SGU Master IT
Data Science
Cyber Security

Empowering the next generation of cyber defenders with data analytics skills

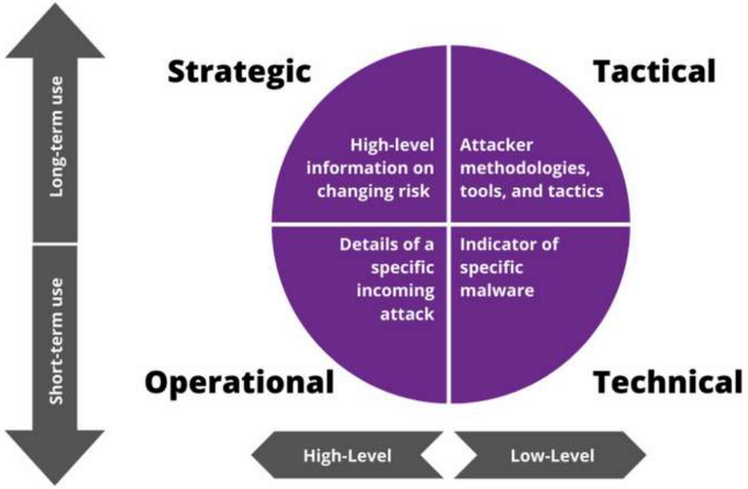
MOU SGU & BSSN | 24 November 2018 | Tangerang, Indonesia

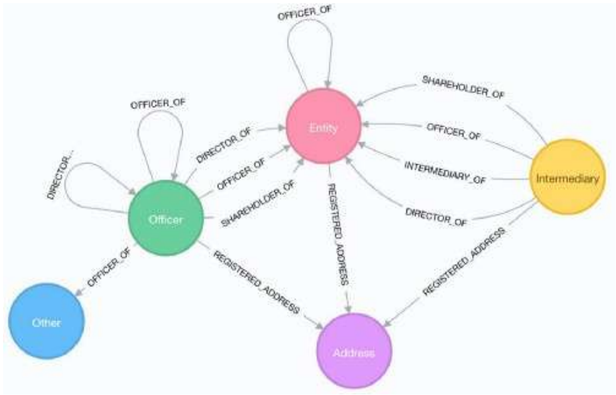
Research Collaboration

- **MALWARE**
- **THREAT INTELLIGENCE**

Our Research Focus

- Malware Analysis
- Threat Detection
- Threat Intelligence
- Knowledge Graph





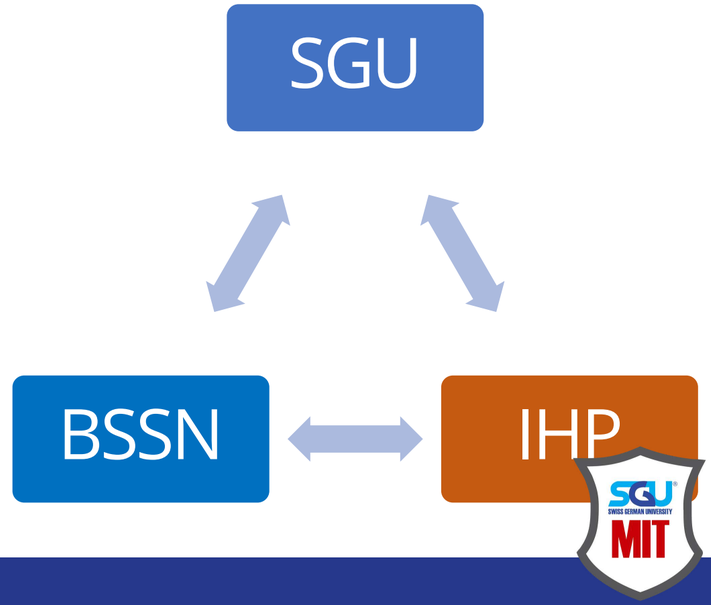
Our Security Operations Center

- SIEM
- Ticketing Systems
- Threat Intelligence Sharing
- Honeynet-based Threat Sharing

ISIF Asia Research Grant 2019, 2020, 2021

This year research:

1. A **more robust repository platform** for processing and storing broader range of honeynet-based threat information
2. A **highly available data lake platform** that allows security analyst perform threat correlation between honeynet threat information with existing OSINT.
3. A **higher quality threat information**, including description, scoring and analysis report, which is generated automatically by system in the help of security analyst, allowing organizations to easily share and exchange security threat information with other organizations.



Industry Scholarship 100%

Our Publications

XT-Pot: eXposing Threat Category of Honeypot-based attacks

Ryandy
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
ryandy@student.sgu.ac.id

Charles Lim
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
charles.lim@sgu.ac.id

Kalpin Erlangga Silaen
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
kalpin.erlangga@lecturer.sgu.ac.id

ABSTRACT

As organization infrastructure is getting more complex to support its business, cyber security threat monitoring on the infrastructure for the emerging threats becomes essential. Honeypot, a decoy system, when properly deployed in the organization's network provides valuable insight into the behavior of attacker to the organization. In this research, we propose a generic framework to analyze and categorize threats collected from honeypots. These threat categories become the building block of threat intelligence to be shared used by security analyst in handling security incidents.

CCS CONCEPTS

• **Security and privacy → Software and application security**; *Distributed systems security; Intrusion/anomaly detection and malware mitigation.*

KEYWORDS

Honeypot, Threat categorization, Threat Analysis, Malware analysis

ACM Reference Format:

Ryandy, Charles Lim, and Kalpin Erlangga Silaen. 2020. XT-Pot: eXposing Threat Category of Honeypot-based attacks. In *ICONETSI '20: The International Conference on Engineering and Information Technology for Sustainable Industry, September 28–29, 2020, Tangerang, Indonesia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Malicious software (malware) has been one of the highest cyber security threats to organizations around the world that rely on Internet to perform their businesses. Purple sec [17] reported an exponential increase of malware volume as well as sophistication of malware used to attack these organizations. The adversaries have been taking advantage of human weaknesses, e.g. the desire of using free software including pirated software, to infect individual or business computers to achieve their goals. In this case, Trojan malware is among the popular malware used by adversaries to disguise itself as legitimate software [6] in the victim computers. The attackers are preparing to launch the next attack from these computers as they work their way into organizations to perform their malicious intents, including compromising business critical systems.

Another common method used by the adversaries is to deploy malware behind the website that looks legitimate for users to explore. With this strategy, malware packed with various malicious payload spreads even faster to many parts of the world. To start the process, attackers use various methods, and among them is to share some interesting link to the community about the latest event and wait for the victim to open the link. With today vast and inexpensive cloud infrastructure, deploying these systems in large scale takes minutes if not few hours of installation and configuration. In addition, these systems become the minefield for the attackers to design more sophisticated malware to probe more information to the targeted individuals or organizations.

To understand these growing threats, honeypot [19] has emerged to be the forefront of security researcher tools to capture the attackers' activities in the organizations. This decoy system, placed in the unused IP address of organization infrastructure, i.e. darknet [4], is designed to lure and capture adversaries' behavior in the acts. Hence, any connection attempt to this unused IP will be deemed as malicious. Honeypots are distinguished by the emulated services they offer to response to any of the possible connection request by the attackers, and they include common well known services such as SMB, HTTP, SSH, FTP, TFTP, MySQL, MSSQL, SIP and others.

Once the threat behaviors are collected, they need to be processed and analyzed to gain insight into what the attackers are up to and possible ramification of the attacks. Possible attack sequence including commands, scripts, and other malicious payload are some of the critical information need to be further analyzed and categorized. This threat information, also called threat intelligence, is useful for threat analyst in the organization to mitigate the threat risks caused. This research will introduce XT-Pot framework to capture attack patterns, analysis of the capability of these threats, and threat categories for each of the attacks we learned from the honeypots.

Following are the contribution of this research:

- We introduce a framework to analyze and generate the security threat information based on the honeypots traffic.
- We propose a taxonomy of threat categories and capabilities, encompassing cyber security attacks to several well known services provided by honeypots.

2 RELATED WORKS

2.1 Honeypot Threat and Processing

The first generation Honeypot was introduced in 1999 [19], with the purpose of capturing malicious activity of the adversaries. The

ICONETSI '20, September 28–29, 2020, Tangerang, Indonesia

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

XT-Pot: EXposing Threat Category of Honeypot-Based Attacks

XB-Pot: Revealing Honeypot-based Attacker's Behaviors

Ryandy[1], Charles Lim[2], Kalpin Erlangga Silaen[3]
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
[1]ryandi@student.sgu.ac.id, [2]charles.lim@sgu.ac.id,
[3]kalpin.erlangga@lecturer.sgu.ac.id

Andi Yusuf
Threat Detection Directorate
National Cyber and Crypto Agency
Jakarta, Indonesia
andi.yusuf@bssn.go.id

***Abstract*—Since its introduction, honeypot has been used by researcher to track and learn the cyberattack into organization infrastructures. With the continuous rise of cyberattack, deception technology, i.e., honeypot, has been eyed by organizations as a prominence tool to provide early detection of attack capability and defense mechanism after learning from the interaction between the attacker and the tool. In this research, a new enhanced framework is introduced to categorize attacker behaviors detected through our honeypots. The framework provides a finer-grained result allowing representation of the actual attacker behaviors as he/she interacts with the honeypot. Complete threat categories both on high-volume and low-volume attack traffic are presented.**

***Index Terms*—honeypot, deception technology, threat detection, mitre attack.**

I. Introduction

Since Honeypot, a decoy system designed to track attacker's interaction with the system under investigation, was introduced in early 2000 [1], it has been used by security researchers to study security attacks into organization digital assets. As the world grappled with the pandemic, organizations around the world are shifting the gear to transform itself to a more digital organization to survive and to be more sustainable in the long term. As reported by Globe Newswire [2], the use of Deception Technology, such as commercial honeypots, in the organizations is on the rise and the market is project to hit around USD 4.7 billion value in 2027. Honeypot allows organization to detect accurately early phase of cyberattack timely, helping organization to reduce overall cyber security risks.

Fan et al. [3] described the building blocks of honeypot: decoy and captor. While decoy is an information system resources that system is trying to furnish, captor is security-related functions, such as profiling, monitoring, response, etc. The degree of decoy is categorized in terms of how system interacts with the attacker: low, medium and high. Low Interaction Honeypot (LIH), also known as facade, emulates certain protocols, provides very small interaction to the attacker and capture network traffic during the interaction. Medium Interaction Honeypot (MIH), on the other hand, implements various applications that provide fake response to the incoming attack. The interaction is recorded by captor into a log file and binaries involved can be saved for further investigation. Finally, High Interaction Honeypot (HIH), furnishes a fully functional system that is capable of capturing the interactions and isolating the outgoing network traffic. LIHs and MIHs could be deployed in a large scale encompassing different countries around the world, HIHs, however, are deployed across different networks in the organizations for a limited time due to its higher possibility of compromise.

As discussed by McGrew [4], it is important not only to understand what the attacker is doing to our honeypots but also to learn the intention of the attack. Our previous work [5] introduced different threat categories for our LIHs and HIHs, i.e. Cowrie and Dionaea, based on network traffic and payload artifacts. This research extends our previous work to include additional services being monitored as well as deeper analysis into the attacker behaviors to provide another level of understanding of the attacker intention.

Attacker may use a more high-volume type of attacks, i.e. to increase the success of the attacks,or the attacker may device a more targeted attacks such as Advanced Persistent Threat (APT) attack, aimed at a certain country, organization or even certain individual. Hutchins et al. [6] discussed Cyber Kill Chain framework to better map out attacker's behaviors and phases during any of these attacks. To provide a more comprehensive model to cover end-to-end attacker activities, mitre developed MITRE att&ck framework [7] to map out tactics, techniques and procedures the attacker took to launch his/her attack.

Since APT attack uses persistence strategy, very low activity can be seen throughout the attack process, hence the term, slow and low. Therefore, it is important for researcher to pay attention not only on the high-volume attack but also on the low-volume, so called long tail activities. Anderson [8] described long-tail as the portion of the distribution of random events away from the "head" portion over time under observation. This concept is especially important for threat analysis, in which many analysts tend to pay attention on the high-volume of events while ignoring low-volume events during the investigation. This research will expose both high-volume and low-volume (long tail) attacks, fostering a complete spectrum of threat analysis of honeypot based attack data.

XB-Pot: Revealing Honeypot-based Attacker's Behaviors

Mapping Linux Shell Commands to MITRE ATT&CK using NLP-Based Approach

Yevonnael Andrew
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
yevonnael.andrew@student.sgu.ac.id

Charles Lim
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
charles.lim@sgu.ac.id

Eka Budiarto
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
eka.budiarto@sgu.ac.id

***Abstract*—Honeypot is a decoy computer resource used to trap an attacker and one of the most common honeypots is a medium to high interaction honeypot that is able to log shell interaction executed by an attacker. In the cybersecurity field, these collected commands can be mapped to MITRE ATT&CK, a knowledge base, and model for cyber adversary behavior. To maximize the utilization of ATT&CK, a good mapping between Linux commands and ATT&CK is desirable. In this paper, we evaluate and measure the performance of Linux commands mapping to ATT&CK Techniques and Sub-Techniques, and ATT&CK Tactics using NLP techniques, such as Bag of Words, TF-IDF, and pre-trained Word Embeddings. Cosine similarity scoring is used to extract the top-n ATT&CK Techniques and Sub-Techniques, and ATT&CK Tactics for each command. The models' performance is evaluated using *recall at n* metrics.**

***Index Terms*—cybersecurity, honeypot, MITRE ATT&CK, machine learning, NLP, word embeddings**

I. Introduction

Honeypot is defined as a decoy computer resource in which the value arises from being probed, attacked, or compromised [1]. One popular honeypot available on the market is a Cowrie [2], a medium to high interaction SSH/Telnet honeypot, which is able to log shell interaction, i.e., commands executed by the attacker. By analyzing honeypot logs, researchers are able to learn malicious attackers' activities, including possibly malicious insiders, to understand their intent.

MITRE ATT&CK [3] is a curated knowledge base and model for cyber adversary behavior and taxonomy across their lifecycle. ATT&CK for Enterprise is a version of ATT&CK that covers behavior against enterprise IT networks and clouds. ATT&CK comprises Tactics, Techniques, and Procedures (TTP): Tactics represent the reason for the attack; Techniques specify how an adversary achieves its goal; Procedures describe the specific implementation of techniques or sub-techniques used during the attack.

To achieve a particular goal, the attackers may use various shell commands inside the system they are to compromise. Attackers usually perform several activities, such as profiling the hardware and software, downloading and executing the payloads, changing permissions, and even removing their footprints before they end the execution [4]. Hence, an accurate mapping of these Linux bash commands to MITRE ATT&CK TTP [5] can precisely describe the attacker's behaviors during its presence in the compromised systems.

In this paper, we evaluate Natural Language Processing (NLP) capability to support our research in mapping Linux bash command to its corresponding MITRE ATT&CK Tactics and Techniques, particularly one-to-many mapping based on text similarity performance. A collection of Linux bash commands and their relevant descriptions and MITRE ATT&CK descriptions are gathered, several pre-processing techniques to remove stopwords and symbols are performed, and the resulting sentences are projected into vectors using various methods: Bag of Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), and Word Embeddings. Finally, the similarity scoring between Linux bash command to ATT&CK Techniques and Tactics descriptions is calculated. With the help of domain expert knowledge, the unsupervised model can be properly evaluated.

The remainder of this paper is organized as follows: In section II, various related research on shell command processing and NLP-based approaches are presented. The next section III discusses the research framework of our work. Section IV and V present the experiment results and discussions on the results respectively. Finally, section VI provides the concluding remarks of our research.

II. Related Works

Boffa et al. [6] treat shell commands like natural language, parse the shell commands, options, and parameters, and project them into Bag of Words and Word2Vec. Trizna et al. [7] extracts features from raw shell commands and use them as input for machine learning models. These research results look promising; however, they do not consider the meaning of that particular syntax.

Hussain et al. [8] take the command description from the documentation and provide scoring on the similarity of the command using the NLP approach. However, this research only measures the similarity between commands, not an external framework like MITRE ATT&CK. Kanakogi et al. [9] map automatically CAPEC from CVE and measure similarity using TF-IDF, USE, and SBERT. Shahid et al. [10] explore the textual description of vulnerability disclosure mapped to CVSS using NLP BERT. For the explainability aspect, the top

Mapping Linux Shell Commands to MITRE ATT&CK using NLP-Based Approach

Our Publications

Mapping Threats in Smart Grid System Using the MITRE ATT&CK ICS Framework

Ahmad Balya Izzuddin
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
ahmad.izzuddin@student.sgu.ac.id

Charles Lim
Information Technology Department
Swiss German University
Tangerang, Banten, Indonesia
charles.lim@sgu.ac.id

***Abstract*—The smart grid system is an integration between power distribution systems with communication networks. A smart grid offers various benefits, but at the same time inherits various vulnerabilities from the implemented information and communication technology (ICT). Many devices in smart grid systems implement the TCP/IP stack to exchange data, which can lead to significant new cyber attack vectors, such as malware, Denial-of-service (DoS), man-in-the-middle (MITM), and replay attacks, as well as various other cybersecurity threats. One approach to deal with these security issues proactively is through threat modeling. We can utilize some tools to gather the threat data targeting the smart grid, such as using honeypots, then analyze the collected threat data to obtain the threat model in order to study the attackers' behavior. In this paper, we collected threat data targeting the smart grid system by deploying GridPot honeypot and analyzed the collected threat data by mapping them to the MITRE ATT&CK for Industrial Control System (ICS) framework. This experiment shows that the threats targeting the smart grid systems are real, and could harm any smart grid system in the world.**

***Index Terms*—smart grid, industrial control system, GridPot, threat modeling, MITRE ATT&CK for ICS**

I. Introduction

A smart grid is an ICT-based system that contributes to the generation, distribution, and consumption of electrical energy [1]. A smart grid creates a broadly distributed and automated system with various new functions and features that allow optimization in the electrical industry by integrating with the communication network to build infrastructure that supports the two-way flow of information and power [1], [2]. By utilizing modern ICTs, such as industrial control system (ICS) devices, the smart grid is able to distribute power more efficiently and to respond to various situations and conditions [3], [4].

vectors, such as malware attacks, Denial-of-service (DoS) attacks, man-in-the-middle (MITM) attacks, replay attacks, and various other cybersecurity threats [1], [2], [4]. Given its status as critical infrastructure, the smart grid system often attracts attackers to initiate massive coordinated malicious activities. These attacks may aim to disrupt the smart grid, obtain sensitive data, and potentially cause massive damage to the distribution, transmission, and generation facilities. The attacks may be motivated financially, politically, or even possibly by terrorist ideology [2], [4].

There had been numerous research that provides various solutions for these security issues. Conventional cyber security techniques such as firewalls, IDS, IPS, malware protection, and many other techniques were proposed as solutions to this problem in many studies [5]–[7]. However, these solutions barely satisfy the requirements for securing smart grid communication systems operated over open networks such as the internet [2]. Another proposed solution is to implement honeypots or honeynet (a collection of honeypots), which is a common technique used in industrial networks to identify attacks, gather attack strategy intelligence, and mislead cybercriminals from attacking actual infrastructure [4]. Integrating honeynet with firewalls and IDSs may provide the opportunity to capture attacker behaviors and actions and to develop methods for system security improvement and attack prevention [8].

One approach to studying attacker behavior is through threat modeling, which includes identifying key assets in the system and threats to those assets [9]. MITRE ATT&CK for ICS is a framework for modeling cyber attacker behavior targeting the ICS domain [10]. This framework describes the various phases of the attacker's attack lifecycle along with the assets and systems identified as targets [10].

Mapping Threats in Smart Grid System Using the MITRE ATT&CK ICS Framework

Threat Analysis on Industrial Control System Based on Attacker's Behaviours using Honeypot

Arssy Hasyir Nursidiq
Information Technology Departement
Swiss German University
Tangerang, Banten, Indonesia
arssy.nursidiq@student.sgu.ac.id

Charles Lim
Information Technology Departement
Swiss German University
Tangerang, Banten, Indonesia
charles.lim@sgu.ac.id

***Abstract*—Industrial Control System (ICS) is a system that has been used in the maintenance and management of critical infrastructures. Which are usually distributed, such as power distribution system, gas, water, transportation, and production. The ICS is commonly having several components such as a Programmable Logic Controller (PLC), Human Machine Interface (HMI), Master Terminal Unit (MTU), and lastly Remote Protocol Unit (RTU). The protocols of Industrial Control System (ICS) is lack of authentication and confidentiality it causes the Industrial Control System easily compromised by the attacker. To get more knowledge about type of attack that compromised Industrial Control System (ICS) we used honeypot to emulate the Industrial Control System (ICS) and collect the data from the attacker. The data from honeypot will be analyzed and map to MITRE ATTACK to get knowledge about the Tactic, Technique, and the Procedure. The results, we can see what is the purpose from the attacker when they are compromise Industrial Control System (ICS) based on the data that has been mapped to MITRE ATTACK**

***Index Terms*—Honeypot, Industrial Control System, Threat Mapping, MITRE ATTACK**

I. Introduction

Since honeypot was introduced in early 2000 as a decoy system that tracked attacker interaction with the system under investigation [1] it has been used by security researchers to study the security attack on organizations' digital assets. When the Compound Annual Growth rate (CAGR) increases from 2016-to 2022, the Global Deception Market Size is predicted to hit 2,068.1 billion dollars in 2022[2]. Deception technology is part of the security techniques and tools used to prevent damage by an attacker. They use a decoy system to manipulate the attacker by preventing the system from deeper. The decoy use emulated either the real operating system or misleadingly tricks the hacker. The honeypot can use to manipulate the phase attacks because the honeypot only captures the attack that redirects to it, the honey pot can help the organization to reduce cyber security risks.

The decoy aims to capture the data from the attacker and use the data to learn about the behavior of the attacker, based on PurpleSec's study cybercrime has increased up to 600 percent during the pandemic COVID-19.

Honeypot was introduced in 2000[1] the honeypot has the aim to let the attacker lure our system and we can collect the data from their activities in our system for further analysis. Honeypot has little traffic in our system, but it has a high value, since the honeypot didn't depend on a signature, advanced algorithm, or rules it can exist an unknown attack by the attacker. Honeypot has two disadvantages first it only captures the data only by the attack that is directed at them and true of all security technology it means honeypot cannot replace all the existing security technology it is only used to improve and complement the security technology.

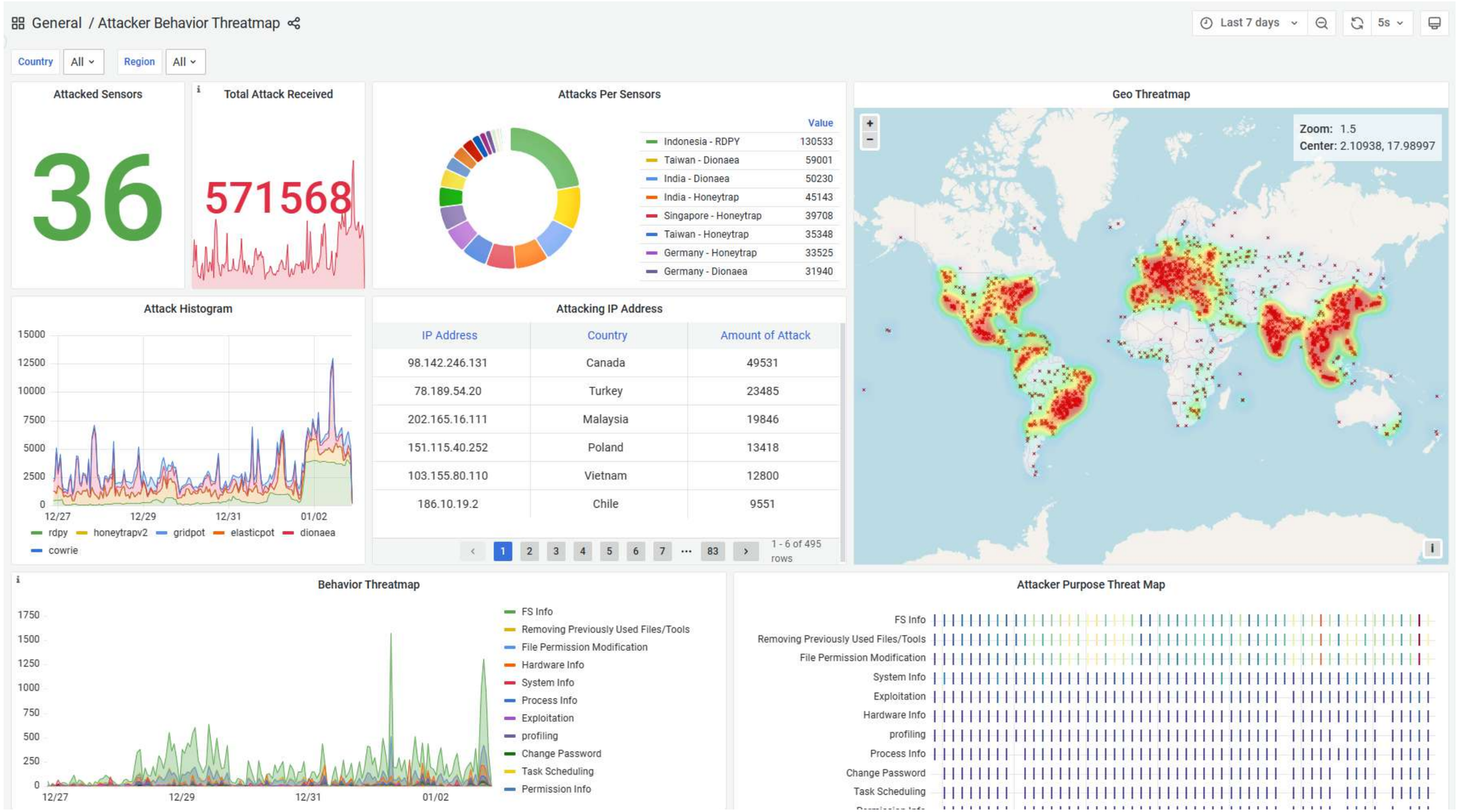
For the interaction, the honeypot is divided into two categories as Low Interaction Honeypot and High Interaction honeypot. The level of interaction stands for the interaction that the attacker has with the honeypot, it means that if the interaction is from the attacker with the honeypot is big enough it can be classified as a High Interaction Honeypot. Since the interaction is high then the data that we have becomes more valuable. Low Interaction Honeypot can emulate the system services. Otherwise, High Interaction Honeypot can emulate the system but by using the system itself they provide a real operating system with which to interact.

The honeypot can emulate the physical system of the Industrial Control System. In this direction, we use the honeypot to emulate the Industrial Control System infrastructure to collect

Threat Analysis on ICS Based on Attacker's Behaviors using Honeypot

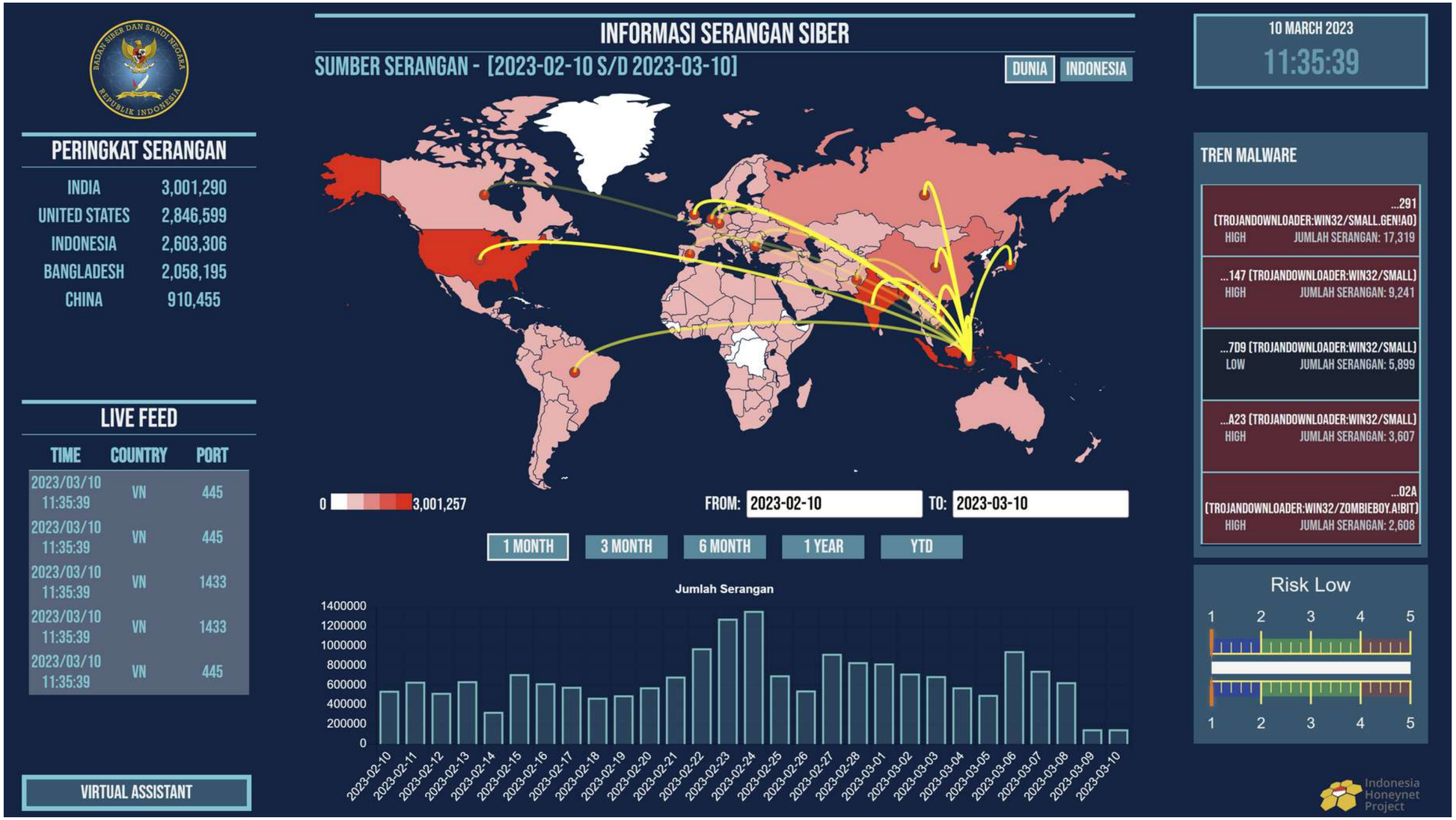
ISIF Award 2022

Our Public Dashboard (https://public.cscisac.org)



Our Public Dashboard (https://public.cscisac.org)

Our Public Contribution (https://Honeynet.bssn.go.id)



Why Data Science Cyber Security Master Program?

As businesses transform their business into digital enterprises, cyber security attacks become the largest threats to businesses and organizations worldwide. With the increasing amount of data and emerging technologies, it is essential to develop a strong defense against cyber threats.

The Data Science Cyber Security Master Program is designed to equip students with the necessary skills to become cyber security professionals who can defend against these threats.

The curriculum covers topics such as **network security, ethical hacking, data science, cryptography, and incident handling skills**. Students will learn through in class capstone projects how to analyze data sets to identify potential threats, design and implement security measures, and use machine learning algorithms to detect and detect/prevent cyber-attacks. The program also covers legal and ethical considerations in cyber security, including privacy laws and regulations.

Graduates of the program will be **well-equipped to work in a variety of industries, including finance, retail, healthcare, telecommunication, manufacturing, energy and government.** The Data Science Cyber Security Master Program is an excellent choice for students who are interested in pursuing a career in cyber security as it provides a strong foundation in data science and cyber security, making them valuable assets to any organization that needs to protect their data and systems from cyber threats.

Distinguished Strengths

- Award winning Cyber Security Lecturers and Researchers
- Industry capstone projects to bring real contributions to the workplace
- World class research projects with industry collaboration research including National Cyber and Crypto Agency (Badan Siber dan Sandi Negara), Indonesia Honeynet Project, Academy CSIRT and others.
- Research topics include: Deception Technology, Malware Analysis, Threat Detection, Threat Intelligence, Vulnerability Analysis, Digital Forensics, Cloud Security, etc.
- Security Operation Centre for handling real national security incidents
- EC Council Special Discounts for cyber security professional certifications

Key National/International Mentions

- ·National Honeynet-based threat map portal (https://honeynet.bssn.go.id)
- ·Honeynet Threat Sharing Platform (Multi Year Research Grants Recipients from The Information Society Innovation Fund - ISIF Asia)
- ·2022 ISIF Asia Award Recipients (https://isif.asia/2022-awardees/)