Acta Informatica Pragensia
2025, Volume 14, Issue 3, 474–488
https://doi.org/10.18267/j.aip.277

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Article Open Access

DORA: Dionaea Observation and Data Collection Analysis for Real-Time Cyberattack Surveillance and Threat Intelligence

Hartinah Hartinah, Andi Syarwani, Ardiansyah Ardiansyah, Irfan Syamsuddin

Department of Informatics and Computer Engineering, State Polytechnic of Ujung Pandang, Makassar, Indonesia

Corresponding authors: Hartinah Hartinah (hartinah@poliupg.ac.id) and Irfan Syamsuddin (irfans@poliupg.ac.id)

Editorial Record

First submission received:
March 6, 2025

Revisions received:
April 8, 2025
April 30, 2025
June 3, 2025

Accepted for publication:
June 19, 2025

Academic Editor:
Kang Leng Chiew
Universiti Malaysia Sarawak, Malaysia

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Hartinah, H., Syarwani, A., Ardiansyah, A., & Syamsuddin, I. (2025). DORA: Dionaea Observation and Data Collection Analysis for Real-Time Cyberattack Surveillance and Threat Intelligence. *Acta Informatica Pragensia*, 14(3), 474–488. https://doi.org/10.18267/j.aip.277

Abstract

Background: As assaults get more sophisticated, honeypots like Dionaea become an essential tool for analysing attack behaviours and detecting weaknesses. Despite their growing importance in cybersecurity, honeypots' role in real-time cyberattack surveillance and threat intelligence is largely unknown. Many studies concentrate on identifying attacks rather than delivering actionable intelligence for defensive solutions. Furthermore, previous research frequently lacks thorough methodology for comparing attack data to real-world incidents and does not investigate the integration of honeypots with external intelligence services.

Objective: This study assesses the Dionaea honeypot's ability to detect and analyse cyberattack trends, with an emphasis on attack patterns, malware dispersion, and geographical threat sources. The project will look into how Dionaea honeypots, when combined with external analysis services such as VirusTotal, might provide more thorough insights into cyberattack tactics and improve proactive cybersecurity defence mechanisms.

Methods: The Dionaea honeypot was used to identify a range of attacks on vulnerable services including Telnet (Port 23), SMB (Port 445), and MySQL (Port 3306). Over a seven-day observation period, 32,395 attack connections from 6,276 distinct IP addresses were detected, yielding 2,892 malware samples. These samples were examined using VirusTotal, and the findings were categorised by malware type, attack vector, and geographical origin. Geospatial and service-specific attack patterns were also investigated to detect emerging trends and high-risk sites.

Results: The investigation identified WannaCry ransomware as the most common malware, accounting for 1,076 incidents, demonstrating the continuous exploitation of the MS17-010 vulnerability in SMB (Port 445). The most frequently attacked ports were Port 23 (Telnet), Port 445 (SMB), and Port 3306 (MySQL), which received 7,988, 6,898, and 3,589 attack attempts, respectively. Geographically, the leading sources of assault activity were China (42%), the United States (17%), and Japan (13%). The findings demonstrate that honeypots are not only effective attack detection tools, but also significant sources of intelligence for understanding cyber threat methods and adversary behaviours.

Conclusion: This study proposes DORA (Dionaea Observation and Data Collection Analysis), an integrated system that enhances the existing Dionaea honeypot by combining its data with external analysis services like VirusTotal. This integration provides critical insights into real-time cyberattack detection, malware analysis, and attack vector identification. The findings highlight vulnerabilities in services like Telnet and SMB, particularly the exploitation of MS17-010. DORA improves threat intelligence workflows, enhancing malware detection accuracy and classifying threats more efficiently. Additionally, it helps identify high-risk attack surfaces, forming the basis for adaptive cybersecurity strategies. This research contributes to developing resilient defence systems capable of addressing emerging threats.

Index Terms

Honeypot; Cybersecurity; Malware detection and analysis; Cyber threat detection; Network security; Real-time threat intelligence; Vulnerability assessment.

1 INTRODUCTION

Cybersecurity risks have increased dramatically in recent years, correlating with the increasing use of internet-based technologies and the rapid expansion of digital infrastructure. According to Indonesia's National Cyber and Crypto Agency's 2024 Cybersecurity Landscape Report, the government recorded approximately 330 million abnormal traffic incidents in 2024. The Mirai Botnet, which is responsible for approximately 81 million malicious operations, is among the most dangerous threats, followed by Trojan RAT-based attacks, phishing, and exploitation of IoT systems (BSSN, 2025). Furthermore, the analysis shows that the government sector is still the leading target of cyberattacks such as ransomware, data breaches, and Distributed Denial-of-Service (DDoS) operations. The intrusion of The Temporary National Data Centre in Surabaya demonstrates how cybersecurity failures can have substantial ramifications for public services (BSSN, 2025).

The increasing sophistication of cyber threats needs a paradigm change away from traditional reactive defence methods like firewalls and signature-based intrusion detection systems and towards more proactive and adaptable strategies. These conventional methods frequently fail to detect emerging threats and sophisticated attack routes. Honeypots are one of the most promising solutions in proactive threat detection, serving not only as decoy systems to lure and deceive attackers but also as strong instruments for gathering important knowledge about attack tactics, exploitation techniques, and adversary behaviour. Honeypots have emerged as valuable assets in cybersecurity research, allowing for real-time analysis of hostile tactics, strategies, and procedures. While prior research has mostly focused on the static use of honeypots to collect attack data, the integration of honeypots with artificial threat intelligence systems has opened the door to more dynamic and actionable cybersecurity techniques. For example, XT-Pot (Ryandy et al., 2020) developed a framework for connecting honeypots with threat intelligence systems, allowing organisations to transform raw attack data into actionable security measures, considerably improving proactive defence capabilities. Simultaneously, Linux-based honeypots such as Cowrie have been shown to effectively track attacker behaviour, with a particular emphasis on SSH and Telnet vulnerabilities, which continue to be popular targets (Maharani et al., 2024). Furthermore, the combination of high- and low-interaction honeypots resulted in the creation of T-Pot, an all-in-one honeypot solution that integrates different honeypot types to provide more detailed insights regarding cyberattack trends (Martínez et al., 2023). T-Pot's adaptability enables it to adapt to a wide range of attack techniques, hence boosting system security. Furthermore, research has shown that honeypots are increasingly being used in cryptojacking detection, exposing how attackers exploit honeypots to repurpose them for illicit cryptocurrency mining, such as the attempted operation of XMR mining installations (Patel et al., 2022). This emphasises the rising complexity of threats and the need to adapt honeypot systems to identify new types of exploitation.

Honeypots have been integrated with Intrusion Detection Systems (IDS) and machine learning (ML) algorithms to address the growing demand for advanced detection and analysis systems as cyber threats evolve. These technologies are increasingly employed to analyse botnet attacks in Internet of Things (IoT) environments, enabling more effective detection of both automated and human-driven attacks. The next step in honeypot-based defence systems is the integration of behavioural analytics with geolocation-based threat mapping, which facilitates real-time threat monitoring and sophisticated assault pattern recognition.

In this context, a honeypot is intentionally designed as a “decoy” or false target, exposed to attract external attacks. This approach allows researchers to observe attack patterns and analyse the characteristics of malware sent by attackers, all while ensuring the core systems remain secure. For this study, the Dionaea honeypot was selected due to its proven effectiveness in detecting and recording a wide variety of attacks (Morić et al., 2025; Saikawa & Klyuev, 2019; Tabari & Ou, 2020a) along with its ease of integration with external analysis platforms (Holbel et al., 2024; Ryandy et al., 2020; Tabari & Ou, 2020a). This integration improves real-time behavioural trend analysis, significantly enhancing threat detection and surveillance. This study introduces Dionaea Observation and Data Collection Analysis (DORA) for real-time cyberattack surveillance and threat intelligence, an innovative approach to honeypot deployment that integrates Dionaea with external platforms like VirusTotal. The primary objectives of this research are:

1. Assess the capabilities of Dionaea honeypots to identify real-time cyber threats and analyse malware samples.

2. Improve threat intelligence workflows by combining honeypot data with external systems for proactive threat mitigation.
3. Identify high-risk attack surfaces and provide dynamic, data-driven defence methods based on empirical evidence.

Using DORA's sophisticated data-driven approach, this study seeks to build a novel paradigm for boosting cybersecurity resilience, particularly in the context of rising threats in networks, cloud infrastructures, and distributed systems.

The paper is structured as follows. The second section consists of related literature review. The next section presents our methodology to conduct the study. Then, section five provides in depths analysis of results followed by discussion in the next section. Finally, conclusion is drawn in the last section.

2 LITERATURE REVIEW

In recent years, various studies have used honeypots to detect and mitigate cyber risks. These research have produced important insights into attack patterns, adversary behaviours, and the changing nature of assaults, paving the path for honeypot inclusion into proactive defence systems. One of the earliest contributions to this sector was a study by Bartwal et al., (2022) who created a Security Orchestration, Automation, and Response (SOAR) engine that dynamically deploys honeypots based on attacker behaviours within an internal network. This engine was able to manage several VLANs, identify botnets and DDoS attacks, and store malware data. The trial findings showed that the SOAR engine could cut attacker engagement time to 3,148 seconds while detecting 7,823 attacks and intercepting DDoS attack packets. The design outperformed prior solutions and has tremendous potential for enterprises looking to secure their internal networks.

Machine learning advancements have also helped to shape honeypot systems throughout time. Huang et al., (2019) proposed an automated honeypot detection approach based on the Random Forest algorithm that classifies honeypot interactions at the application, network, and system layers. The suggested model surpasses existing machine learning techniques, as demonstrated by an Area Under the Curve (AUC) score of 0.93. This model solves several of honeypot systems' present constraints, particularly in terms of simulated service integrity, and serves as a standard for future honeypot technology advances.

The usage of honeypots in critical infrastructures has also drawn attention. Zia et al., (2019) investigated the use of honeypots in Industrial Control Systems (ICS), demonstrating its effectiveness in detecting cyber-attacks targeting industrial environments. The information gleaned from these deployments could be essential in building adaptive defence mechanisms specialised for ICS security. Furthermore, comparative analysis of data from various honeypot systems has resulted in the creation of more efficient data visualisation tools. Another study compared Grafana Loki to the ELK Stack, demonstrating that while Grafana Loki had higher resource efficiency, the ELK Stack had more user-friendly data visualisation capabilities via automated field mapping (Njoera et al., 2024). The comparison emphasises the significance of combining visual tools with honeypot systems to improve real-time monitoring and decision making. The challenges created by increasing cybersecurity vulnerabilities in emerging technologies, such as Software-Defined Networking (SDN), Network Functions Virtualisation (NFV), and cloud/edge computing, have led to the investigation of honeypots as alternative intrusion detection techniques. Likewise, Radoglou-Grammatikis et al., (2024) emphasised the usefulness of honeypots in misleading attackers while also gathering significant threat intelligence in these dynamic settings. The use of wireless honeypots (WH) in ultra-dense network environments was also examined, with Reinforcement Learning (RL) techniques being used to improve these systems' deployment and defence capabilities.

The combination of blockchain technology and honeypots is a novel way to enhance the security and adaptability of such systems. A novel blockchain-based dynamic honeypot system that decentralises attack data storage and adjusts in real time to changing adversarial techniques is proposed in (Shi et al., 2019). Their approach was proven more resistant to anti-honeypot methods because it employs encrypted communication and service transformation, providing a more robust solution than standard static honeypots. Furthermore, research on Dionaea honeypots (Shahrivartehrani & Abidin, 2016) has proved their ability to detect and analyse malware in cloud environments, underlining honeypots' proactive role in reducing cybersecurity risks.

In the context of Internet of Things (IoT) security, multi-phased and adaptive honeypot ecosystems have been created to change in response to attacker behaviour, considerably enhancing the detection of both automated and human-driven attacks. For example, IoTCMal, a hybrid honeypot system that combines low- and high-interaction honeypots, has been demonstrated to improve threat intelligence by providing more thorough coverage of attack methods (B. Wang et al., 2020; M. Wang et al., 2018; W. Zhang et al., 2020; Tabari & Ou, 2020b). Previous research by Naik and Jenkins., (2018) investigated the use of fuzzy logic for identifying spoofing attacks on low-interaction honeypots, which successfully decreased false positives but struggled in more complicated attack scenarios. Zhang et al., (2019) presented pseudo-honeypots that detect and identify spammers on social networks, broadening the scope of honeypots beyond standard network security applications. However, these technologies must be further developed before they can be completely deployed in public network environments. In addition, Thom et al. (2021) used honeypots in numerous global locations to detect geolocation-based attacks, finding that attack types and severity range dramatically between network settings and geographical regions. This study emphasises the necessity of honeypots in discovering global assault trends, as well as the need for adaptive defence measures that can address geographically spread attacks. Siddiqui and Bokhari., (2021) presented a honeypot-based intrusion detection system (IDS) to improve cyber threat detection and classification. The study looked into the capabilities and limitations of honeypots when used in conjunction with IDS systems, revealing prospects for increased detection rates. It discovered that the bulk of assaults were directed at TCP/IP-based protocols, with HTTP and FTP ports being particularly vulnerable. The study also detected proxy scanning, IIS exploits, and Trojan-based Denial of Service (DoS) attacks, indicating the wide range of attack tactics used by cyber adversaries.

As cyberattacks get more sophisticated, there is an increasing move towards AI-powered and machine learning-based honeypots. Liu et al., (2023) created HoneyMustard, a honeypot framework based on GUI emulation, to improve interaction with attackers and make it more difficult for them to distinguish the honeypot from normal computers. Similarly, Reinforcement Learning (RL) techniques like Deep Q-Networks (DQN) and Double DQN (DDQN) have been used to improve honeypot detection of SSH attacks (Kristyanto & Louk, 2024), allowing honeypots to dynamically adapt to changing attack patterns. Furthermore, Commey et al., (2024) investigated game theory approaches in blockchain-based honeypots to improve IoT security, providing a strategic model for honeypot deployment, however its real-world applicability is limited. Syamsuddin and Barukab., (2022) introduced an enhanced k-Nearest Neighbour (kNN) model for identifying botnet assaults in IoT environments, achieving excellent accuracy, precision, recall, and F1 score. However, the model's application is limited to certain datasets, emphasising the importance of broader validation across varied situations. Yang et al., (2023) investigated the use of high-interaction honeypots as a proactive defence mechanism to protect network security, using a modular design that allows for more flexible and extensive data collection. This system has numerous advantages over typical honeypots, particularly in terms of customisation and scalability.

In conclusion, while earlier research has provided useful insights into the usage of honeypots in cybersecurity, there is still a significant gap in integrating these systems into public network environments for real-time monitoring and dynamic defence. The findings of this study contribute to this gap by investigating the use of honeypots in public networks and introducing Dionaea Honeypots, which are integrated with real-time data analysis and malware detection services, providing a novel approach to cyberattack surveillance and proactive defence strategies.

3 PROPOSED METHODOLOGY

We implemented a honeypot system using various hardware and software specifications, as detailed in Table 1. The experimental setup is described in Section 3.1.

Table 1. *Hardware and Software requirements.*

Items	Details
OS	Ubuntu 18.04
Honeypot Version	Dionaea
Computer Memory	RAM 8 GB
Processor version	Core i5

This section describes how we deployed and analysed the DORA for real-time cyberattack surveillance and threat intelligence, which incorporate Dionaea Honeypot and VirusTotal as a novel real-time cyberattack surveillance system. We also go into depth about the integration of external analytic services and the data visualization tools used to process the attack data. Figure 1 depicts a potential architecture diagram. DORA was configured using the public IP address 182.23.83.127 and a firewall with minimal settings to enhance the capture of attack traffic from external sources. The methodological steps are stated as follows.

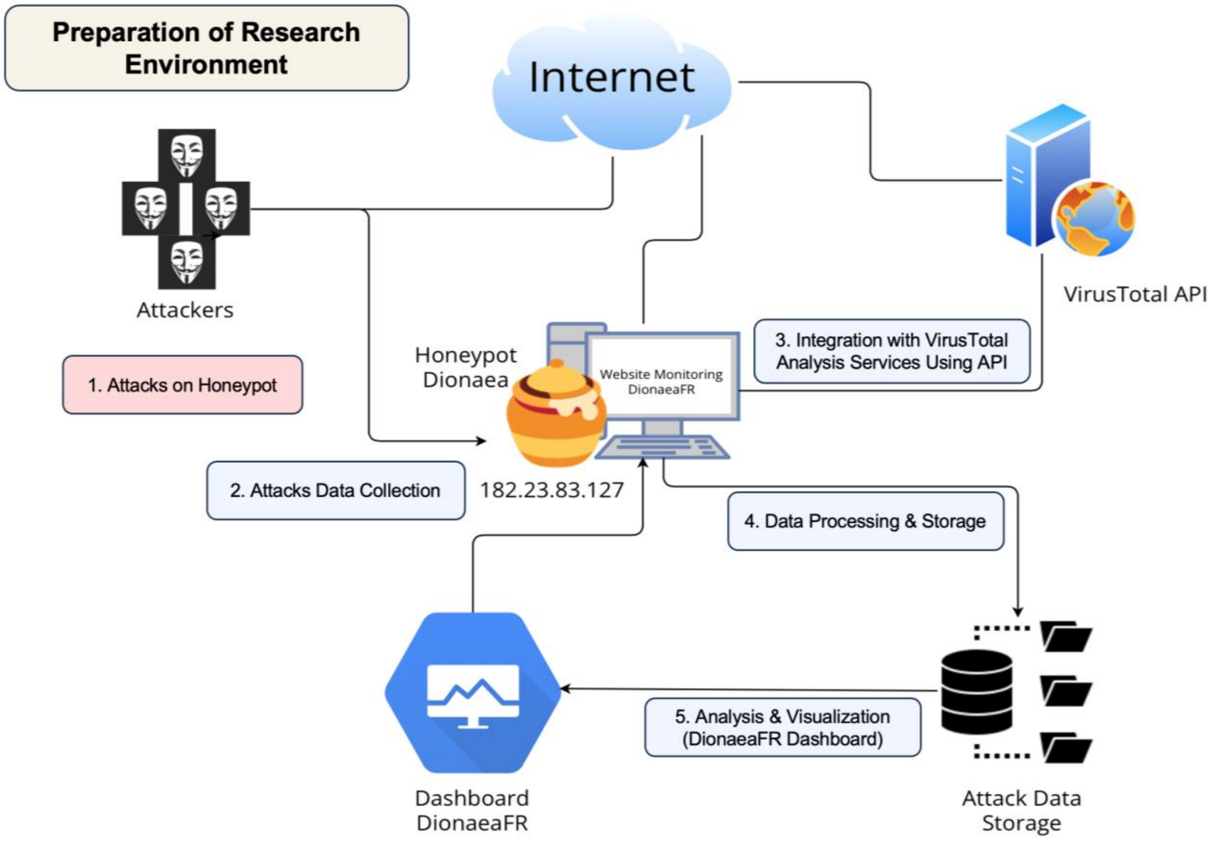

Preparation of Research Environment
Internet
Attackers
VirusTotal API
1. Attacks on Honeypot
Honeypot Dionaea
Website Monitoring DionaeaFR
3. Integration with VirusTotal Analysis Services Using API
2. Attacks Data Collection
182.23.83.127
4. Data Processing & Storage
5. Analysis & Visualization (DionaeaFR Dashboard)
Dashboard DionaeaFR
Attack Data Storage

Figure 1. *System architecture of DORA.*

3.1 Preparation of research environment

Implementation of Dionaea honeypot: The implementation of DORA on critical networks was carried out carefully to ensure that attackers do not realise that this system is a "decoy" within the network while also diverting them away from critical assets. The honeypot in the DORA system was installed on the Ubuntu 18.04 operating system through the "honeynet/nightly" repository, followed by the installation of GeoIP to determine the origin of the attacker's IP address. Additionally, we installed DionaeaFR, a web-based visualization tool that enables the viewing of attack visualizations, including the attacker's IP address, the targeted services, and the malware being delivered. The use of DionaeaFR facilitates efficient monitoring of attack activities against the honeypot.

Security configuration: To keep the DORA system isolated from important networks, it was placed in front of the network firewall. This setting ensures that any attacks aimed at certain ports or services are first directed to DORA, allowing us to study attack trends and identify the most often targeted ports and services, as well as the most typically distributed malware. By placing DORA in front of the firewall, the important network's integrity is preserved, as the firewall rules remain operational without interfering with DORA's data collection activities.

3.2 Attack data collection

Setting digital traps: These setups are required for the Dionaea honeypot to successfully attract attackers by impersonating legitimate systems, files, or data, diverting their attention away from crucial assets. The honeypot configuration leaves all ports open, including port 80 for HTTP, port 23 for SSH, port 445 for SMB, and port 3306 for MySQL, among others. The honeypot is programmed to automatically log every connection attempt and incoming attack via the open ports, providing useful information for analysis.

Log recording: All interactions with Dionaea honeypot are the result of open port configurations, and each attack is scrupulously documented in logs. The logs contain crucial information such as the originating IP address, protocol, timestamp, and any malicious payloads or downloaded files. These logs are an invaluable resource for further analysing attack patterns and finding the most commonly exploited attack pathways.

3.3 Integration with analysis services using API VirusTotal

Malware sample upload: During the installation, we made an account on the VirusTotal platform to receive an API (Application Programming Interface) key, which was then added to the Dionaea honeypot configuration in the "dionaea/ihandlers-available/virustotal.yaml" file. Using this API, the Dionaea honeypot may automatically connect to VirusTotal, allowing each malware sample encountered by the honeypot to be sent for examination in real time. This integration of Dionaea and VirusTotal serves as the foundation for DORA, a comprehensive real-time system for monitoring cyberattacks and generating threat intelligence.

Signature matching: When Dionaea honeypot detects dangerous payloads, such as malware, it automatically forwards the samples to VirusTotal for examination. VirusTotal scans the malware with over 60 antivirus engines and provides a full report for each file. This report contains crucial information such as the type of malware, its threat level, unique file hashes (e.g., MD5, SHA-1, SHA-256), metadata, and the relationships between the malware file and other connected files. This external study provides vital insights into the threat landscape, allowing for the identification of specific malware variants that may be affecting the system.

3.4 Data processing and storage

Storage of attacks and payloads: The Dionaea honeypot has fully automated data collection and storing mechanisms. Each incoming connection attempt, including the originating IP address, protocol, timestamp, and payload or file transferred during the attack, is automatically captured and saved in the database as log data.

Metadata management: The Dionaea honeypot system automatically assigns metadata to each attack, such as the attacker's geographic location, source IP address (IP_SRC), destination port (PORT_DST), and other important information. This data is updated in real time, with new information captured every second when a new threat is discovered. The DORA system also tracks repeated attacks, classifying them as the most common sorts of attacks that enter the system. This functionality is critical for tracking and assessing the changing threat landscape.

3.5 Analysis and visualization (DionaeaFR dashboard)

Data processing in dashboard: DionaeaFR, previously installed, is a web application built using the Django framework that visualizes the collected attack data, providing real-time updates on attack activities. The dashboard presents key metrics such as the total number of attacks, geographical distribution, attack sources, and the most frequently targeted services and ports. This visual representation helps researchers identify attack trends, vulnerable services, and emerging threats in real time. Additionally, the geolocation feature of the dashboard allows us to pinpoint the most active locations involved in attacks, thereby highlighting high-risk areas for potential cyber threats.

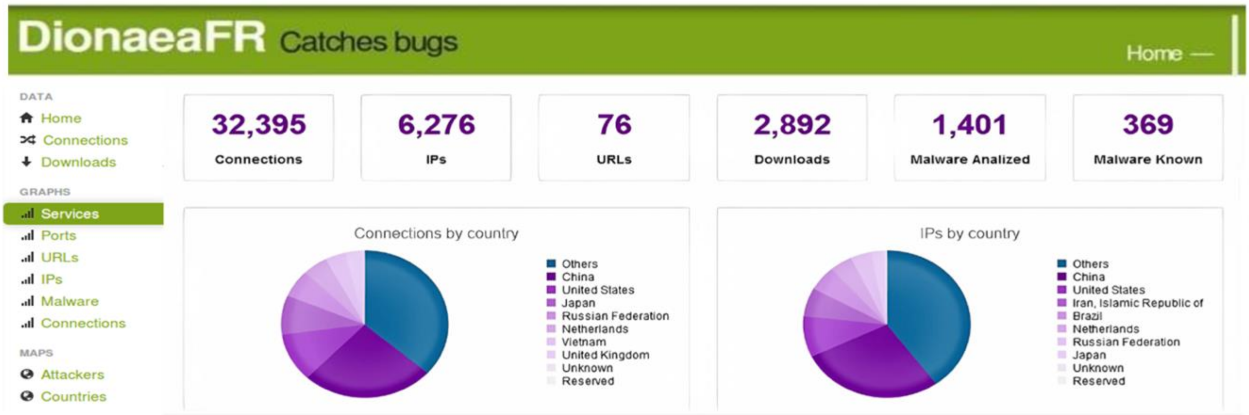
Real-time monitoring: This visualization empowers researchers and practitioners to compare various types of attacks, facilitating the identification of significant attack vectors and paths. By integrating real-time analysis and visualization, the system enables the research and practitioners team to respond quickly and accurately to rapidly evolving threat scenarios, driven by informed insights.

4 RESULTS

The suggested DORA architecture is used to assess and record various sorts of cyberattacks, providing useful information about attack patterns and origins. This approach takes advantage of Dionaea's ability to impersonate vulnerable services such as FTP, HTTP, SMB, and MySQL, allowing it to draw a variety of attacks while also automatically uploading and analysing to VirusTotal. The data gathered throughout the 24/7 observation period is processed and examined to determine the effectiveness of the suggested model. The following are the findings from the Dionaea honeypot:

4.1 Visualization of attacks on the DORA

Figure 2 illustrates that the DionaeaFR visualization web successfully recorded 32,395 connection attempts, 6,276 unique attacker IP addresses, and 76 URLs used by attackers to download malware, access vulnerable services, or issue commands to the system. Furthermore, it collected a total of 2,892 malware samples. Of these, 1,401 samples were analysed, and 369 samples were confirmed as malware through the scanning and analysis process conducted by VirusTotal.



32,395	6,276	76	2,892	1,401	369
Connections	IPs	URLs	Downloads	Malware Analized	Malware Known

***Figure 2**. Honeypot DionaeaFR visualization.*

In the DionaeaFR visualization, two key metrics are displayed: "Connections by Country" and "IPs by Country." While these metrics may seem similar, they serve distinct purposes. "Connections by Country" records the total number of connections, including those from the same IP address. For example, if an attacker from China makes 20 connection attempts to the honeypot in one day using a single IP address, "Connections by Country" will record all 20 attempts from China, as each connection is counted separately, even if they originate from the same IP. In contrast, "IPs by Country" counts only the number of unique IP addresses that have connected. In the same scenario, even though the attacker made 20 connection attempts with the same IP address, "IPs by Country" will record only one IP address from China, as it counts each unique IP address only once.

The Dionaea honeypot can determine the country of origin of an attacker by cross-referencing the attacker's IP address with location data from the GeoIP database. Each device connected to the internet has an IP address that can be mapped to a specific physical location, such as a country or city, using GeoIP. This database is populated with data from various sources, including Internet Service Providers (ISPs), IP address authorities, and public data. Based on the detected IP address, GeoIP identifies the attacker's country of origin. For instance, if an IP address is found to be registered to a range allocated to China, the attack will be recorded as originating from China. However, it's important to consider that the use of VPNs, proxies, or TOR networks may reduce the accuracy of geographic location identification.

In the "Connections by Country" and "IPs by Country" visualizations, the attacker's country categories are displayed based on the highest number of connections and the highest number of unique IP addresses identified through GeoIP. Notably, two additional categories are shown at the bottom of the country names: Unknown and Reserved.

These categories represent connections or IP addresses that cannot be identified due to their absence from the GeoIP database.

- Others refers to connections or IP addresses whose country or region of origin cannot be determined. This typically occurs when VPNs, proxies, or unregistered IP addresses are used to obscure the attacker's true location.
- Reserved refers to private IP addresses that have not been used actively or are not connected to a public network. These addresses are typically part of reserved address spaces for private networks and do not have a geographic location associated with them.

4.1.1 Geographical distribution of attacks

According to the geographic analysis presented in Figure 3, China had the most connections, accounting for 42% of the total discovered connections, followed by the United States (US) at 17% and Japan at 13%. Other countries contributing to the Honeypot system included Russia (9%), the Netherlands (8%), Vietnam (6%), and the United Kingdom (UK) (5%). The significant proportion of connections from China (42%) suggests that the country is the primary source of the detected activities, which include system scanning, vulnerability exploitation, and other sorts of cyberattacks. The United States and Japan also displayed significant activity, indicating that traffic originated from these countries' network infrastructures. The Russian Federation, contributing 9% of the connections, also signals a potential threat that needs attention, particularly given the country's reputation for aggressive cyber activities. The Netherlands and Vietnam demonstrate that attack operations are not centred in a single region, but rather spread throughout multiple countries with varying motivations and assault methods. The United Kingdom (UK), with 5%, indicates that assaults are also coming from networks in Europe. This distribution emphasises the worldwide nature of cyber threats and the need for international cooperation in cybersecurity defence.

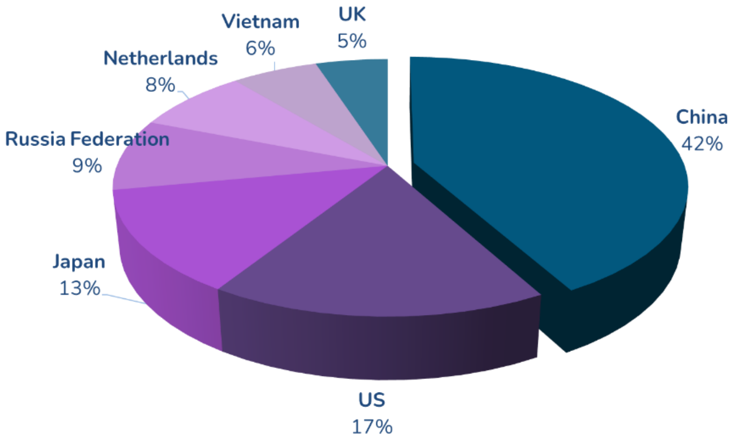


***Figure 3**. Connection by country.*

4.1.2 Malware type analysis

According to the malware detection results in Figure 4, five main types of malwares were found, with WannaCry being the most prevalent threat, accounting for 1,076 incidents. WannaCry, an exploit-based ransomware, spreads via the MS17-010 vulnerability in the SMB service (Port 445), demonstrating that exploitation-based attacks against network protocols continue to pose a substantial danger to cybersecurity. In addition to WannaCry, the "Generic Malware" category comes second with 405 occurrences, indicating the prevalence of common malware variants that are still active and could impact poorly protected systems. Furthermore, Trojan Agent-AYFU was discovered 94 times, indicating the presence of trojans that are most likely utilised for system takeover or data theft.

Two other types of malwares were identified in equal numbers: the MySQL UDF SYS library (PUA) and Conficker-A, each with 87 cases. These findings indicate that both potentially unwanted applications (PUA) and classic worms like Conficker still pose significant threats. The presence of Conficker-A, despite its long-known status, emphasizes that outdated systems remain vulnerable to exploitation. These findings highlight those various types of threats, including ransomware, trojans, worms, and PUAs, are still active and require improved mitigation strategies to reduce risks to systems connected to networks.

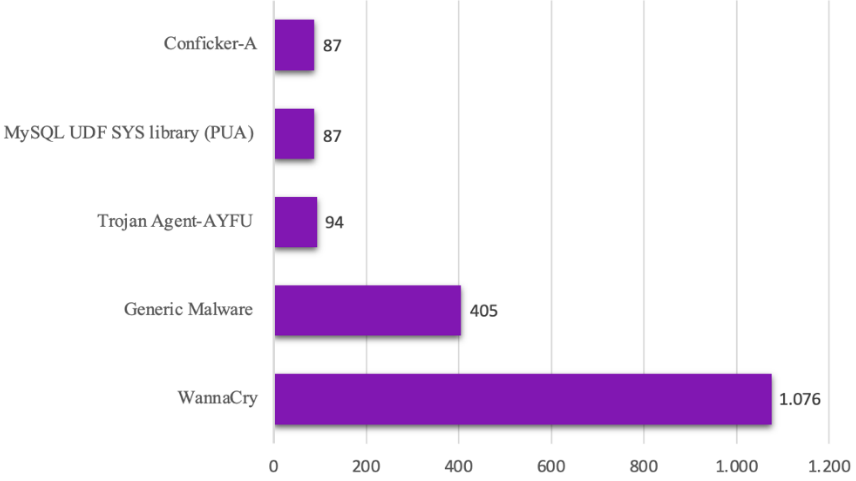
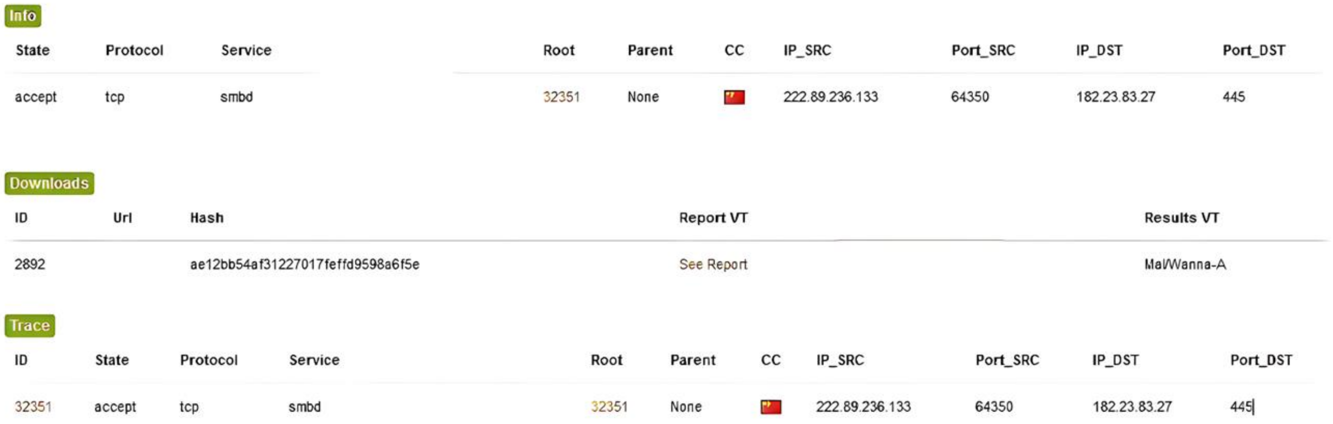


Figure 4. *Top 5 malware detection.*

Additional analysis of one malware sample as seen in Figure 5 revealed that the WannaCry attack started from IP address 222.89.236.133, which was identified as being from China. To deliver and execute the malicious payload, this attack used port 445, the Server Message Block (SMB) service's default port. In this case, the honeypot identified inbound TCP connection activity in which the attacker used source port 64350 to establish communication with the victim at IP address 182.23.83.27, which also utilised port 445 as a destination. The malware analysis system recognised the malicious file that was successfully downloaded as the Mal/Wanna-A variant with the hash ae12b54a1b1227107fefdf95988a8f5. This extensive investigation emphasises the importance of honeypots in detecting exploit-based assaults and giving significant data for identifying malware and its sources.



State	Protocol	Service	Root	Parent	CC	IP_SRC	Port_SRC	IP_DST	Port_DST
accept	tcp	smbd	32351	None	🇨🇳	222.89.236.133	64350	182.23.83.27	445

ID	Url	Hash	Report VT	Results VT
2892		ae12bb54af31227017feffd9598a6f5e	See Report	Mal/Wanna-A

ID	State	Protocol	Service	Root	Parent	CC	IP_SRC	Port_SRC	IP_DST	Port_DST
32351	accept	tcp	smbd	32351	None	🇨🇳	222.89.236.133	64350	182.23.83.27	445

Figure 5. *Malware detection types.*

These findings demonstrate that SMB exploitation remains a significant attack vector in the spread of the WannaCry ransomware, which constantly exploits the MS17-010 vulnerability. This demonstrates that systems that have not been updated or are still vulnerable to this exploit remain good targets for worm-based assaults such as WannaCry.

4.1.3 Number of attacked ports

According to the data in Figure 6, Port 23 (Telnet) was the most targeted port, with 7,988 exploitation attempts, followed by Port 445 (SMB), which received 6,898 attacks. Port 3306 (MySQL) came in third with 3,589 attacks, followed by Port 1433 (Microsoft SQL Server) with 1,623 attacks. Meanwhile, 876 attacks were directed at Port 53281, which is commonly connected with proxy services. These findings show the network's most susceptible ports, which are frequently targeted by attackers exploiting known service flaws.

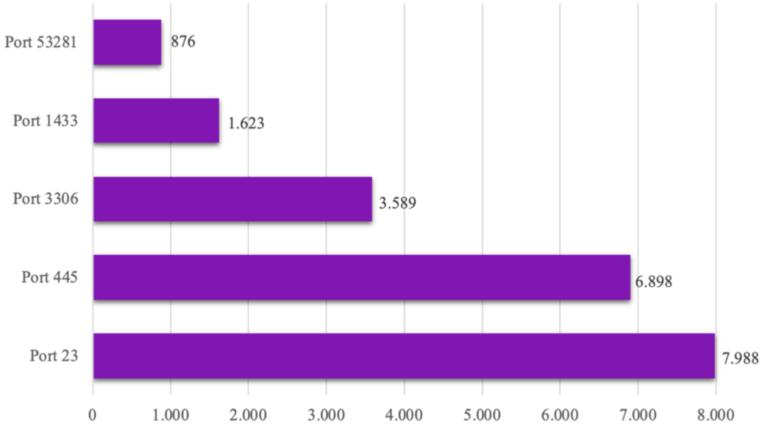


***Figure 6**. Number of attacked ports.*

The high frequency of assaults on Ports 23 and 445 implies that attackers prefer to exploit services that are either under protected or employ weak protocols. Telnet (Port 23), despite being an older protocol, is still commonly used in some situations and is frequently unencrypted, making it an easy target for hackers. SMB (Port 445), on the other hand, is being targeted because of vulnerabilities such as MS17-010, which can be used to transmit malware like WannaCry. Attacks on Port 3306 (MySQL) indicate efforts to access databases, which may result in the theft of sensitive data if database security measures are inadequate. Port 1433, which is commonly used for SQL Server, is also being targeted, presumably due to insecure settings or the continuous use of default credentials. Finally, while Port 53281 receives fewer attacks, it should still be monitored since it may suggest attempts to compromise lesser-known services or applications. These findings highlight the need for stronger security measures, such as fixing vulnerabilities, encrypting data, and safeguarding service configurations, particularly for critical ports that are regularly targeted by attackers.

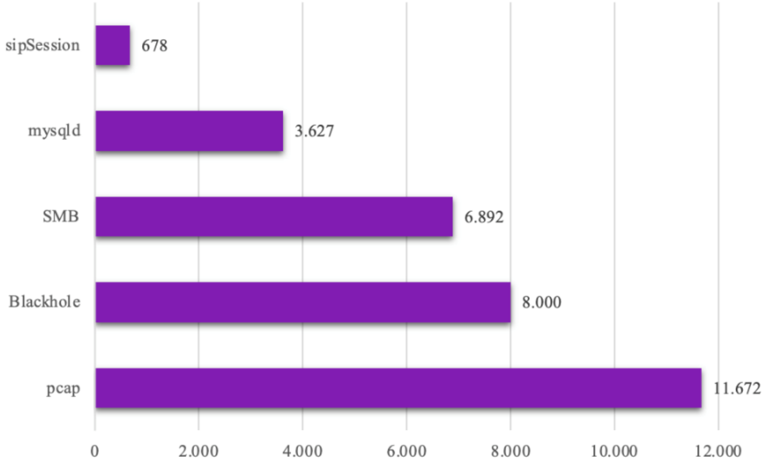


***Figure 7**. Number of services attacked.*

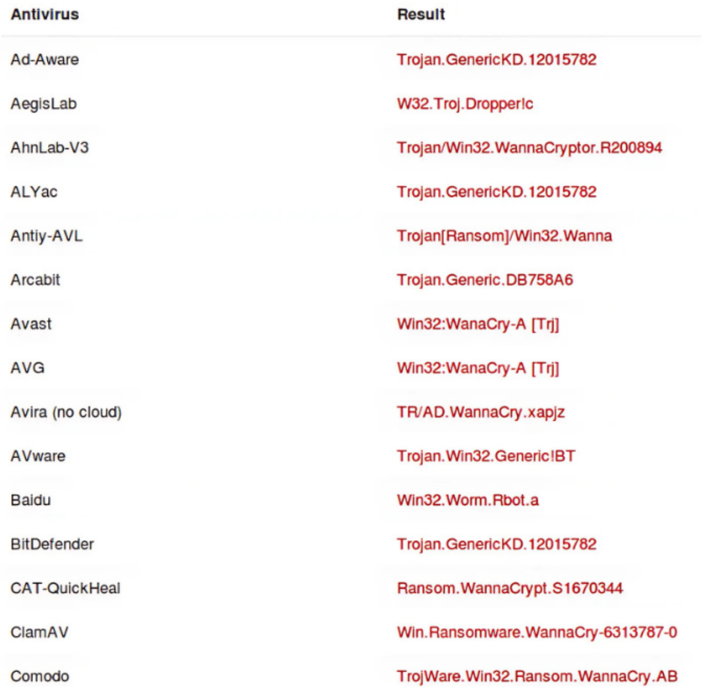
Attack analysis by service in Figure 7 demonstrates that PCAP was the major target, with 11,672 attacks, followed by Blackhole with 8,000 attacks. Furthermore, SMB logged 6,892 assaults, MYSQLD had 3,627 attacks, and sipSession had 678 attacks. The large frequency of PCAP assaults suggests that attackers attempted to collect and analyse network traffic in order to get sensitive information, such as login credentials or unencrypted data packets. Attacks on Blackhole indicate the exploitation of systems with vulnerabilities or that are not constantly updated, which could serve as entry points for future malware distribution.

Meanwhile, the attacks on SMB (6,892 cases) demonstrate that this protocol remains a significant target, owing to its extensive use in enterprise systems and the possibility of exploitation by ransomware such as WannaCry. Attacks against MYSQLD (3,627 incidents) indicate brute-force efforts or exploitation of database flaws, which could result in unauthorised access and data theft. Finally, the use of sipSession demonstrates that IP-based communication infrastructures are open to eavesdropping and abuse for unlawful VoIP calls. These findings highlight the continuous need for strong security practices, such as effective patch management, encryption, and secure communication protocols, to reduce the danger of these assaults.

4.2 Verification and malware analysis with VirusTotal

According to the verification results from VirusTotal connected to the honeypot, one of the analysed samples revealed that 58 out of 67 antivirus engines correctly classified the WannaCry sample as a threat in Figure 8. This score indicates a fairly high detection rate, as most security systems are capable of recognising malware.

The investigation found that different antivirus companies assigned different labels to the WannaCry sample. For example, Avast classified it as "Win32:WannaCrypt-A [Trj]", whilst AVG identified it as "TR/AD.WannaCry.xapz". Other vendors, such as BitDefender and Comodo, successfully detected the virus, labelling it as "Trojan.GenericKD.12015762" and "TrojWorm.Win32.Ransom.WannaCry.AB", respectively. The high detection rate suggests that the WannaCry sample is well-known in the cybersecurity world, allowing many antivirus engines to recognise it correctly. However, 9 antivirus engines failed to detect the sample, demonstrating the disparity in detection capabilities between providers. This highlights the need of employing a multi-layered security strategy to provide more comprehensive protection against evolving threats.



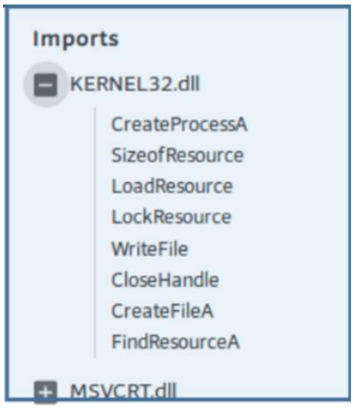
Antivirus	Result
Ad-Aware	Trojan.GenericKD.12015782
AegisLab	W32.Troj.Dropper!c
AhnLab-V3	Trojan/Win32.WannaCryptor.R200894
ALYac	Trojan.GenericKD.12015782
Antiy-AVL	Trojan[Ransom]/Win32.Wanna
Arcabit	Trojan.Generic.DB758A6
Avast	Win32:WanaCry-A [Trj]
AVG	Win32:WanaCry-A [Trj]
Avira (no cloud)	TR/AD.WannaCry.xapjz
AVware	Trojan.Win32.Generic!BT
Baidu	Win32.Worm.Rbot.a
BitDefender	Trojan.GenericKD.12015782
CAT-QuickHeal	Ransom.WannaCrypt.S1670344
ClamAV	Win.Ransomware.WannaCry-6313787-0
Comodo	TrojWare.Win32.Ransom.WannaCry.AB

***Figure 8**. Malware verification using VirusTotal.*

4.2.1 PE Malware sample analysis

Based on VirusTotal's additional analysis of the Portable Executable (PE) structure in Figure 9, the analysed malware sample employs KERNEL32.dll as one of its core libraries to perform a variety of potentially dangerous system activities. Several API methods imported from KERNEL32.dll include CreateProcessA, CreateFileA, WriteFile, and CloseHandle, indicating that this malware can start new processes, read or create files, write data, and control system resources.

This analysis reveals that the malware is designed to interact deeply with the system's core functionality, highlighting its ability to perform malicious activities such as process manipulation, file modification, and unauthorised data handling, all of which are common characteristics of ransomware and trojans.



Imports

⊟ KERNEL32.dll

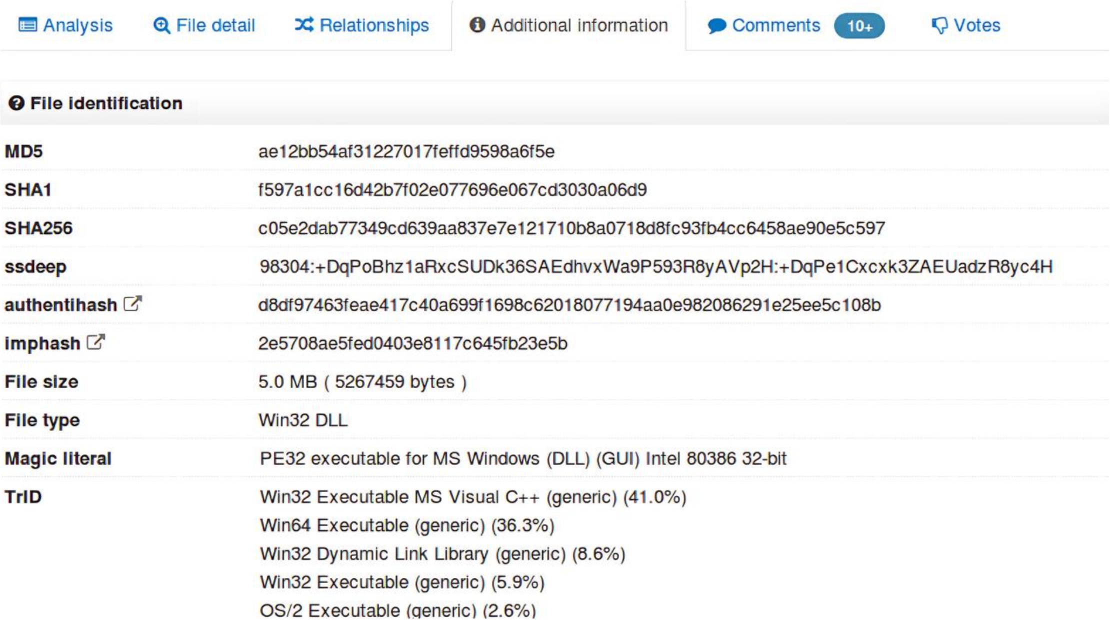
- CreateProcessA
- SizeofResource
- LoadResource
- LockResource
- WriteFile
- CloseHandle
- CreateFileA
- FindResourceA

⊞ MSVCRT.dll

***Figure 9**. Portable executable structure.*

Additionally, the presence of functions such as LoadResource, LockResource, and FindResourceA suggests that this malware may utilize embedded resources within its binary, possibly to extract additional payloads or execute code directly from memory. This is commonly seen in fileless malware techniques, where malicious code can execute without writing additional files to the system, making it harder to detect using traditional file-based detection methods.

Although MSVCRT.dll is also present in the imported libraries list, its specific functions are not shown in the screenshot. However, this library often includes normal C runtime functionality, which is frequently utilised for text and memory processing operations. The presence of these important functions imported from KERNEL32.dll implies that the virus can execute system-level code, modify files, and control memory resources. The use of APIs such as CreateProcessA shows that the malware can generate new processes, a behaviour commonly utilised in persistence or propagation techniques, such as running copies of itself or launching extra processes to escape detection by security systems. These findings underscore the malware's complex nature, as it can function discreetly by using memory-based activities and changing system-level resources to persist or proliferate over the network.



File identification

MD5	ae12bb54af31227017feffd9598a6f5e
SHA1	f597a1cc16d42b7f02e077696e067cd3030a06d9
SHA256	c05e2dab77349cd639aa837e7e121710b8a0718d8fc93fb4cc6458ae90e5c597
ssdeep	98304:+DqPoBhz1aRxcSUDk36SAEdhvxWa9P593R8yAVp2H:+DqPe1Cxcxk3ZAEUadzR8yc4H
authentihash	d8df97463feae417c40a699f1698c62018077194aa0e982086291e25ee5c108b
imphash	2e5708ae5fed0403e8117c645fb23e5b
File size	5.0 MB (5267459 bytes)
File type	Win32 DLL
Magic literal	PE32 executable for MS Windows (DLL) (GUI) Intel 80386 32-bit
TrID	Win32 Executable MS Visual C++ (generic) (41.0%) Win64 Executable (generic) (36.3%) Win32 Dynamic Link Library (generic) (8.6%) Win32 Executable (generic) (5.9%) OS/2 Executable (generic) (2.6%)

***Figure 10**. Detailed information.*

4.2.2 Malware detailed information

Based on further research of the VirusTotal report, the analysed malware sample contains multiple cryptographic hash values, such as MD5, SHA1, and SHA256, as indicated by VirusTotal in Figure 10. These identifications are used to confirm the legitimacy and unique identity of the examined file. The availability of these various forms of hashes is critical in the malware analysis process since it enables for the rapid detection of known variants and comparison of the sample to current threat databases. Using MD5 and SHA256, security professionals can compare this malware to previously detected samples in various cybersecurity archives. Furthermore, the file's metadata indicates that it is 5.0 MB (5,267,459 bytes) in size and is classified as a Win32 DLL, implying that this malware operates in the Windows environment and is most likely structured as a dynamic-link library (DLL) that can be injected into other processes or used to execute malicious code. The Magic Literal and TrID sections show that this file is a PE32 executable for Windows (DLL) with a 32-bit architecture. This shows that the infection was most likely constructed with Microsoft Visual C++, a popular development tool for generating sophisticated malware. These details help to better comprehend the malware's design, capabilities, and possible impact on affected computers.

5 DISCUSSION

In this research, the DORA captured 32,395 connections and 2,892 malware samples, offering valuable information into current cyber threat patterns. For example, the dominance of attacks on port 445 (SMB) and the discovery of 1,076 WannaCry malware infections demonstrate that the MS17-010 vulnerability is still a primary attack vector. These findings confirm that honeypots not only work as detection tools but also serve as significant data sources for understanding the techniques and targets used by attackers. Although honeypots excel in collecting attack data with a low false positive rate, they also have limitations. Since they only capture attacks specifically targeting their system, honeypots may not fully represent the entire range of threats within a network. To enhance their effectiveness, future research could integrate honeypots with machine learning for real-time attack analysis or expand the network coverage by implementing multiple honeypots in various configurations. With this approach, honeypots will continue to be an important tool in understanding and mitigating increasingly complex cyber threats.

Furthermore, the use of honeypots must also be considered within the context of regulations and ethics. Data collection by honeypots should comply with applicable privacy and security regulations, avoiding potential misuse of the collected data. Therefore, honeypot implementation should be supported by clear policies and procedures to ensure responsible use. As such, honeypots will continue to be a valuable tool in efforts to understand and mitigate cyber threats. Through further development and integration with other technologies, honeypots can make significant contributions to safeguarding network security and protecting IT infrastructure from cyberattacks.

6 CONCLUSION

This study introduces and demonstrates the effectiveness of DORA, a new approach to integrating the capability of Dionaea in data collection of malware attacks and VirusTotal’s features of in-depth analysis and reporting of the malware through exploiting VirusTotal API.

Based on our 24/7 observation period, the DORA captured 32,395 connections from 6,276 attacker IP addresses and collected 2,892 malware samples, of which 1,401 were successfully analysed, and 369 were confirmed as malware based on VirusTotal analysis. The five most dominant types of malware identified in the study were WannaCry (1,076 cases), Generic Malware (405 cases), Trojan Agent-AYFU (94 cases), MySQL UDF SYS library (PUA) (87 cases), and Conficker-A (87 cases). These findings highlight that the exploitation of SMB (MS17-010) through WannaCry remains a significant threat, while the presence of generic malware and trojans suggests that network exploitation techniques continue to be diverse. DORA is also able to show a detailed report of each malware attack, such as MD5, SHA1, SHA256, metadata and many more in real time using VirusTotal’s API.

These findings demonstrate that DORA has successfully integrated Dionaea as a beneficial passive monitoring tool with VirusTotal features in analysis and reporting, resulting in a novel real-time malware attack surveillance and threat intelligence platform. Future development will focus on using real-time machine learning approaches to improve DORA's threat detection and response capabilities.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflict of Interests: The authors declare no conflict of interest

Author Contributions: H.H.: Conceptualization, Formal analysis, Methodology, Writing – Original draft preparation. A.S.: Writing – Reviewing and Editing. A.A.: Writing – Reviewing and Editing. I.S.: Supervision, Writing – Reviewing and Editing.

Statement on the Use of Artificial Intelligence Tools: The authors declare that they didn't use artificial intelligence tools for text or other media generation in this article.

Data Availability: The data that support the findings of this study are available from the corresponding authors.

REFERENCES

Bartwal, U., Mukhopadhyay, S., Negi, R., & Shukla, S. (2022). Security Orchestration, Automation, and Response Engine for Deployment of Behavioural Honeypots. In *2022 IEEE Conference on Dependable and Secure Computing (DSC)*. IEEE. https://doi.org/10.1109/DSC54232.2022.9888808

BSSN. (2024). Lanskap Keamanan Siber Indonesia 2024. *Ilmu Bersama*. https://ilmubersama.com/2025/03/30/lanskap-keamanan-siber-indonesia-2024-bssn/

Commey, D., Hounsinou, S., & Crosby, G. V. (2024). Strategic Deployment of Honeypots in Blockchain-based IoT Systems. In *2024 IEEE 6th International Conference on AI Circuits and Systems (AICAS)*. IEEE. https://doi.org/10.1109/AICAS59952.2024.10595866

Holbel, R., Yerby, J., & Smith, W. (2024). Utilizing Virtualized Honeypots for Threat Hunting, Malware Analysis, and Reporting. *Issues In Information Systems*, 25(1), 265–278. https://doi.org/10.48009/1_iis_2024_122

Huang, C., Han, J., Zhang, X., & Liu, J. (2019). Automatic identification of honeypot server using machine learning techniques. *Security and Communication Networks*, 2019, Article 2627608. https://doi.org/10.1155/2019/2627608

Kristyanto, M. A., & Louk, M. H. L. (2024). Evaluation and Comparison of the Use of Reinforcement Learning Algorithms on SSH Honeypot. *Teknika*, 13(1), 77–85. https://doi.org/10.34148/teknika.v13i1.763

Liu, S., Wang, S., & Sun, K. (2023). Enhancing Honeypot Fidelity with Real-Time User Behavior Emulation. In *Proceedings - 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume*, (pp. 146–150). IEEE. https://doi.org/10.1109/DSN-S58398.2023.00041

Maharani, F., Kalsum, T. U., & Alamsyah, H. (2024). Penerapan Honeypot Sebagai Sistem Keamanan Server Berbasis Linux. *Jurnal Amplifier: Jurnal Ilmiah Bidang Teknik Elektro Dan Komputer*, 14(2), 174–183. https://doi.org/10.33369/jamplifier.v14i2.38240

Martínez, S.C.J., Moreno A., H. O., & Hernández A., M. B. (2023). Analysis of Intrusions into Computer Systems using Honeypots. *International Journal of Intelligent Systems and Applications in Engineering*, 11(6s), 461–472.

Morić, Z., Dakić, V., & Regvart, D. (2025). Advancing Cybersecurity with Honeypots and Deception Strategies. *Informatics*, 12(1), Article 14. https://doi.org/10.3390/informatics12010014

Naik, N., & Jenkins, P. (2018). A Fuzzy Approach for Detecting and Defending Against Spoofing Attacks on Low Interaction Honeypots. In *2018 21st International Conference on Information Fusion,* (pp. 904–910). IEEE. https://doi.org/10.23919/ICIF.2018.8455555

Njoera, Y.A.D., Hartawan, I.N.B., Ariana, A.A.G.B, & Krisna, E.D. (2024). The Analysis of Honeypot Performance Using Grafana Loki and ELK Stack Visualization. *Informatika Dan Sains*, 14(3), 297–309.

Patel, P., Dalvi, A., & Sidddavatam, I. (2022). Exploiting Honeypot for Cryptojacking: The other side of the story of honeypot deployment. In *2022 6th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2022*. IEEE. https://doi.org/10.1109/ICCUBEA54992.2022.10010904

Radoglou-Grammatikis, P., Sarigiannidis, P., Diamantoulakis, P., Lagkas, T., Saoulidis, T., Fountoukidis, E., & Karagiannidis, G. (2024). Strategic Honeypot Deployment in Ultra-Dense beyond 5G Networks: A Reinforcement Learning Approach. *IEEE Transactions on Emerging Topics in Computing*, 12(2), 643–655. https://doi.org/10.1109/TETC.2022.3184112

Ryandy, Lim, C., & Silaen, K. E. (2020). XT-Pot: eXposing Threat Category of Honeypot-based attacks. In *Proceedings of the 2020 International Conference on Engineering and Information Technology for Sustainable Industry*, (Article 31). ACM. https://doi.org/10.1145/3429789.3429868

Saikawa, K., & Klyuev, V. (2019). Detection and Classification of Malicious Access using a Dionaea Honeypot. In *10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS).* IEEE. https://doi.org/10.1109/IDAACS.2019.8924340

Shahrivartehrani, S., & Abidin, S. (2016). Dionaea Honeypot Implementation and Malware Analysis in Cloud Environment. *Journal of Computing Technologies and Creative Content*, 1(1), 1–5.

Shi, L., Li, Y., Liu, T., Liu, J., Shan, B., & Chen, H. (2019). Dynamic Distributed Honeypot Based on Blockchain. *IEEE Access*, 7, 72234–72246. https://doi.org/10.1109/ACCESS.2019.2920239

Siddiqui, M. A., & Bokhari, M. U. (2021). Honeypot-Based Intrusion Detection System: A Performance Analysis. *International Journal of Enhanced Research in Management & Computer Applications*, 10(7), 1–7.

Syamsuddin, I., & Barukab, O. M. (2022). SUKRY: Suricata IDS with Enhanced kNN Algorithm on Raspberry Pi for Classifying IoT Botnet Attacks. *Electronics*, 11(5), Article 737. https://doi.org/10.3390/electronics11050737

Tabari, A. Z., & Ou, X. (2020a). A First Step Towards Understanding Real-world Attacks on IoT Devices. *arXiv:2003.01218*. https://doi.org/10.48550/arXiv.2003.01218

Tabari, A.Z., & Ou, X. (2020b). A Multi-phased Multi-faceted IoT Honeypot Ecosystem. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security,* (pp. 2121–2123). ACM. https://doi.org/10.1145/3372297.3420023

Thom, J., Shah, Y., & Sengupta, S. (2021). Correlation of Cyber Threat Intelligence Data across Global Honeypots. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference,* (pp, 766–772). IEEE. https://doi.org/10.1109/CCWC51732.2021.9376038

Wang, B., Dou, Y., Sang, Y., Zhang, Y., & Huang, J. (2020). IoTCMal: Towards A Hybrid IoT Honeypot for Capturing and Analyzing Malware. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. IEEE. https://doi.org/10.1109/ICC40277.2020.9149314

Wang, M., Santillan, J., & Kuipers, F. (2018). ThingPot: an interactive Internet-of-Things honeypot. *arXiv:1807.04114*. http://arxiv.org/abs/1807.04114

Yang, X., Yuan, J., Yang, H., Kong, Y., Zhang, H., & Zhao, J. (2023). A Highly Interactive Honeypot-Based Approach to Network Threat Management. *Future Internet*, 15(4), Article 127. https://doi.org/10.3390/fi15040127

Zhang, W., Zhang, B., Zhou, Y., He, H., & Ding, Z. (2020). An IoT Honeynet Based on Multiport Honeypots for Capturing IoT Attacks. *IEEE Internet of Things Journal*, 7(5), 3991–3999. https://doi.org/10.1109/JIOT.2019.2956173

Zhang, Y., Zhang, H., Yuan, X., & Tzeng, N. F. (2019). Pseudo-Honeypot: Toward Efficient and Scalable Spam Sniffer. In *Proceedings – 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2019,* (pp. 435–446). IEEE. https://doi.org/10.1109/DSN.2019.00052

Zia R.S.M., Uddin, M. J., & Islam, A. (2019). Know Your Enemy: Analysing Cyber-Threats Against Industrial Control Systems Using Honeypot. In *2019 IEEE International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things*, (pp. 151–154). IEEE. https://doi.org/10.1109/RAAICON48939.2019.69

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951