that we put in place to test some of the claimed features, and secondly, we present how we collected the data to be used for testing these features.

#### *3.2.1 Architecture for the Demo*

We installed our FortiNDR as a virtual appliance in a common network environment, connected to a core network switch acting as a central point receiving all ingress and egress network packets. We also have a remote site with a FortiGate firewall, connected to the central site through the internet using site-to-site VPN connection. The FortiGate firewall

is configured in integrated mode as per the Fortinet administration guide (Fortinet, 2023a, pp. 28-29), and it sends all files to the FortiNDR for analysis.

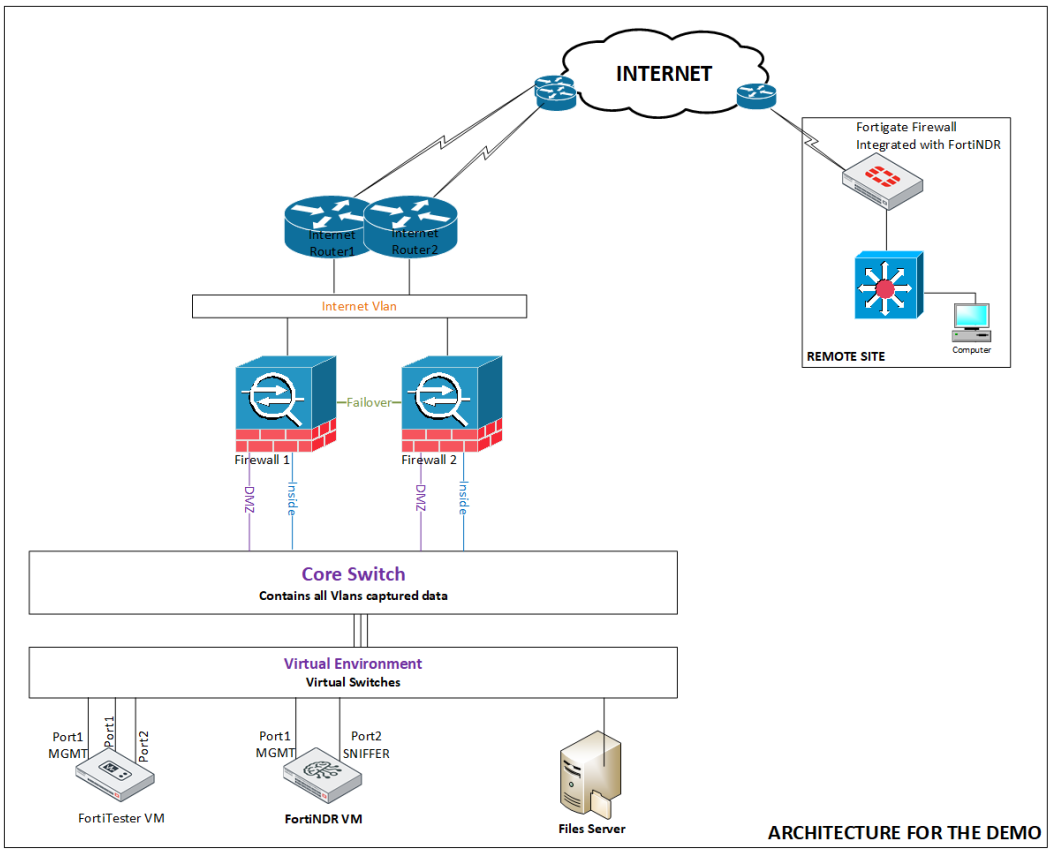



*Figure 4. Architecture for the demo*

### *3.2.2 Data Collection*

The data collection methods vary depending on the type of test (or features) we want to perform. The following methods are used: the first mode is with the FortiGate integrated to the FortiNDR in which all ingress and egress data going in and out of the remote site through the FortiGate. The second mode is by using port mirroring to capture all ingress and egress data from all the Vlans in the network.

#### *3.2.2.1 Through the FortiGate in integrated mode*

The first method of collecting data is through the FortiGate in the remote site. In this mode, FortiGate uses FortiNDR as a sandbox and interactively sends all files to it for analysis. By integrating the FortiGate with the FortiNDR for malware detection, both systems communicate using the Optimized Fabric Transfer Protocol (OFTP) over SSL to exchange

files to be analyzed. OFTP is a proprietary protocol used by Fortinet's device in its security fabric to communicate with each other. As per the Fortinet administration guide 7.1.0, we configured an AV profile in FortiGate which will send all files to FortiNDR for analysis (Fortinet, 2023a, pp. 28-29).

Finally, for our test scenario, we used a copy of the eicar test file that we copied from a computer in the remote site network to the file server in the central site. The eicar test file is an anti-malware test file that was created by the European Institute for Computer Antivirus Research (EICAR) Organization, to test the reaction of computer antivirus software in detecting malware (EICAR, n.d.). The communication between the remote site and the main site is through the FortiGate, the system will be able to intercept the file, analyse it and detect if there is malware.

*3.2.2.2 Using port mirroring*

The second method of collecting data for analysis is by using port mirroring from the core switch to the virtual machine. Interesting Vlans from the core switch were configured as source of the captured packets and the destination was the interface of the virtual switch where port 2 (SNIFFER) of the FortiNDR was connected. The traffic collected from the core switch is external traffic (internet and WAN) and internal traffic from internal Vlans.

### *3.2.3 Machine Learning baseline configuration*

The datasheet of the FortiNDR indicates that it uses "ML-based Traffic Profiling and Malware Detection" (Fortinet, 2022a, p. 1). ML is one of the core features used by the cybersecurity AI solution to profile traffic and detect malwares. The ML Configuration is part of the Virtual Security Analyst menu and can be used to view and modify the baseline machine learning features for detecting traffic anomalies as well as the baseline training status.

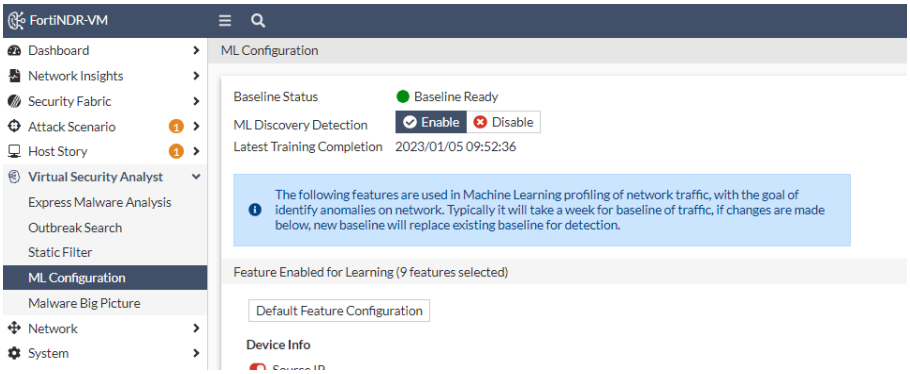



*Figure 5. Machine Learning configuration*

The ML feature is activated by default with a set of default parameters that can be changed depending on what you want to learn and analyze. When the configuration is set, it takes by default 7 days to train the ML engines and create a baseline of the network traffic.

# Chapter 4. Results and Discussions

This chapter presents and discusses the results obtained after putting the chosen AI system in a lab environment and injecting data in it. For each mode of integration or data collection approach, we present the results and examine how the AI technology is used to obtain them. In addition, we discuss the difference between a human security analyst and the virtual security analyst and finally examine other gaps between the claims from the marketing document (i.e., datasheet) and the reality obtained from our evaluation.

## 4.1 Results through the FortiGate in integrated mode

By using this method, the AI solution was able to detect a generic trojan in the infected file and built an attack scenario based on that as you can see in the figure 6 and 7 below.

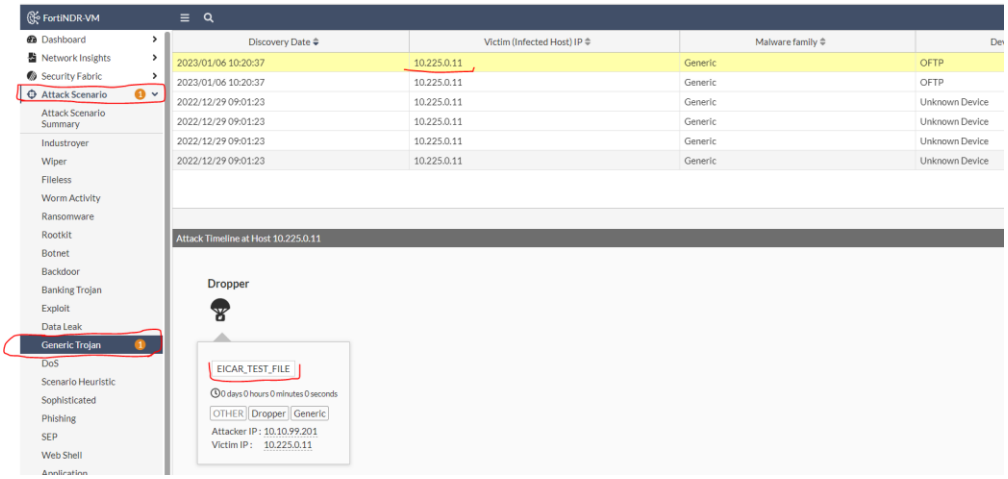



*Figure 6. Attack scenario detected for the eicar file*

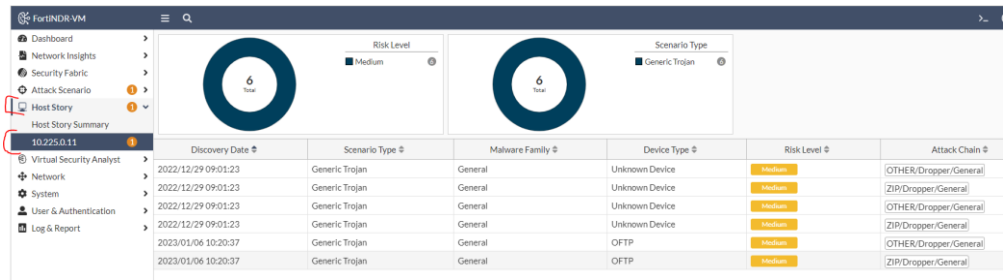



*Figure 7. Host history log for the victim machine infected by the eicar file*

The communication between the FortiGate and FortiNDR uses OFTP. The file scan flow in figure 2 gives no indication if files received through OFTP are scanned by the AI engines after the first phase of scan by the AV engine. Fortinet (2023a, p. 83) in the administration guide of FortiNDR, states that FortiNDR employs attack scenarios to recognise malware attacks and operate as your on-network personal malware analyst by categorising malware attack timings and circumstances scientifically. They continue by saying that most security tools can only inform you, in a vague manner that your network is infected with specific viruses, but FortiNDR goes further to explain in detail what the malware is attempting to accomplish, giving SOC analysts additional useful data for their inquiry and provides the severity levels range as critical, high, medium, or low. In addition, host story menu provides the same information as attack scenario with the difference that it presents malware assaults by host IP address while attack scenario arranges by attack type. For this mode, the AV engine scan the file, but we are unable to confirm whether the file was scanned by the AI engines. In contrast, the scenario AI engine is used to detect the malware in the file and organize it into attack scenario.

### 4.2 Results using port mirroring

Port mirroring was the mode of collection that provided most of the input that we used for our work. Initially, we changed the ML configurations from the default, and it takes 7 days for the ML engine to build a new baseline of traffic after a ML configuration change. The results obtained are based on data collected after the first month of evaluating the AI system. The ML configurations are presented in figure 8 below.

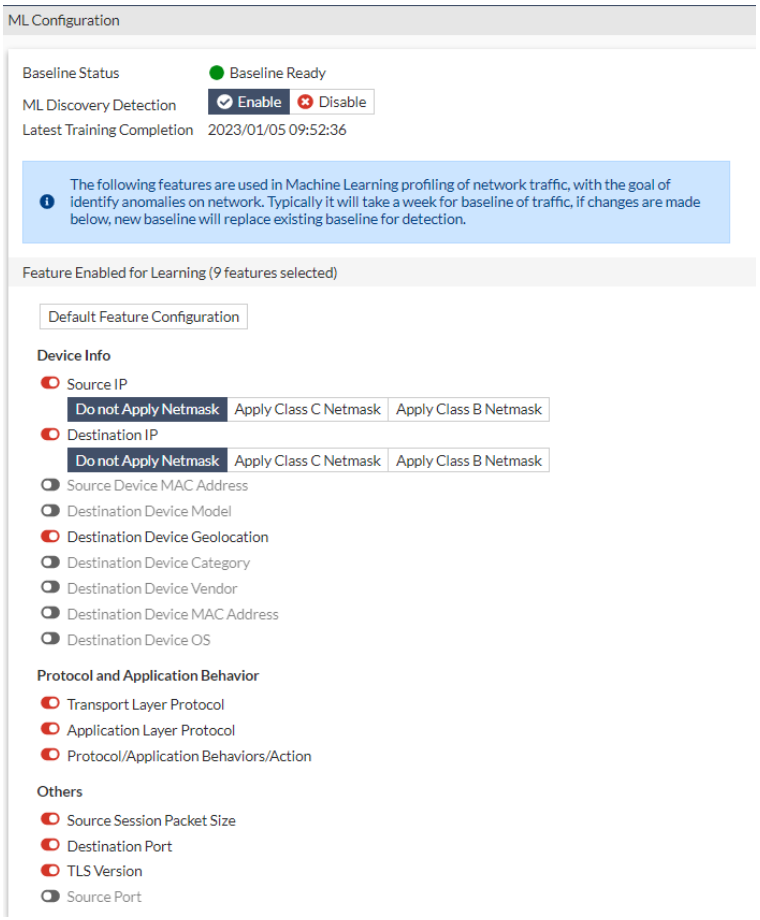



*Figure 8. Machine Learning baseline configuration*

### *4.2.1 Results from the ML training*

We present the results by following the menus of the FortiNDR which provide important outputs related to malware.

Figures 9 and 10 show the NDR overview and malware overview which are subsections of the dashboard menu. For the NDR overview, we will focus on the output from the virtual security analyst insights which include anomaly statistics, botnet connection, encrypted attack, anomaly overview, weak cipher or vulnerable protocol, network attack, and ML discovery. We see that the number of anomalies from ML discovery is different from those in sessions analysed.


*Figure 9. NDR Overview first phase and Virtual security analyst insights*

In our evaluation, malware overview only displays the malware detected in the integrated mode with FortiGate, and there is no occurrence of malware in the port mirroring mode used to feed the FortiNDR with data.

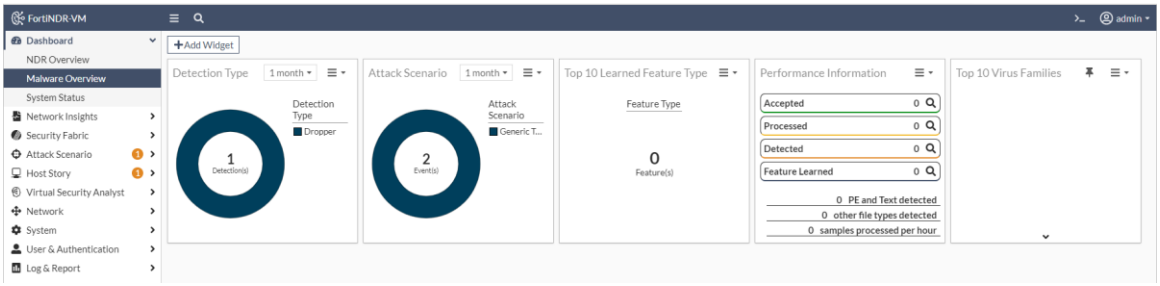



*Figure 10. Malware overview*

Followings in figures 11 to 15 are the sub-sections from the network insights menu which in some extent contains anomaly types and give details about the followings: botnet, FortiGuard Indicator Of Compromise (IOC), networks attacks, weak or vulnerable communication, encrypted attack, and ML discovery.

Botnet page shows botnets detected in traffic and its name is provided if known.

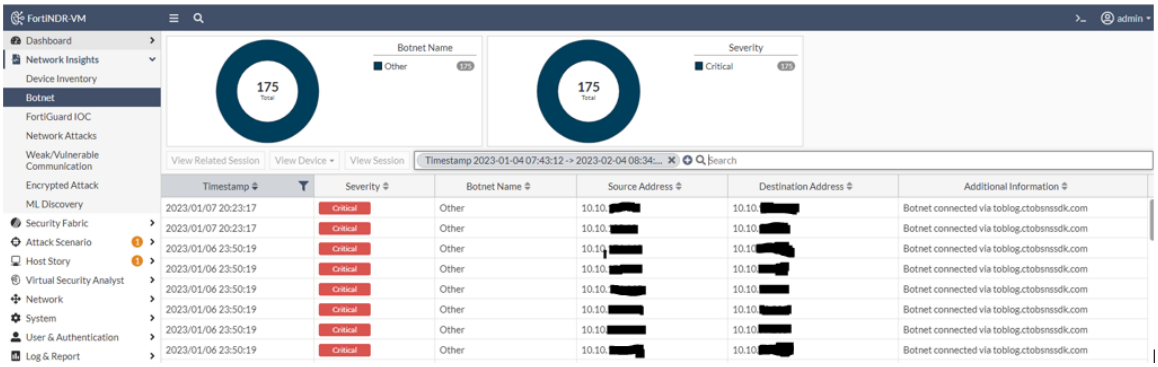



*Figure 11. Botnet detected*

FortiGuard IOC page displays doubtful URLs and IPs that FortiGuard flagged. This anomaly discovery is based on FortiNDR look up into FortiGuard IOC service. Our instance does not detect any anomaly of this type.

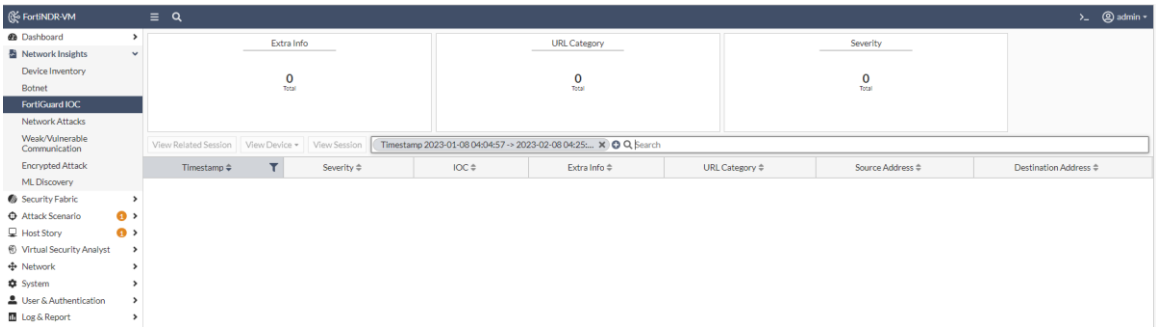



*Figure 12. FortiGuard IOC detected*

Network Attacks page displays known attacks identified by the database of Network Intrusion Protection

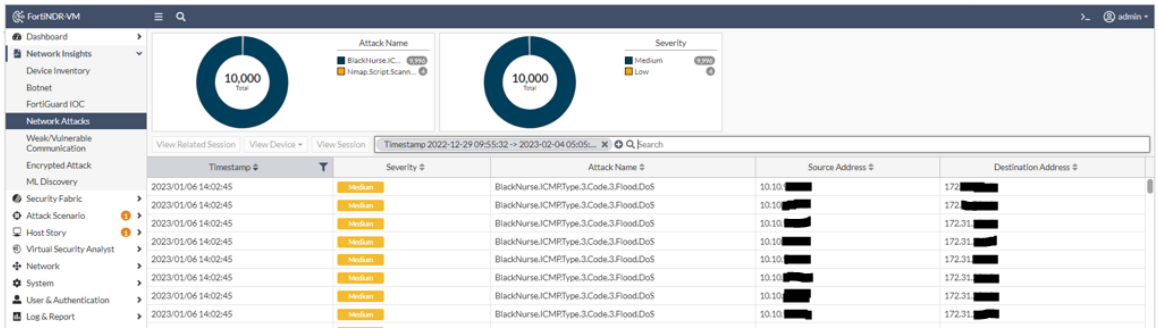



*Figure 13. Network attack detected*

The Weak/Vulnerable Communication page shows the list of weak or vulnerable communication uncovered by the sniffer port of FortiNDR. For instance, a weak cipher used by a previous version of TLS protocol.

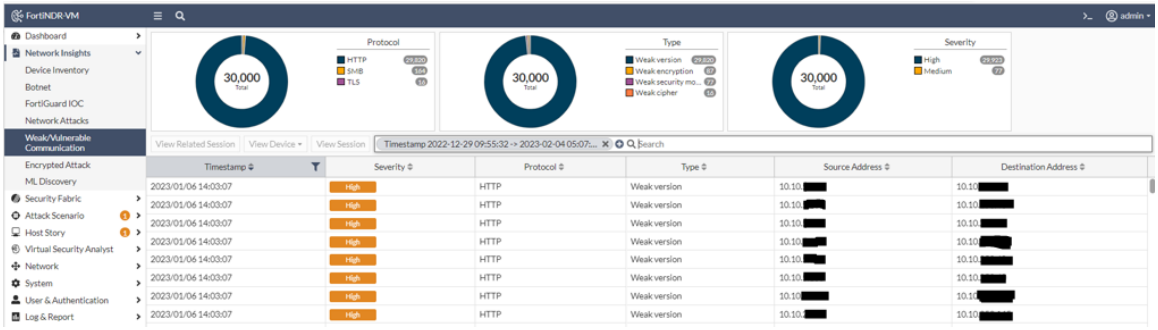



*Figure 14. Weak/vulnerable communication detected*

Encrypted attacks page displays encrypted attacks identified by analyzing JA3 hashes in TLS transactions. Both JA3 client and server SSL fingerprints are used by FortiNDR during detection.

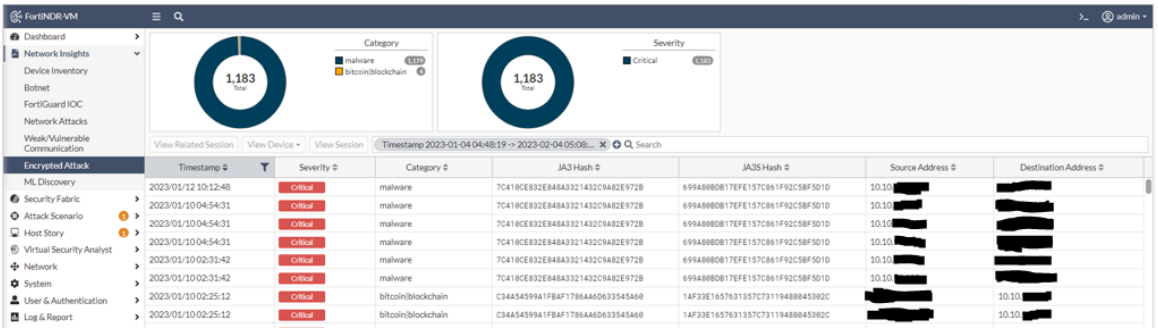



*Figure 15. Encrypted attack*

The ML Discovery page shows a list of anomalies identified by ML configuration as shown in figure 8.

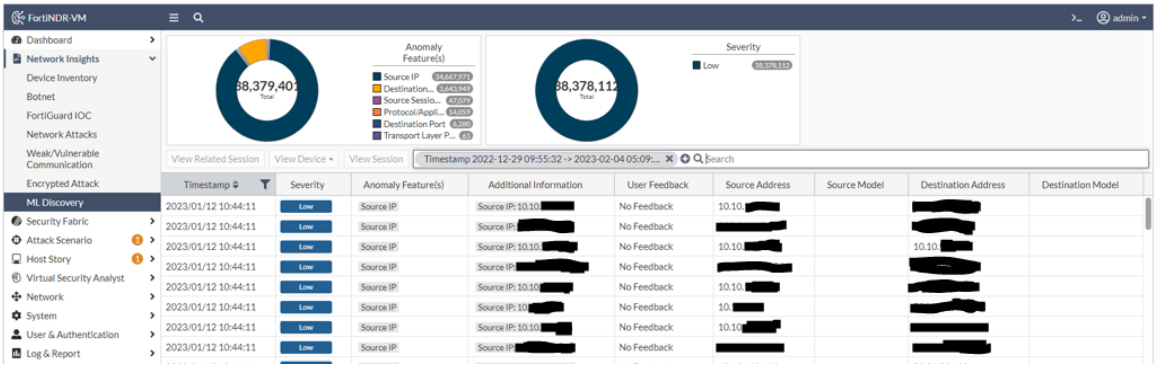



*Figure 16. ML discovery output*

### *4.2.1.1 Analysis of anomaly types*

This sub-chapter describes what FortiNDR considers as anomaly and identifies what types of anomalies are detected by FortiNDR and its ML specifically. FortiNDR groups traffic in two: normal traffic and anomaly traffic. NDR log as you see in figure 16 below collects anomaly traffic which can be one of the following types: botnet, encrypted attack, IOC campaign, network or intrusion attack, weak cipher or vulnerable protocol, or ML discovery. Yang Xu (2020) in the patent document describes malware as "software that is intended to do direct or indirect harm in relation to one or more computer systems. Such harm can manifest as the disruption or prevention of the operation of all or part of a computer system, accessing private, sensitive, secure and/or secret data, software and/or resources of computing facilities, or the performance of illicit, illegal, or fraudulent acts. Malware includes, inter alia, computer viruses, worms, botnets, trojans, spyware, adware, rootkits, keyloggers, dialers, malicious browser extensions or plugins and rogue security software." From this interpretation of what is a malware, we see that botnet, encrypted attack, IOC campaign, and network or intrusion attack are all malwares. Figure 17 contains two pages of the NDR log output, we see a clear difference between the anomalies detected by ML and the other types, this may be an indication that ML and therefore AI is not used to detect all malware types in the FortiNDR.

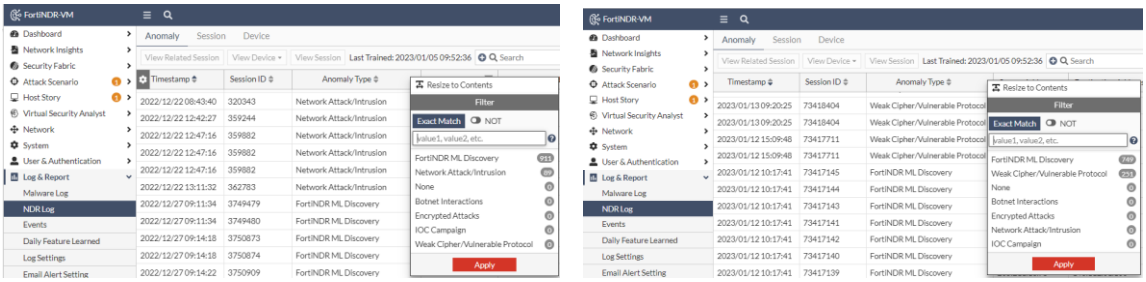


*Figure 17. Type of anomalies*

To explain further in figure 18, we selected two sessions containing attacks, session 10032332 on the left contains a network attack, and session 73258307 on the right contains an encrypted attack, both are not discovered by ML.

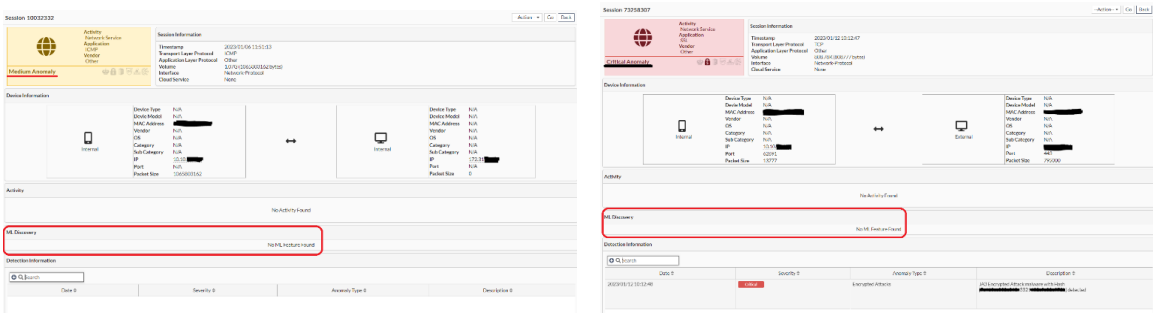


*Figure 18. Detailed session with anomaly not discovered by ML*

*4.2.1.2 Virtual security analyst menu*

The virtual security analyst menu in FortiNDR contains 4 pages, described below as per the administration guide 7.1.0 (Fortinet, 2023a, pp. 88-98):

- Express malware analysis: Use to submit a file to obtain the findings rapidly. We used it on the eircar.com sample infected file and we had the same verdict as in the case in chapter 5.1 earlier. We supposed the scan process follows the flow in figure 2.
- Outbreak search: uses tools to discover outbreak in the network. Logsign (2019) declares that an outbreak happens when a specific malware is detected on more than one device and/or system in our network.
- Static filter: used to handle allow hash list and block hash list.
- ML configuration: used to explore and rearrange the ML baseline features for the traffic anomaly identification, together with the status of the baseline training.

- Malware big picture: used for forensic analysis to evaluate harm to the network, and it may include details such as detection time, detection type and subtype. A capture from our instance in figure 19 below.

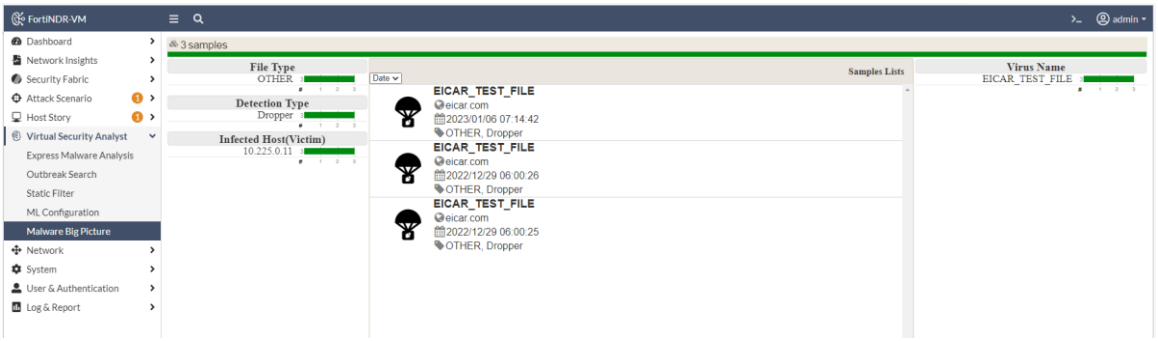



*Figure 19. Malware big picture*

### 4.3 Security analyst versus FortiNDR virtual security analyst

This sub-chapter presents the security analyst role and responsibilities first as human, and secondly as a virtual AI system with FortiAI technology. We continue by comparing and identifying the gaps between the two.

Fortinet (2022a, p. 1) in the FortiNDR datasheet indicates that it is "designed for short-staffed Security Operation Center (SOC) teams to defend against various threats". Therefore, the security analyst it refers to, is the SOC security analyst. As described by Checkpoint (n.d.), one of the leaders in information security industry, the responsibilities of a SOC team include the followings:

- Investigate possible incidents: SOC teams receive many alerts, but not all alerts indicate actual attacks, then SOC analysts are responsible for investigating possible incidents and determining whether they are real attacks or false positives.
- Triage and prioritize identified incidents: All security incidents are not identical, and organizations have limited resources to respond to incidents. As soon as they are identified, incidents should be triaged and prioritized to optimize resource usage and minimize business risk.
- Coordinate incident response: Utilizing several methods and engaging with numerous stakeholders is necessary while responding to an incident, this procedure

must be managed by SOC analysts to prevent errors from leading to a postponed or ineffective remediation.

- Maintain pertinence: SOC analysts must be equipped to handle the most recent threats to the organisation because the cyber threat landscape is always changing, and this entails keeping up with emerging and popular threats and making certain that security systems are equipped with the most recent set of rules to aid in their detection.
- Patch exposed or sensitive systems: Cybercriminals frequently use vulnerabilities in systems as a method of attack, so SOC analysts are in charge of finding, implementing, and testing patches for weak enterprise software and systems.
- Manage the infrastructure: new or up-to-date security solutions are necessary as the cyber threat landscape shifts and the enterprise network develops. The identification, deployment, configuration, and management of the security infrastructure are the responsibilities of SOC analysts.
- Address support tickets: SOC teams being under the IT department, SOC analysts may be asked to handle support tickets related to security from personnel needing assistance.
- Report to management: SOC teams must submit reports to management just like any other department because security is an integral element of the business. This calls for the capacity to persuade business audiences of the security costs and return on investment.

To emphasize on this, Salinas (2021) from Exabeam, another leader in the security industry, indicates 5 responsibilities of SOC analysts: deploy and handle security solutions; scrutinize malicious activities, mitigate, and preclude them; decrease downtime and guarantee business continuity; supplying security services to the remainder of the company; and support for audits and compliance. We see that, in addition to performing triage, investigation, and response to cybersecurity incidents, the SOC analyst provides prevention, tickets support, and reporting.

Checkpoint (n.d.), continues by raising the followings popular SOC challenges:

- ❖ Difficulty to staff important roles: There is a lack of talented security analysts in the cybersecurity industry. Therefore, the consequence is the struggling for organizations to engage and keep them to protect their assets against malwares or attacks.
- ❖ Eliminate false positives: Tens of thousands of alerts are sent to the SOC each day, yet only a small percentage of them are from legitimate threats. It takes precious time and resources for SOC analysts to find the needles in a sizable haystack of logs, alarms, and network traffic.
- ❖ Minimize impact on operations: Not every unusual activity on an organization's network is malware and part of an actual attack. Only factual attacks must be revealed and stopped by SOC analysts, to permit genuine business activity to pursue.
- ❖ Respond faster to attacks: The cost and damage to an organisation increase with the length of time an adversary has access to the network. To reduce the impact on the business, SOC analysts must quickly recognise and address assaults.
- ❖ Collect and aggregate data: There are numerous punctual security options available to enterprises. Effective incident detection and response are hampered by the disconnected and incomplete network visibility that results.

The role and responsibility of the security analyst are presented in the marketing documents (datasheet for example) as features. Following, we will capture these features related to the virtual security analyst and compare them with the role and responsibility of a real security analyst as presented previously in this sub-chapter.

| Security analyst role and responsibility | Virtual security analyst (VSA) features | Comments | Performed with AI |
|---|---|---|---|
| **Investigate possible incidents** | 1. "Helps you identify, classify, and respond to threats including those well camouflaged." (Fortinet, 2022a, p. 1)<br>2. "Provide sub-second investigation by harnessing deep learning technologies that assist you in an automated response to remediate different breeds of attacks." (Fortinet, 2022a, p. 1) | The first is a global feature and incident investigation is part of it. By going through the response process, investigation must take place. The VSA can do investigation as we will see in the lines below.<br>The second feature include investigation with ML | In part. |

| | | | |
|---|---|---|---|
| | "Detect network anomalies where traditional security solutions fail" (under Key features), (Fortinet, 2022a, p. 1). | The VSA detects network anomalies, as you can see in the results presented above, but how do we compare with traditional solutions? To what extent does it detect anomalies where traditional solutions cannot? We did not evaluate that. | VSA uses ML to detect anomalies as you can see in figures 16 and 18. |
| | "Detect encrypted attack, malicious web campaigns, weaker ciphers, vulnerable protocols, IP and DNS-based botnet attacks with advanced analytics" (under Network Detection and Response responsibilities), (Fortinet, 2022a, p. 2). | Supported by FortiNDR, even though it is under the security analyst responsibilities, the datasheet indicates it as a feature of the network detection and response, and not a VSA's feature. | From our evaluation, no ML is used for this feature. |
| | 1. "Profile network traffic with ML models to identify anomalies with user feedback mechanism" (under Network Detection and Response responsibilities), (Fortinet, 2022a, p. 2).<br>2. "ML-based Traffic Profiling and Malware Detection" (Fortinet, 2022a, p. 1) | Supported by FortiNDR, even though it is under the security analyst responsibilities, the datasheet indicates it as a feature of the network detection and response, and not a VSA's feature. | ML used as you can see in our results. |
| | "Reduces the time to identify network anomalies and malicious content on your network." (Fortinet, 2022a, p. 1) | Based on the different database and the ML, anomalies are detected quickly compared to a SOC analyst which must take more time by using tools like Wireshark to capture and analyze network packets. | Contains different database and ML engines that are used to speed up the detection process. |
| | "Mimic experienced security analyst for outbreak, anomalies, and malware detection, processing large volumes of network data" (under Key features), (Fortinet, 2022a, p. 1). | The environment where we collected the network traffic for analysis is a corporate network with large volume of data, as you saw in the dashboard in figures 9, more than 50 million of network sessions were analyzed and more than 8 terabytes of data process by the FortiNDR, this is something that a human SOC cannot do in the same amount of time. | ML was used partially to process the data. |
| | "Analyze zero days scientifically including fileless threats and classify them into 20+ malware attack scenarios" (under Key features), (Fortinet, 2022a, p. 1). | We did not evaluate a zero day. | Not evaluated |
| | "Integrate Fortinet Security Fabric and third party (via API) with FortiGate inline blocking, FortiSwitch/FortiNAC quarantine, FortiAnalyzer, and FortiSOAR" (under Network Detection and Response responsibilities), (Fortinet, 2022a, p. 2). | Our first scenario of test was by using FortiGate. We did not evaluate other forms of integration. | AI engine used for file scan. |
| | 1. "Investigate the attack source by tracking the original source of infection with time stamps" (under virtual security analyst responsibilities), (Fortinet, 2022a, p. 2). | We can see that there is always a timestamp with each session analyzed or anomaly detected. But we did not evaluate if it is using ML or not. | Not evaluated |

| | | | |
|---|---|---|---|
| | 2. "Emulate a FortiGuard malware analyst and scientifically determine the type of malware based on an evolving neural network that constantly learns and matures over time and experience" (under virtual security analyst responsibilities), (Fortinet, 2022a, p. 2). | There are frequent update of the various databases and engines from the FortiGuard which use ML | ML is used by FortiGuard |
| | "Search for outbreaks on networks and look for traces of malware based on hashes and similar variants" (under virtual security analyst responsibilities), (Fortinet, 2022a, p. 2). | This feature was not evaluated. | Not evaluated |
| **Triage and prioritize identified incidents** | This is not supported by the VSA, so FortiNDR. | Not supported by FortiNDR | Not applicable |
| **Coordinate incident response** | 1. Helps you identify, classify, and respond to threats including those well camouflaged." (Fortinet, 2022a, p. 1)<br>2. "AI-Powered Detection and Response for Cyber Attacks" (Fortinet, 2022a, p. 1)<br>3. "Automate and manually respond for quarantine and control" (under Key features), (Fortinet, 2022a, p. 1)<br>4. "Provide sub-second investigation by harnessing deep learning technologies that assist you in an automated response to remediate different breeds of attacks." (Fortinet, 2022a, p. 1)<br>5. "Integrate Fortinet Security Fabric and third party (via API) with FortiGate inline blocking, FortiSwitch/FortiNAC quarantine, FortiAnalyzer, and FortiSOAR" (under Network Detection and Response responsibilities), (Fortinet, 2022a, p. 2).<br>6. "Integrate into Fortinet's Security Fabric by uniting with FortiGate and others to automatically quarantine attacks" (under Key features), (Fortinet, 2022a, p. 1). | The first feature groups other features including response to incident.<br>The response feature of FortiNDR was not evaluated, but based on the administration guide, integration with FortiGate is possible and it can be done automatically or manually. | Not evaluated |
| | "Identify and classify attack scenarios that determine malware attacks with chain-on-infection and big picture analysis" (under virtual security analyst responsibilities), (Fortinet, 2022a, p. 2). | This is especially important for the response process as it provides relevant input for incident response. | Not evaluated |
| **Maintain pertinence** | 1. "Detect encrypted attack, malicious web campaigns, weaker ciphers, vulnerable protocols, IP and DNS-based botnet attacks with advanced analytics" (under Network Detection and Response responsibilities), (Fortinet, 2022a, p. 2).<br>2. "Breach Prevention" (Fortinet, 2022a, p. 1) | Detecting vulnerable protocols is especially useful to take preventive actions. For example, during our evaluation, vulnerable TLS and SMB protocols were detected. By applying corrective measures to the detected vulnerabilities, it can prevent attacks. | ML not used. |
| | "Analyze zero days scientifically including fileless threats and classify them into 20+ malware attack scenarios" (under Key features), (Fortinet, 2022a, p. 1). | This feature was not evaluated. | Not evaluated |

| | | | |
|---|---|---|---|
| **Patch exposed or sensitive systems** | This is not supported by the VSA, so FortiNDR. | Not supported by FortiNDR | Not applicable |
| **Manage the infrastructure** | This is not supported by the VSA, so FortiNDR. | Not supported by FortiNDR | Not applicable |
| **Address support tickets** | This is not supported by the VSA, so FortiNDR. | Not supported by FortiNDR | Not applicable |
| **Report to management** | This is not supported by the VSA, so FortiNDR. | Not supported by FortiNDR | Not applicable |
| **SOC analysts' Challenges** | | | |
| **Staff important roles** | 1. "Designed for short-staffed Security Operation Center (SOC) teams" (Fortinet, 2022a, p. 1)<br>2. "Shortage of Experienced SOC Analysts" (Fortinet, 2022a, p. 1)<br>3. "Mimic experienced security analyst for outbreak, anomalies, and malware detection, processing large volume of network data" (under Key features), (Fortinet, 2022a, p. 1). | Security analysts investigate network traffic to determine which one content anomalies, and this is more difficult and time consuming for huge amounts of data. By using the VSA, it automates this low-level part of the SOC analyst activities by identifying quickly and precisely patterns related to anomalies in data traffic. The security analyst can now proceed to the next level which consists of triaging and prioritization of identified incidents. The analyst accomplishes this process based on a good decision making and good understanding of the business environment and priorities. | Support AI |
| **Eliminate false positives** | "Provide on-premises learning to reduce false positives by analyzing organizational-specific traffic and adapting to newly disguised threats" (under Key features), (Fortinet, 2022a, p. 1). | This feature was not evaluated.<br>From the results obtained, there are many anomalies that are business related traffic, and additional configuration fine tuning could be used to reduce false positives. | Not evaluated |
| **Minimize impact on operations** | "Reduces the time to identify network anomalies and malicious content on your network." (Fortinet, 2022a, p. 1) | This feature was not evaluated. But, as a security analyst, if I must investigate network traffic using Wireshark for anomalies, it takes more time than with the VSA. | Not applicable |
| **Respond faster to attacks** | 1. "Provide sub-second investigation by harnessing deep learning technologies that assist you in an automated response to remediate different breeds of attacks." (Fortinet, 2022a, p. 1)<br>2. "Reduces the time to identify network anomalies and malicious content on your network." (Fortinet, 2022a, p. 1)<br>3. "Reduce malware detection and investigation time from minutes to sub-second verdict" (under Key features), (Fortinet, 2022a, p. 1). | This feature was not evaluated.<br>But, by reducing time to identify malwares, the response process to incident is optimized. In addition, by integrating with FortiGate response can be automated and faster. | Not evaluated |

| | | | |
|---|---|---|---|
| | 4. "Integrate Fortinet Security Fabric and third party (via API) with FortiGate inline blocking, FortiSwitch/FortiNAC quarantine, FortiAnalyzer, and FortiSOAR" (under Network Detection and Response responsibilities), (Fortinet, 2022a, p. 2).<br>5. "Integrate into Fortinet's Security Fabric by uniting with FortiGates and others to automatically quarantine attacks" (under Key features), (Fortinet, 2022a, p. 1). | | |
| **Collect and aggregate data** | 1. "ML-based Traffic Profiling and Malware Detection" (Fortinet, 2022a, p. 1)<br>2. "Provide on-premises learning to reduce false positives by analyzing organizational-specific traffic and adapting to newly disguised threats" (under Key features), (Fortinet, 2022a, p. 1)<br>3. "Emulate a FortiGuard malware analyst and scientifically determine the type of malware based on an evolving neural network that constantly learns and matures over time and experience" (under virtual security analyst responsibilities), (Fortinet, 2022a, p. 2). | There are two levels of data collection and aggregation. First at FortiGuard and secondly in the environment where the system is installed as we did in the second mode of data collection for our analysis. For each of them ML is used to identify anomalies and malwares in those data. | ML is supported. |

Table 2. *Security analyst versus FortiNDR virtual security analyst*

From this comparison we see that, amongst the eight roles and responsibilities of a SOC analyst, FortiNDR provides features in the following three groups only: incidents investigation, incident response coordination, and pertinent maintenance. Most of the features provided by the system falls under the incidents investigation group, especially at the low-level of the investigation where security analysts investigate massive amounts of network traffic to look for anomalies or malwares. Followed by the incidents investigation, some features of the incident response coordination group are beneficial for the automation of the response process; but the problem is the limitation of the integration with non-Fortinet systems for incident response. Another relevant aspect relates to the features of the network detection and response compared to the ones of the virtual security analyst. Network detection and response is part of the responsibilities of a SOC analyst, and it brings confusion when differentiating the responsibilities of the virtual security analyst with the network detection and response responsibilities. The affirmation that VSA could replace a human does not indicate to what extent that will be possible, as we saw from the comparison table that only 3 of 8 skills of a human security analyst are covered by the VSA. Even by analyzing these 3 skills deeply, we see that the VSA does not handle them completely in all types of environments. For example, based on the documentation as we did not evaluate it, the response provided by VSA is made easy and automated in an

environment with Fortinet equipment, there is no guarantee to have the same response apply in a non-Fortinet environment. In addition, the VSA cannot operate by itself as it still needs a human to manage it. FortiNDR contributes with some of its features, to meet the challenges encountered by SOC analysts in their day-to-day job in some ways, and this is an important benefit as it reduces the effort they deploy for incident investigation and response. Lastly, we see that some of the features provided by FortiNDR are performed with the AI technology, in the next sub-chapter, we will continue the comparison by focussing on the supported features and their use of AI or not.

### 4.4 Other gaps between AI claims in marketing document and findings

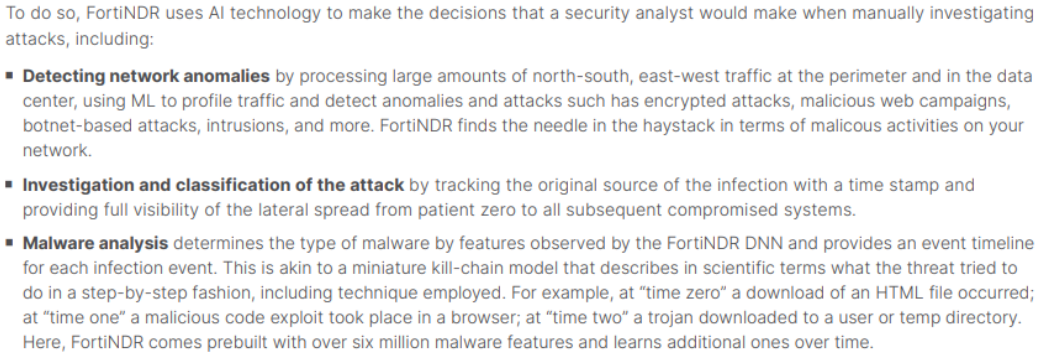
The chapter presents other gaps between the claims related to AI in the datasheet and the finding during our evaluation.

To begin with, let capture some of the claims from the datasheet and the solution brief documents for FortiNDR :


Source : FortiNDR Datasheet document

*Figure 20. AI claim from marketing document group1*

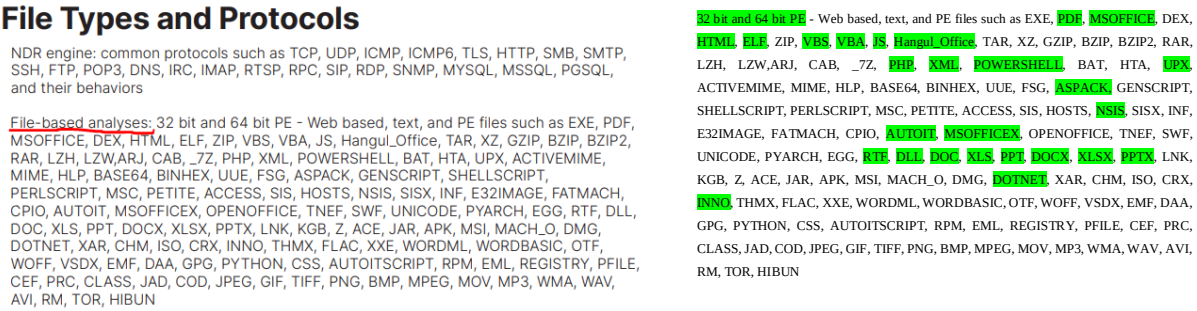
Claim 6




Source : FortiNDR Solution Brief document

*Figure 21. AI claim from marketing document group2*

Claim1 is the introduction to the datasheet, it lets us understand that FortiNDR only works with AI technology, based on a patented technology that we presented in chapter 4.1.1. The other claims relate to traffic profiling, anomalies or malwares detection, and malware analysis, except claim 2 which also deals with response to cyberattack. What each of these claims have in common is their use of AI or ML to provide the expected features.



**File Types and Protocols**

NDR engine: common protocols such as TCP, UDP, ICMP, ICMP6, TLS, HTTP, SMB, SMTP, SSH, FTP, POP3, DNS, IRC, IMAP, RTSP, RPC, SIP, RDP, SNMP, MYSQL, MSSQL, PGSQL, and their behaviors

File-based analyses: 32 bit and 64 bit PE - Web based, text, and PE files such as EXE, PDF, MSOFFICE, DEX, HTML, ELF, ZIP, VBS, VBA, JS, Hangul_Office, TAR, XZ, GZIP, BZIP, BZIP2, RAR, LZH, LZW,ARJ, CAB, _7Z, PHP, XML, POWERSHELL, BAT, HTA, UPX, ACTIVEMIME, MIME, HLP, BASE64, BINHEX, UUE, FSG, ASPACK, GENSCRIPT, SHELLSCRIPT, PERLSCRIPT, MSC, PETITE, ACCESS, SIS, HOSTS, NSIS, SISX, INF, E32IMAGE, FATMACH, CPIO, AUTOIT, MSOFFICEX, OPENOFFICE, TNEF, SWF, UNICODE, PYARCH, EGG, RTF, DLL, DOC, XLS, PPT, DOCX, XLSX, PPTX, LNK, KGB, Z, ACE, JAR, APK, MSI, MACH_O, DMG, DOTNET, XAR, CHM, ISO, CRX, INNO, THMX, FLAC, XXE, WORDML, WORDBASIC, OTF, WOFF, VSDX, EMF, DAA, GPG, PYTHON, CSS, AUTOITSCRIPT, RPM, EML, REGISTRY, PFILE, CEF, PRC, CLASS, JAD, COD, JPEG, GIF, TIFF, PNG, BMP, MPEG, MOV, MP3, WMA, WAV, AVI, RM, TOR, HIBUN

32 bit and 64 bit PE - Web based, text, and PE files such as EXE, PDF, MSOFFICE, DEX, HTML, ELF, ZIP, VBS, VBA, JS, Hangul_Office, TAR, XZ, GZIP, BZIP, BZIP2, RAR, LZH, LZW,ARJ, CAB, _7Z, PHP, XML, POWERSHELL, BAT, HTA, UPX, ACTIVEMIME, MIME, HLP, BASE64, BINHEX, UUE, FSG, ASPACK, GENSCRIPT, SHELLSCRIPT, PERLSCRIPT, MSC, PETITE, ACCESS, SIS, HOSTS, NSIS, SISX, INF, E32IMAGE, FATMACH, CPIO, AUTOIT, MSOFFICEX, OPENOFFICE, TNEF, SWF, UNICODE, PYARCH, EGG, RTF, DLL, DOC, XLS, PPT, DOCX, XLSX, PPTX, LNK, KGB, Z, ACE, JAR, APK, MSI, MACH_O, DMG, DOTNET, XAR, CHM, ISO, CRX, INNO, THMX, FLAC, XXE, WORDML, WORDBASIC, OTF, WOFF, VSDX, EMF, DAA, GPG, PYTHON, CSS, AUTOITSCRIPT, RPM, EML, REGISTRY, PFILE, CEF, PRC, CLASS, JAD, COD, JPEG, GIF, TIFF, PNG, BMP, MPEG, MOV, MP3, WMA, WAV, AVI, RM, TOR, HIBUN

*Figure 22. File type supported by FortiNDR – Captured from the datasheet*

The scan flow in chapter 4.1.2 has two steps, the first step is the scan by the Antivirus Engine (AV static engine) and the second step is the scan by either the text AI engine or the binary AI engine based on the following file types : text file types contain PDF, MSOFFICE, HTML, VBS, VBA, JS, PHP, HWP, Hangul_Office, XML, Powershell, MSOFFICEX, RFT, DOC, XLS, PPT, SOCX, SLSX, PPTX, IFRAME; and binary file types are : 32-bit and 64-bit PE, ELF, UPX, APACK, NSIS, AUTOIT, DLL, DOTNET, INNO. Combining these two lists of supported files in the second step of scan, it looks like some types of files are not supported by both AI engines while on the other hand the antivirus engine scan (and support) all types of files. For example, image (JPEG, GIF, TIFF, PNG, BMP, MPEG) and audio/video (MOV, MP3, WMA, WAV, AVI) file types are not scanned by any AI engines, but they are scanned by the antivirus engine. To continue further, the administration guide presents the following list of files as only supported by ANN :32-bit and 64-bit PE, PDF, MSOFFICE, HTML, ELF, VBS, VBA, JS, PHP, HWP, Hangul_Office, XML, POWERSHELL, UPX, ASPACK, NSIS, AUTOIT, MSOFFICEX, RTF, DLL, DOC, XLS, PPT, DOCX, XLSX, PPTX, DOTNET, INNO, IFRAME (Fortinet, 2023a, p. 17). Figure 22 presents the file type from the datasheet on the left, and on the right amongst those file type the ones highlighted in green are supported

by the ANN (the ML technology used by FortiNDR), this gives about 25 percent of file types declared in the datasheet sent to the AI for processing. Unfortunately, the marketing document namely the datasheet, presents the list of file types supported by the AI system and makes no mention that only a quarter of those file types goes through the AI for analysis.

Another difference is the real usage of AI to identify malwares. Claim 6 in figure 21 says that "using ML to profile traffic and detect anomalies and attacks such as encrypted attacks, malicious web campaigns, botnet-based attacks, intrusions, and more". In chapter 5.2.1.1 where we analysed the types of anomalies, we see that amongst the types of malwares collected by FortiNDR, only ML discovery anomalies are detected by the AI, the other malwares such as botnet, networks attacks, weak or vulnerable communication, and encrypted attacks are not processed by the AI. For a period of one month, FortiNDR detected 175 botnets, 48.31 million counts of network attacks corresponding to 10000 network attacks, about 92 million of weak/vulnerable protocols corresponding to 30000 weak/vulnerable communications, 1183 encrypted attacks, and 38.38 million of counts of anomalies found by the ML discovery. Therefore, a relevant part of the malware profiling and detection is performed by non-AI components in the FortiNDR.

### 4.5 Limitations

During the evaluation of the AI system, we encountered some limitations that we are going to present in this chapter, followed by the consequences on our work and results. The first restriction is the availability of technical documents from Fortinet about the FortiNDR. Our research is based on evaluating the claims compared to reality, but we also needed technical documents to understand how the solution works and how the inner components in the solution work together to provide the announced features. Unfortunately, these details were not available in the public documents we have access to. The consequences to our work and results may be a wrong understanding and deduction, biases in analyzing some results. For example, figure 3 shows different engines and databases included in FortiNDR, neither we found nor obtained their description from the manufacturer, finally we provided some meaning based on our research and understanding.

The second limitation is the scope of our research. We did not evaluate all features related to AI and we focused on malware profiling and detection. Features under the virtual security analyst menu such as express malware analysis, outbreak search, static filter, and features related to response and remediation were not evaluated. Moreover, we did not finetune the configurations or apply any mitigation methods to reduce false positives or anomalies from the collected traffic. By doing it, it would have removed all legitimate business traffic detected as anomalies and consequently reduced the list of anomalies or malwares detected. The goal was to see to what extent anomalies can be detected in a common network environment with the system evaluated. Other studies could be based on the response and mitigation of AI in cybersecurity.

# Chapter 5. Conclusion

Our study shows as stated by Lewis et al. (2021) that we must make a distinction between AI as a "speech act" and AI as a technology with realistic implementation (p. 6). Based on the features evaluated, our investigation reveals many inconsistencies between the features stated in the marketing claims and the reality from implementation in a real-live environment.

The first inconsistency is that the virtual security analyst with the built-in AI may replace an experienced human security analyst. The benefit of using the VSA is its capability to process huge amounts of data with AI to identify and detect anomalies and to participate in the response process. In that scope, we can say that it is helpful for understaffed SOC team at level 1 support, as AI is used as a tool to reduce manual effort work of a human security analyst and automate this work at machine speeds so that the security analyst could progress in their job to the next level of support in the SOC team. In contrast the VSA cannot replace a human security analyst because amongst the eight responsibilities presented, it covers only three of them: investigation of possible incidents, coordination of incident response, and maintenance of pertinence. The other five: triage and prioritization of identified incidents, patching of exposed or sensitive systems, infrastructure management, support tickets handling, and report to management are not supported by the VSA.

The second inconsistency is the hyperbole regarding the features provided by AI in the system. By going through the marketing documents, we have the impression that the system only uses AI to provide all the marketed features, but the reality is completely different. The system contains two groups of scan engines, the first group is the antivirus engine which performs advanced analysis using non-AI technology, we suppose it uses traditional antivirus technology based on signature-based because no detail is given about in the documentation we accessed. This antivirus engine scans all files entering the system. The second group contains two AI engines, one for text file type and the other for binary file type. These two engines use AI and support only 25% of file types supported by the entire system, which is quite a low percentage of file types supported and no indication is done in the datasheet about this. In the list of file types in the datasheet, there is no

differentiation between the ones processed by the antivirus and those by AI. Additionally, there is no occurrence of the word “antivirus” either in the datasheet or other marketing documents that we accessed. We found only one occurrence of the term “advanced analytics,” when it talks about “detect encrypted attack, malicious web campaigns, weaker ciphers, vulnerable protocols, IP and DNS-based botnet attacks with advanced analytics” (Fortinet, 2022a, p. 2). Our results indicate that the system uses this advanced analytic, which is not based on AI, and uses a signature-based approach to detect known malware such as botnets, network attacks, weak or vulnerable communications, and encrypted attack, based on their corresponding databases in the system.

Finally, based on these findings, our recommendations are towards the following categories of users of AI systems or products: the creators, the consumers, and the researchers. To the creators, we know competition is part of the industry, everyone wants to be the first to create an AI-based system with outstanding and revolutionary features and they tend to market new features as built with AI technology. As a result, the reality regarding the evolution and the actual capabilities of AI is biased and overestimated. In addition, they don't aggregate their effort and knowledge to make the AI field grow quickly and consequently like the AI features they claimed and marketed. Creators must put their effort together to build a framework that will make the AI technology grow faster and provide systems with real AI-based features. To the consumers, we will recommend not to trust what is sold and marketed about the revolutionary features built with AI, the most important are the features they have in the products or systems either made with AI or not. Lastly, the researchers must continue their work on AI which helps first in identifying the pitfalls that the industry and market say about AI and its evolution, and secondly by developing new AI standards and frameworks that will help to the growth of the AI technology and its applications.

## REFERENCES


[1] Azimi, S., Pahl, C. (2021). The Effect of IoT Data Completeness and Correctness on Explainable Machine Learning Models. In: Strauss C., Kotsis G., Tjoa A.M., Khalil I. (Eds), *Database and Expert Systems Applications* (DEXA 2021. Lecture Notes in Computer Science, vol 12924. Springer). https://doi.org/10.1007/978-3-030-86475-0_15

[2] Bass B. & Bloomberg (2023, February 16). Buzzy ChatGPT chatbot is so error-prone that its maker just publicly promised to fix the tech's 'glaring and subtle biases.' https://fortune.com/2023/02/16/chatgpt-openai-bias-inaccuracies-bad-behavior-microsoft/

[3] Boden, M. (2018). Artificial intelligence: a very short introduction. Oxford University Press.

[4] Checkpoint (n.d.). Security Operations Center (SOC) Roles and Responsibilities. https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-soc/security-operations-center-soc-roles-and-responsibilities/

[5] Conti, M., Dargahi, T., Dehghantanha, A. (2018). Cyber Threat Intelligence: Challenges and Opportunities. In: Dehghantanha A., Conti M. & Dargahi T. (Eds), *Cyber Threat Intelligence* (Advances in Information Security, vol 70. Springer, pp. 1-6). https://doi.org/10.1007/978-3-319-73951-9_1

[6] EICAR (n.d.). Eicar. https://www.eicar.org/

[7] Fortinet (2022a, August 15). Datasheet FortiNDR. https://www.fortinet.com/content/dam/fortinet/assets/data-sheets/fortindr.pdf

[8] Fortinet (2022b, April 6). Solution brief FortiNDR. https://www.fortinet.com/content/dam/fortinet/assets/solution-guides/sb-fortindr.pdf

[9] Fortinet (2023a, January 18). Administration Guide FortiNDR 7.1.0. https://fortinetweb.s3.amazonaws.com/docs.fortinet.com/v2/attachments/328ad77a-50a4-11ed-9d74-fa163e15d75b/FortiNDR-7.1.0-Administration_Guide.pdf

[10] Fortinet (2023b). FortiNDR Service: https://www.fortiguard.com/services/fortindr

[11] Fortinet (n.d.). FortiAI Next Generation Network Security Artificial Intelligence in Cyber Security.

[12] Grosse, R. (2018). [Lecture 9: Generalization]. Department of Computer Science, University of Toronto. https://www.cs.toronto.edu/~rgrosse/courses/csc321_2018/readings/L09%20Generalization.pdf

[13] IBM (n.d.). Artificial intelligence (AI) for cybersecurity. https://www.ibm.com/security/artificial-intelligence?utm_content=SRCWW&p1=Search&p4=43700074592765696&p5=e&gclid=Cj0KCQiArsefBhCbARIsAP98hXTfOjmENFzUL7D_hZJOC1xk-2bK0lmDoW4sp33sO1OD8lfb-ZIQxQsaAqGxEALw_wcB&gclsrc=aw.ds

[14] ISACA (2021). Technology AI Uses in Blue Team Security. Emerging technology. https://store.isaca.org/s/store#/store/browse/detail/a2S4w000004L3CjEAK

[15] Kaloudi, N. & Li, J. (2020, February). The AI-Based Cyber Threat Landscape: A Survey. *ACM Computing Surveys*, 53(1), 1–34. https://doi.org/10.1145/3372823

[16] Lewis, P. R., Marsh, S., & Pitt, J. (2021). AI vs "AI": Synthetic Minds or Speech Acts. *IEEE Technology and Society Magazine*, 40(2), 6-13. https://doi.org/10.1109/MTS.2021.3077052

[17] Logsign (2019, November 20). What is Malware Outbreak? https://www.logsign.com/blog/what-is-malware-outbreak/

[18] Malle, B. F., Fischer, K., Young, J. E., Moon, A., & Collins, E. (2020). Trust and the discrepancy between expectations and actual capabilities of social robots. In D. Zhang and B. Wei (Eds.), *Human-robot interaction: Control, analysis, and design*. New York, NY: Cambridge Scholars Publishing.

[19] Marcus, G., & Davis, E. (2020). Rebooting AI - Building Artificial Intelligence We Can Trust. Pantheon Books New York

[20] No stockbroker uses AI, most `powered by AI' are marketing claims: Nithin Kamath [Stock in news]. (2021, Nov 26). The Economic Times.

http://search.proquest.com.uproxy.library.dc-uoit.ca/newspapers/no-stockbroker-uses-ai-most-powered-are-marketing/docview/2602113480/se-2

[21] Nguyen, T. N. & Choo, R. (2021). Human-in-the-Loop XAI-enabled Vulnerability Detection, Investigation, and Mitigation. *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 36, 1210-1212. https://doi.org/10.1109/ASE51524.2021.9678840

[22] Pieters, W. (2011). Explanation and trust: what to tell the user in security and AI? *Ethics and Information Technology*, 13(1), 53-64. https://doi.org/10.1007/s10676-010-9253-3

[23] Pillsbury (2021). Artificial intelligence & cybersecurity. https://www.pillsburylaw.com/images/content/1/5/155129/Report-AI-Cybersecurity-Sept-2021.pdf

[24] Redman, T. (2018). If Your Data Is Bad, Your Machine Learning Tools Are Useless. Harvard Business Review. Retrieved from https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless

[25] SANS (n.d.). 20 Coolest Careers in Cybersecurity. https://www.sans.org/cybersecurity-careers/20-coolest-cyber-security-careers/

[26] Stolte, R. (2018, May 15). Short on security analysts? AI can help. CSO (Online), http://search.proquest.com.uproxy.library.dc-uoit.ca/trade-journals/short-on-security-analysts-ai-can-help/docview/2039008809/se-2

[27] Xu Y. (2020, February 6). Malware Identification Using Multiple Artificial Neural Networks. https://uspto.report/patent/app/20200042701#D00001

# APPENDICES

## Appendix A. Approval to install and collect real-live data

For my research, I received the approval from a company to use their infrastructure to install, operate, and evaluate FortiNDR in a real-live environment.

I also received demo licence from Fortinet to evaluate the FortiNDR system for the purpose of my master's degree project research.

## Appendix B. Updates on the FortiNDR datasheet

The first draft of this report was shared with Fortinet teams. While reviewing and proofreading the report before final approval and submission, we noticed that Fortinet did major changes in the content of the datasheet made publicly available on March 2, 2023. The documents we used for our study and evaluation have the following dates as indicated in the references :

- Datasheet FortiNDR : 2022, August 15. https://drive.google.com/file/d/1Sglo5_SU4itT58RgUu1ITldTW1BwYuU2/view?usp=sharing
- Solution brief FortiNDR. 2022, April 6. https://drive.google.com/file/d/1jhILYBzXdG2lduERQFiMsKl4lju5tFRF/view?usp=sharing
- Administration Guide 7.1.0 FortiNDR. 2023, January 18. https://drive.google.com/file/d/14Qxmq7HtTDqYz38nHpXfHJCYLC3DiyTR/view?usp=sharing