11-14-2024

Relationship of Cyber Threat Intelligence and Critical Infrastructure Assets on Information Technology Critical Infrastructure Attacks

Albert Su
Walden University

Walden University

College of Management and Human Potential

This is to certify that the doctoral study by

Albert Su

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Jon McKeeby, Committee Chairperson, Information Technology Faculty
Dr. Cheryl Waters, Committee Member, Information Technology Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2024

Abstract

Relationship of Cyber Threat Intelligence and Critical Infrastructure Assets on Information Technology Critical Infrastructure Attacks

by

Albert YL. Su

MS, Illinois Institute of Technology, 1994

BS, Purdue University, 1992

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Information Technology

Walden University

November 2024

Abstract

Ransomware cyberattacks on critical infrastructure can disrupt critical services with the capacity to impact human lives. Some security operations center (SOC) analysts feel overwhelmed by the influx of threat indicators of compromise (IoC), making it challenging to discern which alerts are significant for their organization. Grounded in John Boyd's OODA (Observe, Orient, Decide, Act) theoretical framework, the purpose of this quantitative correlational study was to examine the relationship between cyber threat intelligence (CTI) and critical infrastructure assets on information technology critical infrastructure attacks. Data were collected by ingesting information from a U.S. CTI vendor and alerting IoCs related to critical infrastructure assets from trusted open-source intelligence sources. The analysis using the Jaccard index and Cosine similarity revealed a strong, positive association between the size of CTI data and the number of matching Alert IoCs. When the CTI data increased from 8K to 328.7K IoCs, the number of matches rose significantly from 55 to 138. This relationship is supported by a high Jaccard index of 0.793 and a cosine similarity of 0.891, indicating strong similarity between CTI data and the resulting Alert IoCs. A key recommendation for IT leaders is integrating CTI into the organization's cybersecurity application with OODA loop automation, CTI threat monitoring, and meaningful alerts. The implications for positive social change include the potential to provide SOC analysts with fast-acting mitigation responses, preventing disruptions in critical infrastructure services to the local community.

Relationship of Cyber Threat Intelligence and Critical Infrastructure Assets on Information Technology Critical Infrastructure Attacks

by

Albert YL. Su

MS, Illinois Institute of Technology, 1994

BS, Purdue University, 1992

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Information Technology

Walden University

November 2024

Dedication

I want to dedicate this research to those who have served and their families for their support. A constant reminder of what a family means to me, to a daughter I strive to find, and to a place I call home.

Acknowledgments

I humbly thank my chair, Dr. Jon McKeeby, whose guidance was the North Star of my research, and respectively, the 2nd committee member, Dr. Cheryl Waters, and the program director, Dr. Gail Miles. A heartfelt acknowledgment to the community of support that allowed me to run each step of this long journey.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study

When the U.S. Energy Secretary Jennifer Granholm was asked in June 2021 if adversaries could disrupt the U.S. electric power grid, her reply was yes (Smith, 2021). Thousands of cyberattacks occur to key critical infrastructure such as the electric power grid consisting of over 7,000 power plants, 55,000 substations, 160,000 miles of high-voltage distribution lines, and millions of miles of low-voltage distribution lines. The foundation of the study was cybersecurity and the energy sector protection from evolving cyber threats.

Background of the Problem

The increasing concern about cyberattacks on U.S. energy infrastructure in early May 2021 with the Colonial Pipeline Company shutdown led to policy changes by President Biden (Smith, 2021). The resulting ransomware attack disrupted fuel supply to the U.S. East Coast for several days, affecting millions of people. Former director of the U.S. Department of Homeland Security (DHS) Cybersecurity and Infrastructure Security Agency (CISA), Chris Krebs, referred to the incident as cybercrime and a threat to national security. Colonial Pipeline paid a ransom of $4.4 million to cyber criminals to restore the system. In light of the attack, President Biden issued an executive order (EO) improving the nation's cybersecurity. EO 14017 and a House bill, the Industrial Control Systems Capabilities Enhancement Act of 2021 (H.R. 1833), were enacted in the United States in July 2021 and required guarding U.S. supply chains. The EO includes legislation for facilities that provide critical services and operational technologies (OT) (Ashley et al., 2022), addressing the issue of resilience. The issue of resilience includes

securing industrial control systems (ICS), supply chain compromises, and attacks on OT while promoting a zero-trust security approach. The zero-trust conceptualization removes a trusted internal network boundary, causing a significant increase in work for the attacker (AlQadheeb et al., 2022). The EO also emphasizes breaking down barriers for threat information sharing. These information barriers occur with the federal government and information technology (IT) and OT service providers that have contracts limiting the sharing of incident information (e.g., threats and risks). The sharing of cyber threat intelligence (CTI) on internally generated intelligence to counter critical infrastructure attacks was previously kept a secret. In response to the latest cyberattack, the EO allows sharing of CTI and attack methodologies (e.g., ransomware, advanced persistent attacks [APT]) to the energy sector. Sharing meaningful CTI information to organizations on IT platforms is critical for security personnel to prevent future attacks.

In the case of the Colonial Pipeline attack, the management lacked clarity on what network, systems, or devices the ransomware attack had infected or could infect (Ashley et al., 2022). The ransomware impacted the IT network, not the OT network connected to the pipeline equipment. The OT network was never infected, but the company prematurely shut down operations because of the fear that the IT attack would infect the interconnected OT network.

Problem Statement

The DHS in the United States observed that 54% of the utility sector expected a cyberattack on critical infrastructure in 2020. CTI could no longer rely solely on internally generated intelligence (Nweke & Wolthusen, 2020). By 2018, 72% of 585

organizations were producing or consuming sharable CTI to help organizations detect and prevent cyberattacks by sharing threat information (Shin & Lowry, 2020). The general IT problem is that the reports of cyberattacks on the utility sector's critical infrastructure and the critical infrastructure industry (e.g., electricity, gas, water) at risk have not been closely studied by academia. The specific IT problem is that some security operations center (SOC) analysts do not know the relationship between CTI attack identification and critical infrastructure assets to predict cyberattacks in the critical infrastructure industry.

Purpose Statement

This quantitative correlational study examined the relationship between CTI's attack identification and critical infrastructure assets that some SOC analysts use to decide whether to take action in a cybersecurity attack. A SOC analyst uses CTI's threat indicators called indicators of compromise (IoC) to identify potential cyberattacks along with other types of threat intelligence (TI) such as tactics, techniques, and procedures (TTP) to build security information and reports (Schlette et al., 2021c). Overwhelmed is the SOC analyst who receives large amounts of indicators from CTI data providers, which may not be actionable to their industry (Liu et al., 2022). As a result, a quantitative research approach to this data mining problem uses a correlation methodology to understand better the relationship between variables. The strength of this correlation between these two variables, CTI attack identification and critical infrastructure assets, can determine how effective CTI is to a particular industry. The potential contribution to positive social change from this study involves critical infrastructure organizations in the

U.S. energy industry using these findings to evaluate the effectiveness of CTI in predicting cyberattacks and applying mitigation strategies. These mitigating actions may help prevent disruptions to business operations, allowing customers to access critical infrastructure companies' essential resources (e.g., electric, water, gas).

Nature of the Study

To study the nature of this relationship, I chose a quantitative correlational research design applying an ex post facto approach. Through this ex post facto study, I aimed to evaluate CTI cyberattack datasets and target-rich critical infrastructure datasets. I did not choose the other methodologies because the research problem involved finding the relationship between two data points. A descriptive qualitative study does not test specific relationships between variables but describes specific or general behaviors in a phenomenon (Mohajan, 2020). The strength of this relationship between these two variables using statistical methods is to be measured using more than one dependent variable (e.g., firewall, VPN, Internet of Things [IoT] devices) for comparison. Although researchers use an experimental approach to find the cause and effect between two variables and to test a hypothesis empirically (Ross-Hellauer), a nonexperimental ex post facto study can be equally as effective. A study by Bryan (2020) measured the dependent variable, information systems quality, and the independent variable, a population of small businesses on the U.S. East Coast. Bryan sought to determine whether there was a significant relationship between policies and the positive use of small business information systems. A multivariate statistical approach by Bryan was used to determine the strength of the association between variables. The net benefit of this approach is that

the resulting data used more than one dependent variable to measure the strength of the various associations with the independent variables.

This approach was appropriate for this study because the relationship between CTI attack identification and critical infrastructure assets to IT critical infrastructure attacks is not well known. This study used a quantitative methodology to examine the available data feeds (CTI) and the correlation to infrastructure assets (e.g., industrial equipment, systems) relevant to the critical infrastructure industry in the United States. Similarly, other works use contextual correlation, where a similarity-based correlation analysis method for cyber alerts uses a causal link to infrastructure assets (Sen et al., 2022). A researcher can better understand the data analysis by quantifying the strength of causal links (Rohrer, 2024) between the data feeds and infrastructure assets.

Research Question

What is the relationship between CTI attack identification and critical infrastructure assets on IT critical infrastructure attacks?

Hypotheses

Null hypothesis (H0): CTI attack identification and critical infrastructure assets have no significant relationship with IT critical infrastructure attacks.

Alternative hypothesis (H1): CTI attack identification and critical infrastructure assets have a statistically significant relationship with IT critical infrastructure attacks.

Theoretical Framework

In 1942, fighter pilot John Boyd rationalized that the theory of the observe-orient-decide-act (OODA) time loop should be faster than the opponent's OODA loop to win a

conflict or game (Boyd, 1987). The strategic importance of Boyd's OODA situational awareness model is known for the strategic speed loop in the military. However, applications beyond the military (e.g., healthcare, risk management) have found great value in the comprehensive nature of the elements to orient to new information and experience (Ryder & Downs, 2022). The OODA loop offers practitioners a view of new information that allows them to change and adapt, resulting in faster actions to overcome a situation.

An element of the OODA model of orientation that makes this theoretical framework adaptable to practitioners provides resilience to uncertainties. The ability to orient and make adjustments concerning the world is part of the OODA's feedback loop. This feedback loop directs the orientation of individual and organizational learning, according to Ryder and Down (2022). As a result, the model is constantly adapting to the environment. A closer examination of John Boyd's OODA loop, the orient element contains analysis and synthesis of previous experiences, cultural traditions, genetic heritage, and new information (Ryder & Down, 2022). The observation element is to orient the decision and action elements of the OODA model. While the decision and action elements are used as feedback into the observation element, supplying dynamic information to the orientation element.

The categorization between the theoretical framework presented as Boyd's OODA model is situational awareness, and the applicability to cybersecurity is risk management (Jabar & Singh, 2022). In cybersecurity, the capability to have situational awareness of the threat landscape gives the security practitioner a helpful tool to mitigate (i.e., adapt

and overcome) a cyberattack. For instance, an analyst examines the relationship between attack identification and high-value targets that the organization considers high-risk. Using the OODA process, analysts would need to observe the situation, orient themselves to new information, decide if this new information has potential consequences, and act based on all information (Ryder & Down, 2022). The speed of operation affects not only the mitigating outcome, but also the effectiveness.

A variant of the OODA loop is a cyber cognitive situation awareness (CCSA) with cognitive aspects of the decision-making process oriented to perception, comprehension, and projection (Andrade et al., 2022). One focal point of decision-making is the operational activity of the SOC analyst or security incident response team (CSIRT) (Schlette et al., 2021b). These security professionals must effectively analyze large amounts of data and determine possible anomalies that may negatively affect the organization. Situational awareness in the SOC (Sreedevi et al., 2022) is defined by three elements: (a) perception, the information within cyberspace (e.g., firewall, threat news); (b) comprehension, the current situation analysis based on the attack, threat level, and risk; and (c) projection, prediction of types of attacks (Radanliev et al., 2020), threats, and vulnerabilities. For instance, technology solutions such as security information and event management (SIEM) correlate logs and events that facilitate tasks for SOC analysts to assess abnormalities. These correlations are critical because they allow the detection of threats and attacks with purposeful alerts meaningful to the organization (Albasheer et al., 2022). One such relationship between these two variables is the attack identification and infrastructure assets. The analyst must understand the situation's impact (Andrade et al.,

2022) and may need to decide whether to take mitigations to orient and act with active defense (Jabar & Singh, 2022).

Several techniques were developed based on Boyd's OODA loop to address cybersecurity situational awareness. An OODA approach developed by IBM in 2001 called MAPE-K (Andrade et al., 2022) uses a cognitive process of monitoring, analysis, planning, execution, and knowledge. This process uses a technique for feedback controls where self-awareness and situational awareness are designed into knowledge-based computer systems. These cognitive aspects support a cyber cognitive situation awareness (CCSA) model (Jabar & Singh, 2022) where perception, comprehension, and projection form relationships with the identification of relevant data, interpretation of data into knowledge, and prediction of possible impacts. In another model, a mathematical convergent spiral fractal using multidimensional and multi-variability analysis (Tiwari & Tiwari, 2020) measures time impacts in the OODA operating process. Each variant of Boyd's OODA theory in a fractal-OODA environment gives each loop different timings. They found that each phase of the loop of the OODA cycle gives complexity of decisions particularly useful in artificial intelligence (AI) research. This AI technique was used to classify an operational process where cybersecurity applications identify specific attacks.

A study by Jabar and Singh (2022) recommended the OODA framework to minimize the risk to internet-connected mobile devices and systems (e.g., IoT, computers, cyber-physical systems [CPS], 5G edge-cloud systems, autonomous vehicles). They proposed an OODA situational awareness system for enterprise-wide collection using game theory with access control to support decision-making in a monitored environment

providing analysis and reporting. Also noted within the risk model was the National Institute of Standards and Technology (NIST) framework (Kaur et al., 2023) to improve critical infrastructure cybersecurity with five categories (identity, protect, detect, respond, and recovery). NIST, situational awareness, OODA, and game theory frameworks encompass behavior modeling.

Definition of Terms

Critical infrastructure: Critical infrastructure is systems of assets that make up the proper function of society needed for public health, national security, and economic well-being. These systems include vital sectors such as water, energy, communications, and transportation (Makrakis et al., 2021).

Cyber-physical systems (CPS): Smart systems that interact with networks of physical and computational components in the environment. CPS devices (e.g., hardware, software, actuators, sensors) have been used increasingly in critical infrastructure, transportation, healthcare, electric power, and manufacturing (Malatji & Von Solms, 2021).

Cyber threat intelligence (CTI): CTI is actionable threat information relevant to an organization and is designed to be proactive, preventive, and timely (Shin & Lowry, 2021).

Indicators of compromise (IoC): IoC is a threat-sharing format related to CTI with IoC types (e.g., IP, URL, MD5; Ramsdale et al., 2020).

Internet protocol (IP) address: An IP is a unique identifier of a host (e.g., server, device) that can be used to establish behavior patterns and time series of behavior characteristics presenting similarities for network analysis (Huang et al., 2022).

Mitigations: A countermeasure that could prevent adversaries from achieving their tactical object by addressing what can be done about the TTPs. (Georgiadou et al., 2021).

Supervisory control and data acquisition (SCADA): SCADA is a control system that integrates manufacturing, automation, and computing systems with interconnected devices, networking architectures, and services increasingly through the internet. These industrial control systems (ICS) are located in various industrial sectors (Kayan et al., 2022).

Security information and event management (SIEM): SIEM is a system that can query large amounts of data, including aggregation, correlation, alerting, and forensic analysis, for cybersecurity (Rawat et al., 2021).

Security operations center (SOC): SOC bridges CTI and organizational elements that provide security roles, services, and tools (Schlette et al., 2021c).

Structured Threat Information Expression (STIX)/Trusted Automated Exchange of Indicator Information (TAXII): STIX is a format used in sharing platforms (e.g., TI; see Appendix E) supporting automated technologies containing IoC repositories used in CTI (Schlette et al., 2021a). Used along with STIX is TAXII, a transport protocol used by a client to access CTI sources or feeds (Ramsdale et al., 2020).

Tactics, techniques, and procedures (TTPs): TTPs are the adversary's behavioral objectives of performing an attack. Tactics describe the contextual category for actions of why adversaries performed the attack, techniques describe how the action was performed, and procedures are the specific implementation of those techniques (Georgiadou et al., 2021).

Assumptions, Limitations, and Delimitations

Assumptions

The assumptions were that the dataset structuring had the relevant information in the data and that the data in the filtering process were relevant (Miranda-Calle et al., 2021). What researchers in data science want to know is whether the aggregation of cyberattack data provided by a company's technology department is meaningful and measurable. Data scientists use manipulation steps (e.g., natural language processing [NLP]) to remove redundant information during the dataset cleansing and transformation process. The assumption is that the data cleaned up for high-quality analysis will have no missing values of importance disregarded during this filtering process.

Limitations

The limitation of the study was the critical infrastructures' lack of available datasets (Maglaras et al., 2022). Some of the limitations related to critical infrastructure were real-world experiences and innovative dataset solutions. At the same time, other limitations to data were privacy and security on the cloud, edge, and fog computing.

Delimitations

The delimitations included the scope of the critical infrastructure sector relating to utility companies with an emphasis on electric power stations. The three hierarchic levels of critical infrastructure include a system, sector, and element (Rehak et al., 2022). At the system level, the focus is on energy; at the sector level, it is electric power; and at the element, it is the distribution system (e.g., SCADA, controllers). Other delimitations include the type of datasets from CTI feeds, emphasizing the STIX 2.0 format and TAXII sharing protocols (Ramsdale et al., 2020).

Significance of the Study

Contribution to Information Technology Practice

In 2018, the estimated losses of more than $45 billion worldwide were due to 2 million cyberattacks dealing with ransomware incidents (González-Granadillo et al., 2021). These ransomware attacks resulted in a 2-day shutdown of the U.S. natural gas pipeline in February 2020 (González-Granadillo et al., 2021). As analysts prepare for cyberattacks in the utility sector (e.g., electric, water, gas), technical strategies incorporating SIEM tools are used to analyze big data (González-Granadillo et al., 2021). The significance of using SIEM tools with CTI data feeds on a threat intelligence platform (TIP) in cybersecurity is risk management. Applying an OODA model in a SOC operations center gives cybersecurity professionals the tools to provide timely decisions against potential threats by disrupting or mitigating the attack vector. This technology contributes to the IT field by finding meaningful correlations between cyberattacks (e.g., viruses, malware) and the organizations' infrastructure assets (e.g., IoT, CPS; Rantos et

al., 2020). This relationship between IT (e.g., data processing) and OT (e.g., physical device monitoring and controlling) often operates and responds side by side (Kayan et al., 2022).

Implications for Social Change

The implications of this study are to provide meaningful insights for IT leaders regarding protecting IT critical infrastructure assets from cyberattacks. These meaningful insights will provide situational awareness to the critical infrastructure community defending against cyberattacks, such as an early warning system. This preventive approach to attain resilience in critical infrastructure will make it possible to anticipate, adapt, and quickly recover from cyber terrorism, threats, and cybercriminal activities (Malatji et al., 2021). The resulting positive social benefit is that the customers of critical infrastructure utilities experience fewer business disruptions affecting industries (e.g., healthcare, manufacturing; Corallo et al., 2020) and individual customers (Cybersecurity and Infrastructure Security Agency [CISA], 2021).

A Review of the Professional and Academic Literature

In this literature review, I examine the relationship between CTI's attack identification and critical infrastructure assets through the theoretical lens of Boyd's OODA model. The first part of this model (observe and orient) that I researched consists of attack methodologies and identification (cybersecurity roadmap, OODA model, research variables, TTP methods, CTI platforms, formats, indicators). The second part of the OODA model (decide and action) refers to cybersecurity mitigation and risk

(correlation between research variables, critical infrastructure elements, CTI experiments, analysis, risk assessment methods, SOC environment applicability, effectiveness).

Search Strategy

The search strategy was broken down into various themes to conduct this search. I used a spreadsheet to categorize most peer-reviewed articles from the past 5 years related to the research problem, question, and theoretical framework. Four significant categories or themes emerged and were later used as search criteria: cybersecurity, cyber threat intelligence, critical infrastructure, and data analysis. I selected these themes based upon a more extensive keyword search: *Boyd OODA*, *intrusion detection system (IDS)*, *CTI*, *threat intelligence (TI)*, *critical infrastructure*, *SCADA*, *artificial intelligence (AI)*, *machine learning (ML)*, *SOC*, *analysts*, *attack methodologies*, *MITRE ATT&CK*, *risk mitigation*, and *policy*.

I used Google Scholar, the Walden library database, and the peer-reviewed Ulrich database to look for sources. These searches resulted in forming categories of topics when analyzed from the spreadsheet, such as theoretical frameworks, CTI, TI, AI technologies, IDS, cybersecurity, critical infrastructure, datasets, correlations, resilience, TIPs, data analysis, research methodologies, policy, cognitive science, security practitioners, cloud platforms, instruments, and technical vocabulary.

The total number of references was 144, and the total in the literature review section was 106. The peer-reviewed journal articles were 109 or 76% of the total references. From the anticipated graduation date of 2024, the total number of references within 5 years was 121 or 84%. The literature review section contained 86 peer-reviewed

articles, or 81%. Within the 5 years of graduation, the total number of references in the literature review was 88 or 83%. In the sections after the literature review (*Presentation of Findings* to *Conclusion*), the number of peer-reviewed articles within 5 years of graduation was 18.

Of the total number of references, there were 33 non-peer-reviewed articles and other sources. These non-peer-reviewed articles were from conferences that addressed CTI, IDS, cybersecurity, critical infrastructure, correlation, resilience, TI, data analysis, and policy. The other non-peer-reviewed sources included references to public sector media: statistical modeling books, cloud architecture books, research methodologies books, government websites, and university datasets.

Cybersecurity Roadmap

Over the years, cyberattack methodologies and defense technologies used to protect the IT infrastructure have changed. Since 2000 (Andrade et al., 2022; Andrade & Yoo, 2019), cybersecurity professionals (e.g., analysts, IT personnel, and SOC operators) have used cybersecurity devices (e.g., firewalls and IDSs) to detect and defend against cyberattacks by analyzing malicious code (e.g., malware). Cybersecurity professionals analyze this malicious code to detect and identify cyberattacks. In 2005, the defense against APT, zero-day, and critical infrastructure protection (e.g., ISO27001) became more prevalent. This shift to advanced cyberattacks led cybersecurity professionals to correlate cybersecurity events (e.g., the integration of logs from devices, SIEM) in 2010. By 2014, TI emerged from these cybersecurity events, where analysts started correlating cyberattacks with cyber incidents (Andrade et al., 2022). A new generation of TI tools in

conjunction with cybersecurity devices (e.g., firewalls) had been developed to mitigate cyberattacks (e.g., ransomware, social engineering, and cloud attacks) in 2015. A new set of cyberattacks in 2018 caused cybersecurity professionals to take on user privacy attacks, blockchain, IoT, and social media (Admass et al., 2023). During this time, TIPs enabled organizations to collect CTI information effectively against malicious zero-day exploits and APT cyberattacks (Alaeifar et al., 2024). By 2021, ransomware cyberattacks on critical infrastructure facilities disrupted the U.S. gas supply chain, leading to critical infrastructure legislation (Smith, 2021) promoting zero-trust architecture (AlQadheeb et al., 2022).

Observe, Orient, Decide, and Act Model

The dynamics of the attacker and defender in cyberspace is an integral part of the observe, orient, decide, and act process of Boyd's OODA theoretical framework. In the role of a security analyst, the situational awareness of cybersecurity within the organization is a process of detection and response to security incidents (Andrade et al., 2022). The OODA model's effectiveness in cybersecurity is relevant to this research.

The first element of OODA is observe (Andrade et al., 2022; Thulfiqar & Manmeet, 2022). During this phase, the observation of the data collection process starts with monitoring. The second phase is orientation, where the culmination of information builds situation awareness from the data gathered. The third phase, decision, is where the final decision is made based on the previous process and all of the hypotheses. The fourth phase is action, the moment that defines the analytical reasoning process by interacting with the environment (Andrade et al., 2022; Thulfiqar & Manmeet, 2022).

In the OODA model, the detection of sophisticated APT cyberattacks is effective due to comprehensive monitoring of the device's behavior (Thulfiqar & Manmeet, 2022). From the security analyst's perspective, observation is the first phase of the data collection process (Andrade et al., 2022). By applying a cognitive science approach, the OODA model is a process focusing on the tasks and skills of the SOC analyst. Cybersecurity analysts apply these tasks and skills in SOCs using human and technological data collection solutions (Agyepong et al., 2020). These data collection systems that security analysts use are big data, ML, and support decision systems. The workflow process for the security analyst includes three steps. The first step is detection, in which the analyst inspects and associates malicious or suspicious activities with data to detect. This detection of data is then converted into information. The second step is evaluation, during which the analyst needs to take the data and recognize the state of organizational security. Once established, this information is then converted into knowledge. The third step is the evaluation of threats, in which the analyst then correlates the threats to identify the adversary (i.e., attackers) by looking for motives and support. This information is then converted into knowledge and prediction.

The analyst generates knowledge from information based on the evaluation of threats by using different sources (Andrade et al., 2022). These sources include security information from threat databases, reports (vulnerability, security websites), events, user activity, configuration information, vulnerability results, and logs (system and application). The analyst's cognitive association between the tasks and processes is categorized into perception, comprehension, and projection relative to the OODA model.

In the second phase, orientation is the fusion of information where building blocks of situational awareness are formed (Andrade et al., 2022). The security analyst integrates experience and practical knowledge to interpret the observations and then evaluates them to generate a hypothesis about the events. These events can be possible attacks from multiple sources of data. The analyst's skill should establish a correlation between the organization and the digital environment. Andrade and Yoo (2019) proposed a cognitive security model for cognitive tasks and processes for cybersecurity awareness. To generate situation awareness, the cognitive security model considers different sources of information from human intelligence (HUMINT), signals intelligence (SIGINT), open-source intelligence (OSINT), measurement and signature intelligence (MASINT), and geospatial intelligence (GEOINT). A human-generated intelligence is HUMINT, where human sources are gathered through interviews, conversations, or forums. SIGINT is intelligence produced by intercepting signals (i.e., messages) from computer, network, or telecommunications equipment. OSINT is information from open sources such as news, social media, commercial, and technical databases (e.g., WHOIS) (Yadav et al., 2023). MASINT is data from sensing instruments (e.g., sensors) used tactically or strategically. GEOINT consists of geospatial systems derived from satellite or aerial imagery.

The amount of information the security analyst can analyze may be overwhelming. As a result, analysts are aided by expert decision support systems to support the decision-making processes (Andrade & Yoo, 2019; Andrade et al., 2022). The cognitive security model's security indicators support information management using technological solutions. These technology solutions help contribute to the cognitive

functions of the security analyst, where the knowledge domain in intelligence (HUMINT, SIGINT, OSINT, MASINT, GEOINT) is passed to the information domain (observe, orient) and to the cognitive domain (decide). Another model for sharing cybersecurity intelligence relates to these knowledge domains in intelligence. According to Shin and Lowry (2020), the CTI sharing platform uses several taxonomy schemes from data sources such as OSINT, HUMINT, media intelligence, technical intelligence, and intelligence from the Dark Web (i.e., Onion services; Moraliyage et al., 2022), counterintelligence, and threat actor groups. The intelligence gathered from publicly available sources as data feeds can drive active intelligence that helps to manage analysis in the form of signatures, stateful analysis, ontologies (Mkuzangwe & Khan, 2020), cross-layer analysis, and anomaly detection (Kaur et al., 2023).

The third phase, the decision, is the final decision made from the process of analysis based on all of the hypotheses (Andrade et al., 2022). For instance, U.S. Air Force pilots determined that the military established situational awareness based on Boyd's four phases (observe, orient, decide, action). Similarly, cybersecurity professionals (e.g., SOC security specialists, analysts) working on a cybersecurity incident response team (CSIRT) can establish situational awareness by understanding the gravity of the situation and evaluating the likelihood of the impact accordingly. Understanding these environmental trends in the threat landscape can impact the decision-making process affecting the OODA loop in two critical aspects: time constraints and information uncertainty.

Time pressure can negatively impact human cybersecurity (HCS) behavior. A review of HCS behavior (Chowdhury et al., 2020) was conducted based on a theoretical framework of previous psychological constructs (contexts, psychological constructs, nonsecure HCS behavior) found the impact of time pressure and cybersecurity in the context of human effects. The effect of human stress (e.g., cognitive load, pressure, fear) in the decision-making phase (Chowdhury et al., 2020) resulted in a lack of understanding, disclosing sensitive information, disregarding procedures, and bypassing policies, influencing peer behavior, and an overreliance on infrastructure. To meet these challenges, researchers have proposed variants of the OODA loop (Andrade & Yoo, 2019; Andrade et al., 2022): extended OODA loop, iterative OODA loop, CECA model, OODA loop adapted to network-centric warfare, OODA loop adapted to effect-based operation, modular OODA, and cognitive OODA loop.

In the extended OODA loop version, the decide phase is a cognitive domain where a human acts according to the cognitive security model (Andrade et al., 2022). The cognitive security model is a three-phase workflow to establish situational awareness based on prior OODA models. Phase 1 is detection, Phase 2 is situation evaluation, and Phase 3 is evaluating threats (Husák et al., 2022). Compared to the cognitive security model extended, it emphasizes the MAPE-K model of self-awareness and Boyd's OODA model of situational awareness. The focus on the cognitive security model advanced features two workflows. A situational awareness workflow features Boyd's OODA model of process observe, orient, and decide along a mental process of perception, comprehension, and projection. Andrade and Yoo (2019) suggested that evaluating these

parameters of self-awareness and its situation-awareness environment based on security indicators needs further investigation (i.e., future work). Chowdhury et al. (2020) suggested further research regarding HCS behavior and the effectiveness of countermeasures (e.g., operational, technical, human, physical) and the element of the time pressure affecting the cybersecurity domain context.

The fourth phase is action, the moment that defines the analytical reasoning process (Andrade et al., 2022). In the cognitive security model extended, the act phase of the OODA loop is in the cognitive domain, where the human uses actuators to execute a decision process. The OODA situational awareness process's result is the act. From the security analyst's perspective, the association between the cognitive tasks and the cognitive functions within the OODA model in cybersecurity operations shows the mental processes of perception, comprehension, and projection. These mental processes are the operational activities of a SOC in incident response. The security specialist must analyze a large amount of data (e.g., logs, SIEM) and determine how this information may negatively affect the organization.

Similarly, statistical techniques or data mining can determine patterns of attacks or threats. Technology has allowed automated responses and actions, but according to Andrade and Yoo (2019), the system generated many false positives due to misclassifying everyday temporal events. As a result, human-in-the-loop is needed based on experiences for decision-making. Human-in-the-loop decision-making processes allow organizations to set up mechanisms to detect security incidences and define recovery actions to determine security responses in a phased approach. Organizations with incident

response references include CERT, ENISA, SANS, ISO, and NIST (Benyahya et al., 2022). The European Union Agency for Network and Information Security (ENISA) provides guidelines to secure IoT and CPS devices by proposing a structured classification of critical assets to be protected against cyberattacks (Rosado et al., 2022). Another guidance for securing critical infrastructure assets is the National Institute of Standards and Technology (NIST) SP 800-82 Rev. 2, which supports a risk management process to assess, respond to, and monitor critical infrastructure assets (NIST, 2015).

In various cybersecurity scenarios, large volumes of data, and different responses to attack vectors, the security analyst can be overwhelmed when establishing actions with less time. Andrade and Yoo (2019) proposed that cognitive sciences can enhance this process in security operations using a model that integrates technology. Several cognitive security models integrate technological solutions such as big data, ML, and support decision support systems. These information technology systems are increasingly interconnected to devices such as IoT, cloud computers, big data, and cloud computing. In 2017, 8.3 billion connected devices (Andrade & Yoo, 2019) interconnected people, things, and processes worldwide. With management and development applications in smart cities, mobility, transportation, and consumption of natural resources, some models that use technology to monitor cybersecurity include MAPE-K and human-in-the-loop.

MAPE-K uses a cognitive model and preprocesses large amounts of data to determine a level of cybersecurity situation awareness and some automation of the security analyst's tasks. The feedback control system in MAPE-K proposed by IBM in 2001 contains five phases (Andrade et al., 2022): monitoring, analysis, planning,

execution, and knowledge. MAPE-K and the human-in-the-loop approach to situational awareness use some automation to aid the security analyst. These human-assisted machines are recommendation systems and expert decision support systems. While this approach uses technology to assist human decision-making, Tiwari and Tiwari (2020) used a more automated process taking humans out of the loop. Using large datasets to provide situational awareness, Tiwari and Tiwari proposed using AI technologies to automate many of the analyst's tasks in the OODA action phase. During the action phase, this equipment could neutralize the threat, like the new generation firewall in 2014 that immediately acts when a particular threat is detected using TI (Tiwari & Tiwari, 2022). Whether the approach uses some automation to aid the security analyst or a fully automated system, both needs analyzed data to support the decision-making processes. The decision to act phase of the OODA model is tactical, where a go or no-go decision is time-sensitive, while the observe and orient phase is strategic, where timing is at an organizational level. Tiwari and Tiwari concluded that a requirement for faster OODA loops is automation and human-out-of-the-loop in complex environments. However, the quality of timely decision making requires more time. Incorporating big data analysis and AI techniques into cognitive models aids the OODA decision-making loop.

Different phases of the OODA model represent different situations. In a strategic situation, the OODA model represents research topics of real-time situation awareness, knowledge representation, threat impact measurements, risk modeling, and security decision-making (Andrade & Yoo, 2019). As for tactical situational awareness, the research topic measured the impact of threats using automated security aids and

collaborative defense approaches. While at an operational level, the research topics include computer forensics, the detection of viruses, malware, threats, and risk management.

Literature Related to Research Variables

Related to these research variables are data sources used for cybersecurity research to answer research questions related to the variables in the OODA model and the protection of information for critical infrastructure and users. The literature focuses on data sources related to identifying security event fields and what datasets were used in correlation with the existing information security gaps (Kotenko et al., 2022). The security event fields consist of finding relationships between events that belong to the same cyberattack.

Cyberattacks on the electrical power grid can cause damage not only to the electrical system, such as brownouts, but unauthorized access to the system can shut off the lights, communications, and essential services (Dawson et al., 2021). These cybersecurity challenges of threats to CPS, along with attacker motivations, are related to the research variable: the target of critical infrastructure cyberattacks and cyber threat TI detection. Literature gathered focuses on malicious behaviors used on critical infrastructure and the users where two topics emerged: Human behavior in cybersecurity and cybersecurity situation awareness in operations (e.g., security procedures, operational tasks).

According to the cybersecurity roadmap timeline (Andrade & Yoo, 2019), the cybersecurity challenge in 2010 was the correlation between cybersecurity events (i.e.,

attacks) and the user (i.e., security analysts, defenders). These cyberattacks were logged from multiple devices (e.g., firewalls, IDS) and then combined into a SIEM software application (i.e., appliance). The SIEM software generates reports (e.g., auditing, compliance) and can be enhanced to produce alerts based on rules (Albasheer et al., 2022), policies, and the significance of the relationship. Depending on the organization, these data points can be alerts using various variables, from IP addresses to malicious domains. In 2015, the threat landscape included ransomware attacks, social engineering, and cloud attacks (Andrade & Yoo, 2019). By 2018, attacks on user privacy, blockchain, IoT, and social attacks became more complex involving human and cognitive behaviors in cyberspace. The process of detecting these malicious behaviors on an information system and the security interest of protection is an essential focus of this research. One approach was cognitive security through the security analyst's perspective, identifying the relationship between the research variable of attack identification and the assets the user is protecting critical infrastructure.

CPS in critical infrastructures is increasingly being targeted using sophisticated cyber-attacks (Settanni et al., 2021a). Although critical infrastructure networks have strict security measures and multiple security products, these products often fail in APT attacks (Lu et al., 2021). These multistage APT attacks evade traditional detection and management systems (e.g., IDS, SIEM) because of their benign behavior and hidden malicious payloads. Although SIEM products can detect clues, there is a lack of efficient correlation analysis approaches (Sancho et al., 2020). One method for APT detection is to analyze all security incidents and data mine for correlations to identify hidden threats.

This real-time technique to identify threats in the infrastructure network requires efficient data processing, understanding of building correlation rules for an analysis system, and understanding of critical infrastructure networks. The insufficient knowledge of critical infrastructure networks is different from traditional networks, where a security breach can compromise systems' operations and disrupt businesses. The disruption of business operations in critical infrastructure such as gas, electricity, water, financial, and intellectual property can severely impact a nation's security.

The former director of the National Security Agency (NSA), Gen. Keith B. Alexander referred to securing our nation's network as a team sport (Zrahia, 2018) at an RSA conference in 2011. This speech referred to sharing knowledge between attackers to establish a time and technological advantage while many defenders operate in silos where information is not shared. A trend of sharing technology and culture in cybersecurity has emerged to produce real-time actionable details referred to as CTI (Rantos et al., 2020). CTI is a dynamic feed of constant threat information on attack-related objects useful for decision making. Relative to the theoretical framework, Boyd's OODA domain of orient is the fusion of information where situational awareness is built (Andrade & Yoo, 2019) and where CTI provides TI. TI is an emerging defense concept related to real-time actionable cyber threats and focusing on detection and mitigation (Rantos et al., 2020).

Research Variables

Organizations are developing CTI to combat cyberattacks, which cost the global economy over $450 billion annually (Samtani et al., 2022). The CTI capabilities provide organizations with threat information, data formats, and application domains (Schlette et

al., 2021c). One of the data formats used in CTI (e.g., MITRE ATT&CK, STIX) is an indicator of a cyberattack where a malicious actor uses an Internet protocol (IP) address (see Appendix D). An IP address is a unique identifier of a host (e.g., server, device) that can be used to establish behavior patterns and time series of behavior characteristics presenting similarities for network analysis (Huang et al., 2022). These behavior characteristics are a fingerprint of activities from the host at different times contained in the packet contents in the identification process (Miranda-Calle et al., 2021). In the CTI framework, one of the many types of TI research variables range from IoC that use the attributes such as IP addresses, file hashes, and domain names to identify a malicious actor (Serketzis et al., 2019) to TTP that describe the attack attribution (Schlette et al., 2021c).

Since the other variable of the research question focuses on the attack identification of critical infrastructure, researchers can set up experimental observation to see if these malicious actors are attacking the physical infrastructure. Four cybersecurity weights are computed for threats (Jeyaraj et al., 2022): physical, personnel, communication, and data. The physical threat is defined as potential intrusions or insufficient physical protection of infrastructure that can be damaged related to information resources within the organization (e.g., natural disasters, physical assault, software services, malfunctions). The potential intrusions and poor handling, training, or monitoring of personnel with information resources describe the personnel threat (e.g., password of users, false identities, extorting others). Describing the communication and

data threats are attacks on data, the research model of compromising confidentiality, integrity, and availability (e.g., computer viruses, malicious software, worms).

The control variables Jeyaraj et al. (2022) uses are locations of the headquarters (e.g., east, west coast), business-to-consumer organizations (e.g., cybersecurity infrastructure), and information-intensive organizations (i.e., IT, finance, healthcare sectors). In Jeyaraj's research, a fourth variable was introduced to distinguish between organizations that had previous security breaches as opposed to others that did not. Jeyaraj's experiment used text mining and statistical analysis for the research method by collecting data from 10-K reports (e.g., Edgar database), analyzing data using the statistical tool R (e.g., EdgarWebR package), and using NLP algorithms for organizing the data into topics from predefined keywords (e.g., topic modeling). The data analysis measured the dependent variable (e.g., cybersecurity response, categories) and the four independent variables (e.g., physical personnel, communication, and data, operational). Using an XT regression model, Jeyaraj's results showed variance in the cybersecurity response of information-intensive organizations represented in time-invariant and time-varying categories. The resulting experiment showed that the cybersecurity response was higher for information-intensive organizations in the United States coastal region from 2015 than other organizations. The overall variance in the cybersecurity response was 32.7%. This research methodology and use of variables are vital to the research question and paralleling my study's independent (e.g., attack identification, threats) and dependent variables (e.g., physical infrastructure, critical infrastructure). My research problem

variables include CTI attack identification, critical infrastructure assets, and IT critical infrastructure attacks.

Research Variable Cyber Threat Intelligence

In the first variable of CTI attack identification, the research question asks whether there is a significant relationship between the organization's infrastructure assets and the cyberattacks. The family of critical infrastructure threats (e.g., damage to IT assets, loss of information, malfunctions) focuses on deliberate physical attacks (Rosado et al., 2022). Depending on the stakeholders at each level of the organization (e.g., executives, security managers, incident response teams, operators, analysts), the CTI attack identification has different meanings. By mapping out the research variables to who is consuming the CTI information attack, identification of research variable subtypes has been related to the organizational structure (Tuma et al., 2021; Zrahia, 2018): Consumed by the organization's board for strategic information (high-level), operational information about potential attacks on the organization for security manager, TTP about threat actors conducting attacks for the incident response teams, and the technical information feeds for operators in enforcement and analysts in analysis systems.

Subtype four of CTI focuses on technical information feeds consumed by the operators in enforcement and analysts (Zrahia, 2018; Borges Amaro et al., 2022). The technical information structure from CTI provides data objects, including IoCs such as IP, email, and DNS addresses. When these CTI objects (Zrahia, 2018; Borges Amaro et al., 2022) correlate with the organization's network traffic, the SOC personnel can decide whether these alerts should be actionable. These actions include a decision by SOC

personnel on whether to mitigate a cyberattack based on TI analysis by humans. Human cyber experts performed TI analysis manually or automatically, providing CTI data. This CTI data, when used together with products such as firewalls as an enforcement point, directly affects a mitigating action such as blocking of an IP address. Additionally, these CTI data sources can be shared within a community on a platform where CTI objects can be acted upon immediately.

This CTI platform focuses on the capabilities, motivations, and the adversary's intentions (Ramsdale et al., 2020), where the analyst gathers information about the adversary through observation and analysis. When shared in a community, the observation and TI analysis results provide meaningful intelligence by alerting them of an impending cyberattack. These security alerts, incident reports, and CTI objects includes threat indicators or observables associated with the attack. Understanding these attack variables or indicators can better protect the network, knowing the relationship between attackers, network defenses, and automated mechanisms. These mechanisms include a multilayered defensive system correlating cyber-threat detection and dictionary-based or word-based approaches in the collaborative analysis engine for situational awareness called the European control system security incident analysis network (ECOSSIAN; Wallis & Leszczyna, 2022). This mechanism used a correlation of word-based approach called Vector space models (VSM) to derive actionable intelligence from similarities processed from security-related documents. Documents imported are security-related documents from various sources in various formats (e.g., STIX, IODEF, JSON). This

heterogeneous data needed to be standardized by analysts into a format allowing for community interoperability.

The standardized format of CTI in the literature (Ramsdale et al., 2020) is called STIX/TAXII, and the cyber observable expression (CybOX). The most popular standard for describing TI is STIX (see Appendix E). An exploratory study of software vendors from a literature review by Ramsdale identified 22 TI sharing platforms. This broad literature review compared protocols and methods used for sharing CTI. The analyzed results of this sharing model are to find the strongest relationship with the output variable (Suryotrisongko et al., 2022). The input variables were the types of CTI from correlated sources (i.e., blogs, forums, vendors) in structured (or unstructured) formats. This survey focused on the types of intelligence and the quality and quantity of CTI sharing (Ramsdale et al., 2020). The results were quality verses the quantity of CTI's IoC data applicability at the organization's strategic, operational, and tactical levels (Zibak et al., 2022).

Independent Variables

An independent variable can be manipulated to explore its effects without being influenced by other variables (Kong et al., 2023). The researcher can vary the independent variable, such as CTI data types, and explore the effects on the critical infrastructure assets. These data types are related to the research question and these cyberattacks' relationship with critical infrastructure assets. Researchers can evaluate the significance of the independent variable by analyzing various variables (Ramsdale et al., 2020) such as comparing cyber intelligence sources, formats, and languages.

What was investigated are vendors with CTI data feeds in several format types (e.g., single-column text feeds, multicolumn, rich CSV feeds, STIX, RSS) (Ramsdale et al., 2020). The range of CTI types analyzed by Ramsdale includes the most common type is IP addresses, followed by the description of the threat type (i.e., malware type) and the URLs. The common CTI feeds with one or two CTI types are sparse, while data feeds with more than two CTI types are rich and use complex data formats such as structured threat information expression (STIX). The CTI types represented include the sparse and rich types: Details link, Threat type, Labels, IP, Port, URL, Domain, Email, Win RegKey, CVE, Filename, Hash, Status, Count/score, RDNS, DN registrar, ASN, Country, Cert CN.

The IP data type is the most common of the CTI types represented (Ramsdale et al., 2020). An IP address is a numeric address that identifies a computer on the Internet and within network transmission packets contain source IP and destination IP information (Kotecha et al., 2021). The name associated with IP is the domain name system (DNS) that maps hostnames in the Internet infrastructure and domain-based applications (Ren et al., 2020). This IP address is an essential variable found in the DNS hostname of command and control (i.e., C2) nodes associated with APT, malware, botnets, and ransomware.

According to Ramsdale et al. (2020), the STIX 2 data format provides a rich and sparse representation of the CTI types (see Appendix D). Various providers use a common source retransmitted by STIX 1 and 2 formats (see Appendix E). Microsoft SIEM solution (Sentinel) (Microsoft, n.d.) can import threat indicators (i.e., CTI data)

from STIX and Trusted Automated Exchange of Intelligence Information (TAXII) servers (see Appendix H). One of the CTI data sources used by Microsoft is the Limo dataset by Anomali, where these CTI data feeds provide threat indicators to detect, provide context, and inform about potential cyber threats. Ramsdale compared different formats, such as STIX 1.1 and STIX 2, using CTI source, PickUpSTIX, and Limo. The STIX sources from the literature examined samples of over 275 data feeds. Identified from the first 125 data feeds were based on the STIX standard. They analyzed from the 125 feeds 22 CTI sources (i.e., providers) and the formats used (text, CSV/RSS, JSON/XML, STIX1.x, STIX2.x, MISP, IDS, DNS). The CTI source, Limo used the STIX2.x format supporting as many as 11 feed types. Limo endorsed a mix of sparse and rich CTI types. A mix of CTI types from various data sources are represented in the CTI data feed (see Appendix A).

Correlation Approach Between Research Variables

To uncover the correlation between the data and timely attacks, researchers have introduced various methods as countermeasures. Introduced is one method that used CTI and security information to uncover attacks using a word-based methodology (Schlette et al., 2021a). Researchers mapped these attacks in the CTI feed from known data feeds using word-based methodologies to make these correlations. These correlations are essential to the research question related to the relationship between the research variables CTI and critical infrastructure cyber-physical systems. The metrics used to assess the accuracy of the correlation methods implemented a cyber TI analysis engine called Collaborative Automated Course of Action (CACAO; Schlette et al., 2021b).

When applied to SOCs, this CTI analysis engine can help analysts support incident handling for cyber situational awareness. Incident handling CACAO provides analytical support to SOCs, correlating the attack to an incident (Schlette et al., 2021b). A correlation VSM technique draws similarities between security-relevant documents through CTI security information relating to CPS in critical infrastructures. This approach analyzed large amounts of data from security incidents and cyber-physical systems CPS operating modern critical infrastructures. Once the data was analyzed, the analysts extracted meaningful information about what was compromised.

The CACAO application was used to assess the performance of these correlation methods by importing security data from diverse input formats such as STIX, IODEF, and JSON (Schlette et al., 2021a). The analyst then selects a document and extracts artifacts, or words linked to the knowledge base. The method used to link documents for analysis was word-based. This method could identify 85% of documents while 20% were identified as different documents. The resulting F-measure for the word-based method was around 0.81, providing high precision.

The architecture presented by Schlette et al. (2021a) was used in a national security operations center (N-SOC) to acquire threat data, aggregate information, and share information while protecting privacy. This practical CTI analysis and correlation using this N-SOC architecture provided efficient threat data acquisition, visualization, and evaluation. This architecture used the CACAO applications' process diagram that consisted of a data source importer from incident reports (IODEF), threat information (STIX), and OSINT (JSON/text). The following process features extraction of these data

sources submitted to the document-feature mapping component that inputs that information into the knowledge database. The knowledge database has document-linking elements that help the N-SOC analyst visualize, provide feedback, and handle incident information quickly. The resulting OSINT feeds using this CACAO system have proven high-quality correlation when applied across large, diverse datasets using VSM. This VSM technique processed the CTI data to identify significant similarities to security-relevant documents (Schlette et al., 2021a). These techniques are relevant in applying correlation approaches between research variables.

In relevant work, similar matching methods and correlation techniques helped SOC analysts identify meaningful similarities between research variables on attacks and critical infrastructure systems. Classification systems helped incident response architected in network zones and on asset types using a prioritization scheme for alerts (Makrakis et al., 2021). These alert schemes used Bayesian techniques and learning-based anomaly detection in their algorithms. Detection used SNORT signatures specialized in detecting critical infrastructure intrusion protocols (e.g., Modbus, DNP3, PCS, ICCP) and sensors (e.g., RTU, ICS, SCADA) that send security events to SIEM server (Rakas et al., 2020). These security events allow the ICS security operator to detect traversal attacks across network zones using IP address matching. A sequential correlation algorithm, Detection and Analysis of Threats to the Energy Sector (DATES), scores each connection towards the high-value target. However, no standard correlation analysis technology meets the realistic environment solving real-time analysis.

A correlation algorithm called Greedy Correlation Analysis System (G-CAS) addresses this real-time analysis problem (Lu et al., 2021). Using a correlation analysis approach, Lu based the G-CAS formula on previous studies in three categories: similarity-based sequence-based, and case-based methods. A similarity-based correlation method based on the alerts' similarity and merging these alerts to reduce the number of alerts resulted in greater efficiency. The computer community commonly uses a Greedy algorithm to optimize tasks by dividing a significant problem into multiple minor problems where the subproblem can be solved. Lu combined a Red-Black-Tree (i.e., decision tree) with a greedy-tree algorithm, utilizing a meta-matching technique to analyze these correlation methods related to dependent variables (e.g., critical infrastructure assets) by calculating the best performance. Of all these methods (e.g., word-based, similarity-based) and techniques (e.g., VSM, Bayesian) used for correlation, the researchers use algorithms (e.g., CACAO, DATES, G-CAS) to produce meaningful results. It is easy to see from these studies that the researchers developed algorithms concerning the dependent variable (e.g., critical infrastructure) using various correlation methods and techniques.

Dependent Variable Critical Infrastructure Elements

In previous discussions about correlations and independent variables, researchers used various independent variables to experiment on the dependent variable. The independent variable (CTI) can be changed or varied, while the dependent variable (critical infrastructure) remains unchanged so that researchers can see the impact on the dependent variable. These critical infrastructure variables include a family of assets such

as devices (e.g., actuators, sensors), an ecosystem of devices (e.g., IoT, embedded systems), communications (e.g., network, protocols), infrastructure (e.g., gateways, power supply), platforms (e.g., cloud, web-based services), decision making (e.g., algorithms, data mining), applications and services (e.g., analytics, visualization), and information data (e.g., databases, network services) (Rosado et al., 2022). These CPS dependent variables are within the critical infrastructure system, with elements requiring resilience and service reliability.

According to DHS *resilience* is described as quality that reduces the vulnerability of a system and "the ability to absorb, adapt to, and rapidly recover from a potentially disruptive event" (Rehak et al., 2022, p.1). When these disruption events occur, the affected organization assesses the security and documents the incident. Several approaches to resilience researchers used to assess infrastructure resilience include indexes, frameworks, and methodologies.

A Resilience Measurement Index by Petit et al. (2013) from Argonne National Laboratory in the US provided resilience indicators by measuring critical infrastructure components. Researchers classified the components into four primary groups for disruptive events (Rehak et al., 2019): preparedness of subsystems, mitigation measures, response capabilities, and recovery mechanisms. The impact of this critical infrastructure resilience cycle model is the ability of these subsystems to mitigate the damage caused by disruption.

A critical infrastructure resilience study by Sathurshan et al. (2022) evaluated the literature on frameworks' impact on resilient infrastructure. A survey of frameworks were

evaluated post-disaster (i.e., ex-post) and categorized based on critical infrastructure resilience: social serviceability (e.g., response strategies, restoration), functional measure (e.g., recovery, damage assessment), and indicators (e.g., seismic repair rate, downtime). The impact of this framework is the methodology used to review scholarly literature: (1) identification, (2) screening, (3) eligibility, and (4) inclusion. A flowchart was used for studies that identified resilience performance indicators, meta-analysis, and damage assessment to critical infrastructure.

These impacts on critical infrastructure had dependent links to critical infrastructure sectors. These sectors were categorized into three hierarchic levels of critical infrastructure (Rehak et al., 2016): system, sector, and element level. Rehak described the top level; the *system* categories the critical infrastructure's functions (e.g., energy, transport, water, food processing, agriculture, industry, communications, and information systems). Middle of the chain, at the *sector* level, describes the individual sectors that make up the critical infrastructure (e.g., electric power, water supply). Next is the *element* level at the bottom of the chain; the sector-related infrastructure comprises elements or subsystems (e.g., distribution system, SCADA). These three critical infrastructure levels are part of a technical and organizational resilience assessment called the Critical Infrastructure Elements Resilience Assessment (CIERA), funded by the Ministry of the Interior of the Czech Republic (Rehak et al., 2016).

This case study describes the electric power sector containing the control room of the distribution system operator using the CIERA method (Rehak et al., 2019). The CIERA method comprises three cyberattack assessments: robustness, recoverability, and

adaptability. The first step was to assess the current robustness of individual variables (i.e., preparedness, redundancy, detection capability, responsiveness, and physical resistance). The second step is to assess the recoverability, consisting of individual variables of material, financial, human, and recovery processes. The third step is adaptability, where the last resilience component consists of individual variables of risk management, innovation, education, and development processes. The resulting assessment was the calculation of the element's resilience as the arithmetic means of the components. For example, when the final total of variables was calculated to be 69% or more, it is considered an acceptable level of resilience, and below 69% are considered weak points (e.g., crisis preparedness at 60%, innovation processes at 60%, and risk management at 34%). These low levels of resilience scores are areas for incident reporting and improvements to organizational resilience. These resilience evaluations are essential to understanding what variables or elements are meaningful to the organization's critical infrastructure (Rehak et al., 2022).

Related to the research question, meaningful to the organization's critical infrastructure assets are the element's resilience to cyberattacks. These sector-related infrastructure elements, including the subsystems (e.g., controllers, sensors), are evaluated for risk. Kure and Islam (2019) proposed a formula to determine the level of risk to assets. A malicious actor may threaten these assets, such as people, data, services, and facilities. It is essential to understand the impact on assets in critical infrastructure instead of other nonessential infrastructures because of the consequences associated with the threat. The consequence of an attack on critical infrastructure assets that support a

nation's utility (e.g., electric, gas, water) is higher than an asset that supports a local noncritical system.

Application to the Applied IT Problem

This quantitative research approach aims to understand better the strength of a correlational relationship between variables (e.g., CTI attack identification and critical infrastructure industry assets) effectively applied to the IT problem. The specific IT problem is that some SOC analysts do not know the relationship between CTI attack identification and critical infrastructure assets to predict cyber-attacks in the critical infrastructure industry. According to Agyepong et al. (2020), one reason is the impact of time constraints and information uncertainty on SOC analysts. A lack of time can negatively impact HCS behavior (Chowdhury et al., 2020), such as avoiding awareness. This negative impact on HCS behavior during the decision-making process in the OODA cycle can result in delayed actions by the SOC analyst and security operators. The security operators' choices as defenders must decide if the attack vector is effective enough to circumvent the defenses (Maccarone & Cole, 2022). The attacker must decide which attack vector to take accordingly, such as the default configuration of an industrial controller (e.g., PLC, IoT) that supports a critical infrastructure (e.g., nuclear power plant). The dynamics of this relationship between the attacker and defender are similar to game theory (e.g., Bayesian game).

The application of the OODA theoretical framework provides a methodology where the CTI attack identification and critical infrastructure assets can be significantly correlated (Aldhaheri et al., 2023). According to Settanni et al. (2021a), SOC analysts

who know the correlation between cyber-attacks and CPS in critical infrastructures are better prepared. Relative to the OODA cycle, a faster decision-making process leads to fast mitigating action (Kotsias et al., 2023). Overall, when the security personnel are given actionable, timely TI (Andrade et al., 2022) through CTI feeds and correlated with the organization's critical infrastructure asset elements (Aldhaheri et al., 2023), the resulting alerts can be enough to find immediate mitigating actions preventing an attack (Rehak et al., 2022). Furthermore, this data collection of attack correlations can lead to predictions, allowing SOC operators to understand attack methodologies better, as enumerated in MITRE's attack identification method.

MITRE ATT&CK Method

According to Mitre (Andrade et al., 2022), since the 1990s, cybersecurity events have changed how the SOC operators deal with operations as the growth of APT, consolidation of IT resources into cloud-based computing, the proliferation of mobile technologies, and the transition of attacks to the network to attacks from the client-side. To meet these changes in cyber-defense, situational awareness in cyberspace involves the understanding of investigating concepts, technologies, and analysis of attacks. These concepts are categorized in a knowledge base of data into a framework called MITRE ATT&CK (Georgiadou et al., 2021). This framework comprises a comprehensive matrix of cyber-attack procedures using TTP principles. In the Mitre's attack framework, security professionals such as threat hunters, red teamers, and blue team defenders use these tactics and techniques as a standard classification so that organizations can quickly identify the risks applicable in CTI analysis using mapped data.

According to Georgiadou et al. (2021), Mitre's attack framework addresses a knowledge gap in the literature review and research analysis using a holistic approach to cybersecurity from the adversary tactics and threat perspective. What makes Mitre uniquely suited to CTI is the assessment of the malicious techniques within the MITRE ATT&CK matrices. These matrices explain the attack chain by describing the events of the adversary attacks. The adversary attacks, Mitre ID, were identified as providing a referenceable knowledge base suited for CTI. Exploiting the MITRE ATT&CK matrices, the scientific community has also been using this framework for accurately modeling cyber-attacks, detecting, re-generating, and simulating cyber-attack chains (Georgiadou et al., 2021). The MITRE ATT&CK matrices identify multidimensional security gaps such as human factors, insider threats, policies, procedures, and infrastructure of an enterprise (Georgiadou et al., 2021) by analyzing attacks/tactics on the organization and human behavior. These human behavior factors leading to attacks are suited to the critical infrastructure where the security operator (i.e., analyst) can identify neglected security mitigations and understand the possible risks (e.g., insider threats). In a survey by McAfee and the University of California in September 2020 (Georgiadou et al., 2021), 80% of enterprises used MITRE ATT&CK for threat protection, 57% for detecting gaps in security policy, 55% for security policy implementation, 54% for threat modeling. According to the survey, the applicability of MITRE ATT&CK primarily relates to threat protection. Secondly, this knowledge gap relates to the IT problem of understanding the correlations between cyberattacks and ICS controls in critical infrastructure (see Appendix D).

ATT&CK for industrial control systems (ICS) is a knowledge base to cover the adversarial techniques, sub-techniques, and specific procedures organized into a model suited for critical infrastructure by Mitre (Georgiadou et al., 2021). Mitre ICS in critical infrastructure elements from various sectors (e.g., energy, gas, water) has a vital feature in the MITRE ATT&CK framework that categorizes post-compromised ICS network attacks. In January 2020, the publicly released ICS attack had 11 tactics, 81 techniques, and 50 mitigations (Georgiadou et al., 2021).

Applicable to the IT problem are these enumerations of ATT&CK for ICS, where the correlation of ICS elements to the CTI knowledge base is mapped so that security professionals can analyze ICS exploited networks' vulnerabilities (Chen, Vaitheeshwari et al., 2024). For instance, the category of MITRE ATT&CK with Mitre ID: M0804 (Human User Authentication) describes an attack on a VPN network that allows access to the ICS critical infrastructure assets.

MITRE ATT&CK and Critical Infrastructure Assets

A supplier of industrial cybersecurity software in 2020 surveyed 1000 IT professionals from security positions around the globe (United States, United Kingdom, Germany, France, and Australia) and found that respondents were more concerned 65% about attacks on critical infrastructure as compared to 35% enterprise data breaches (Georgiadou et al., 2021). More than half of the survey's participants believed that the industrial networks were not adequately protected from cyberattacks. Of those IT professionals surveyed, the trends of IT and OT networks show a co-existence. Strong correlations exist between electrical power and energy systems sector (EPES)

organizations using the *hybrid MITRE ATT&CK for Enterprise and ICS* approach to detect external threats in the MITRE ATT&CK matrix.

Organizations have utilized the MITRE ATT&CK matrix and developed a cybersecurity culture framework to co-exist in the IT and OT dimensions of critical infrastructure and domain elements (Georgiadou et al., 2021). This cybersecurity culture framework was designed by the EnergyShield project and co-funded under the European Union's Horizon 2020 research and innovation program. The framework consists of security culture indicators on the organizational and individual levels. In the first level, dimensions address security aspects of the organization (Assets, continuity, access and trust, operations, boundary defense, and security governance). Within the organizational category is *Assets*. These assets include application software security, data security and privacy, hardware configuration management, information resources management, network configuration management, network infrastructure management, software assets management, personnel security, and physical safety and security. The second or individual levels are analyzed into four dimensions: Attitude, awareness, behavior, and competency. Dimensions can be further analyzed to address security policies including hardware assets management, hardware configuration management, network infrastructure management, and information resources management.

One case using the cybersecurity culture framework during the COVID-19 pandemic assessed the cybersecurity readiness of critical infrastructures (Georgiadou et al., 2021). The survey found the framework's methodology complex because of the relationship between the organizational and individual indicators. When analyzed, the

application scenario showed that the technical nature of critical infrastructure assets (e.g., software security, data, hardware) affected the overall security readiness relative to dimensions (e.g., attitude, awareness, behavior, competency) in the security policy of domain assets (e.g., buildings, machines, systems, information, people).

In an energy sector case, a scenario of the cybersecurity culture framework was applied in the primary electricity substations of a transmission systems operator (TSO) company (Georgiadou et al., 2021). This scenario was assessed for "Physical Safety and Security" and "Personnel Security" for the operational security of the power plant. The cybersecurity culture framework aims to assess the specific domains of the "Assets" dimension. A discovery related to the "Physical Safety and Security" domain, the framework suggested re-examining mitigations "M0805—Mechanical Protection Layers" and "M0812—Safety Instrumented Systems" and alerting for cyber-threats such as "T0979—Damage to Property" and "T0880—Loss of Safety." In this simple example, the MITRE ATT&CK knowledge matrix and the mitigation table led the cybersecurity operator "X" to act. The operator took mitigating action and improved the power plant's operational security and safety.

The sector’s organizational assets (Safety Instrumented Systems, ICS) aligned well with the cybersecurity culture framework and the specific IT problem. The specific IT problem states that some SOC analysts need to gain knowledge about the relationship between CTI attack identification and critical infrastructure assets. In this application scenario by Georgiadou et al. (2021), the problem was to protect the power plant. At the first level, organizational dimensions address security aspects of the assets. The second

level was individual, where operator "X" competency was demonstrated using MITRE ATT&CK with knowledge of TSO domains. The operator utilized MITRE ATT&CK for Enterprise and ICS knowledge base to fill in the knowledge gap that led to mitigating action. These mitigating actions can be used to predict cyber-attacks in the critical infrastructure industry by identifying security deficiencies using a threat intelligence platform.

Threat Intelligence Platform

The data gathered early in Boyd's OODA framework is complex, using various intelligence methodologies (e.g., HUMINT, SIGINT, OSINT, MASINT, AND GEOINT). Some organizations may not want to share intelligence problems outside their organization. The complexity of gathering CTI data to be shared with the community relates to security threats linked to actual organizations (Shin & Lowry, 2020). The idea of sharing CTI information with the community is an issue with the organization, extra-organization, and nation/law (Dykstra et al., 2023). Related to sharing CTI data is the taxonomy and the effort required for empirical data to be conceptualized, operationalized, and instrumented. In the conceptualized space, stakeholders measure the data on a multilevel and longitudinal level relative to the actual organizational conditions (Shin & Lowry, 2020). These different levels of sharing TI can occur at a national, international, or local level where stakeholders have a trusting relationship with partners (e.g., Information Sharing and Analysis Center [ISAC]; Dykstra et al., 2023). The effect (longitudinal) in organizations is to create situational awareness so that stakeholders can

take the actionable intelligence directly from CTI data and alert about the threat to take mitigating action as soon as possible (Sun et al., 2023).

Compared to traditional forms of manual dissemination of information (e.g., emails, phone calls, meetings), the CTI sharing platform features automation, collaboration, indicators, industry sector sharing, and some human elements of cultural and language barriers (Sun et al., 2023). In a Poneman Institute survey (2014), 39% of participants replied that an impediment to full CTI exchange participation is that slow and manual sharing processes, such as copying and pasting spreadsheets, were time-consuming (Dykstra et al., 2023). The information became obsolete when the analyst had to prepare the report for sharing, analyze the data, inspect for context relevance, and format for stakeholders. A more effective method was needed to reduce miscommunication and improve the efficiency of sharing. An automated data analysis system developed by the U.S. Government and the Mitre company using the protocols called TAXII provided a more efficient mechanism for cybersecurity needs using a standardized approach (Sun et al., 2023). This structured approach to cyber threats includes analysis, specifying indicator patterns, managing responses, and sharing CTI information. According to a SANS survey in 2020 (Dykstra et al., 2023), a CTI industry surveyed respondents and found that 82% believed CTI provides value, but only 4.2% believed in its effectiveness.

Organizations must collect data effectively, share real-time CTI, and use actionable TI to prevent attacks. By using a threat intelligence platform (TIP), organization can dynamically response to cyber threats (Zibak et al., 2021) and automate

defense in response. CTI information can be correlated and visualized when connecting to CTI repositories. In the process of using CTI information to mitigate a threat, such as malware downloaders targeting banks within the URL (e.g., https://funds.com/images/Nu48jdx), the TIP system uses a script to mitigate downloads from the malware site found in the URL (https://funds.com). The timely mitigating action prevents future attacks before malware from the malicious site can be downloaded.

Cyber Threat Intelligence Indicators

An *indicator* indicates something new in a changing situation (Meland et al., 2022). These indicators in CTI, called IoC, identify threats such as malicious IP addresses, hashes, descriptions of threat actors, campaigns, and motivations. Security professionals used CTI indicators in TI applications and TIMP systems to detect threats, such as using IDSs to detect malicious IP addresses. The actionable intelligence from indicators leads to identifying compromised hosts and facilitates blocking websites, as well as downloading malware (Aldhaheri et al., 2023). At the same time, when IDS detects threats, analysts create CTI libraries of indicators. The quality of indicators is good at identifying malicious actors at the IT enterprise level (Meland et al., 2022). However, it neglects the device level, the Internet of things (IoT), and industrial Internet of things (IIoT).

CTI libraries may help link historic enterprise IT indicators to new critical infrastructure IoT and IIoT indicators (Radanliev et al., 2020). These embedded devices may help the stakeholder's area of responsibility of CTI applied to relevant work, such as software, hardware, and networks. This data relevance is essential to data quality, where

stakeholders must filter high-level CTI data by selecting what is meaningful. Stakeholders receive meaningful data when filtering TI data through search functions in CTI or TIP platforms (Chen et al., 2024). Further correlating this threat data with their organization's system of inventory assets (e.g., IoT, IIoT), the stakeholder can map business processes to geographical, political, and industry sectors (Meland et al., 2022). Nonmeaningful threats or irrelevant information are discarded like irrelevant emails (SPAM) are not shared, and only genuine emails are shared. In this scenario, CTI data sources are relevant to the organization's objectives, while other data sources are appropriate.

Indicators in the industry sector share similar CTI attributes because similar TTP may have similar IoC (Rantos et al., 2020). Sharing similar IoC has a cost-benefit because of the investment in prevention. Knowing a cyberattack before it happens by indicator has the advantage of timeliness, where course of action information can prevent or mitigate an event. In one scenario, when stakeholders shared CTI information, it was not immediate due to an incomplete course of action. In the second scenario, when someone in the organization clicked on a malicious download, the CTI information that was shared immediately could mitigate the risk, thereby blocking the email link. Information sharing is a policy considered by the critical infrastructure industry as a success factor (Atkins et al., 2021), with the dependent variables of firms' capabilities, collaboration with the government, and collaboration with other firms.

In contrast to CTI structured data sources, an unstructured social media data source is a good indicator of the changing cyber landscape. A CyberTwitter framework

presented by Fang et al. (2020) used OSINT collected from Twitter (i.e., X) feeds to generate alerts. These alerts are gathered from a thematic analysis of cybersecurity keywords (Kaur et al., 2023). A TI tool SYNAPSE filters Twitter streams and categorize these IoCs using an AI classification technique (Browne et al., 2024). A quantitative study of Tweets by Alves et al. (2021) contained more than 195,000 from 80 accounts and showed that most tweets over eight months related to IT infrastructure security (above 90% true positive). Out of 9,093 vulnerabilities analyzed, 89 were referred by tweets before being published. These early alerts on Twitter are good indicators of future vulnerabilities. The alerts posted by 53 different accounts used in 89 alerts showed a diverse source of accounts that complemented each other. In this case, the alerts were about a cybersecurity discussion about an exploitation (CVE-2017-0144) called "wannacry" (Fang et al., 2020).

Relevant to the research question are meaningful CTI data to the organization. In this study by Wagner et al. (2019), organizations that handled large volumes of TI data, such as shared threat alerts (i.e., CTI indicators), found that relevant alerts reduced SOC personnel workload. In scenario A, the stakeholder receives 15,000 alerts per week from a threat monitoring system. Of those 15,000 alerts, only 400 alerts are relevant. Since the organization has limited resources, a shortage of employees can investigate only 60 relevant alerts. The stakeholders receive an additional 10,000 high-risk threat alerts but need to know if these are relevant to the system. In scenario B, the number of weekly alerts is the same as in scenario A, with 15,000 alerts weekly. The difference is that

scenario B used a tagging system to render CTI content relevant to the stakeholders. This approach effectively saves time, resulting in a practical system.

CTI Experiment and Analysis

Using approaches that use CTI, sharing platforms located in various geographic areas, analyzed are the results of an experiment about attack behaviors. This attack behavior can be best analyzed by correlating the overall number of attacks with the number of IP addresses from CTI's IoC. In one experiment by Thom et al. (2021), the set-up was six multiservice honeypots to collect traffic in locations worldwide over months. Malicious activity such as attacks on VPN was monitored and recorded. Analyzed are the tactics, and techniques, facing services such as SSH, Telnet, HTTP, and FTP, capturing characteristics of the source and destination IP addresses and port numbers. The collected data included usernames, passwords, executed commands, and files between the honeypot sites, correlated and analyzed attack and traffic patterns (Aldhaheri et al., 2023). The log data was analyzed for attack methodologies from various geographic locations, and various approaches were developed to identify malicious behavior to construct cyber-TI feeds. Explained is an experiment by Thom et al. (2021) and applicable to the theoretical.

The log data of the Cowrie dataset (Thom et al., 2021) was reconstructed from the attacker's sessions. By gathering the attacker's intent and methodology, malicious behaviors were identified and used as CTI feeds. The CTI feeds from various sources were then aggregated by TI vendors. The logging activity in this experiment demonstrated the deployment of multiple data collection points using lightweight

multiservice honeypot virtual machine servers (i.e., Raspberry Pi). During the experiment, services from applications were purposefully left open, exposing the service port to attackers. Within minutes, attackers performed brute-force attacks using dictionaries of usernames and passwords to gain access. Data from six Digital Ocean data centers in London, New York, Toronto, Amsterdam, Bangalore, and Singapore gathered attack behaviors related to higher education domains. During the study, these attacks attracted specific types of traffic between March 2020 and December 2020. The data amalgamated correlated to the overall number of attacks per unique IP address gathered from Apache, FTP, and Linux kernel logs captured in (.json, .log) files. These files contained tags, event-ID, session-ID, source-IP, destination-IP, username, password, source-ports, destination-port, messages, file hashes, and other data points. After several months, the attacker's insights, objectives, and methods (TTP) were observed.

Analyzed from the Cowrie data, insights generated from event ID defined a command executed by the attacker. The experiment's most common types of activity (Thom et al., 2021) were login attempts and requests to forward traffic to another device. Observed in stage 1 of these attacks is where the initiating machine contact is randomly scanning the network. In stage 2, the initiating machine (infected by a bot) performs additional scans and performs brute force attacks. The victim is unaware that the attacker is logging in and passing traffic through the honeypot to hide their location. The attacker's machine is usually hidden, masked by one or more proxy, VPN, or VPS devices. As a result, the actual IP address may not be the real identity of the attack origin. There are 170,865 recorded unique source IPs from all of the honeypots. An average of

28,478 unique IP addresses were seen at each honeypot. These addresses appeared globally random, except that 6527 addresses were common to all globally diverse six honeypots.

This experiment's analysis of the destination IP addresses and port numbers attempted to map the attackers to their targets. Thom et al. (2021) mapped the attackers to targets and ports to better understand the IP address forwarding behavior from the Cowrie data. The top 20 attackers were listed per target mapped, suggesting that there are more targets than attackers. Some targets were compromised and used to host an attack proxy shared in the botnet network. Geographically, the London honeypot saw fewer unique destination IPs (12,746) than Bangalore (100,252), which saw over seven times as many destination IPs. This behavior of fewer unique destination IPs in London is probably an indicator of attacks such as IP tunneling activity identified in the honeypot data. The unique destination port numbers per honeypot location data showed 1038 destination port numbers were targeted from Bangalore and 718 from Amsterdam, while other honeypots ranged between 20-57. This type of behavior could be attributed to scanning activity from Bangalore. At the same time, New York was targeted with standard services such as web, telnet, SMTP, and SSH. Specific services targeted with noncommon port numbers (i.e., bo2k, Ghidra). According to Thom et al. (2021), future work is to understand better the attack patterns and behavior of malicious activity in actual servers or IoT devices.

Methodologies

Various methodologies show proposed frameworks from these studies and the impact of these solutions have on cyberattack incident response. A survey by Borges

Amaro et al. (2022) found that 38% of the respondents used ad hoc methods to update CTI requirements, weakening cyberattack mitigation and defenses against vulnerabilities. Borges Amaro developed an eight-step cyber methodological threat intelligence framework to collect data, organize, share, visualize, and analyze CTI threat data from heterogeneous sources (Borges Amaro et al., 2022). Cybersecurity analysts use this CTI tool to insert threat data and analyze and visualize timelines for better insights.

The eight-step threat intelligence model contained steps for the methodological framework of CTI solutions (Borges Amaro et al., 2022). The first six steps in the CTI model used a proprietary CTI tool for TI functions such as threat information collection, preparation, and sharing. The last two steps, visualization, and analysis, of the CTI model are to create a threat timeline of IoC information. These eight steps of the CTI model include (a) management, (b) indexer, (c) collect, (d) generate, (e) search, (f) share, (g) visualization, and (h) analysis. In step 1 (Management), stakeholders are the interaction between users and applications controlling the data flow and permissions. Step 2 (Indexer) supports data collection and storage (e.g., relational database). The mechanism of the storage structure supports ingestion of data from CTI data feeds. Step 3 (Collect) provides the external gathering techniques to allow the application to ingest the data. The collection from a different format must be processed into a standardized format so Step 2 can be in the same format. Step 4 (Generate) is the normalization of the internal data, and in Steps 2 and 3, the same pattern can be converted into data feeds. Step 5 (Search) is the data manipulation and exploration of data using fast indexing mechanisms to find the visibility of stored data. Step 6 (Share) allows for internal data such as collections, data

feeds, and threat campaigns to be shared with users with third-party tools using the STIX language. Step 7 (Visualization) is the add-on feature that presents CTI data (e.g., IoC) in a timeline format that enriches the threat footprint. This feature helps interested parties to share a threat picture. Step 8 (Analysis) is part of Step 7, allowing data analysis using functions to obtain the best threat data. Cybersecurity analysts use this eight-step methodology in an architecture to understand the threat environment better.

Applicability of Work

To demonstrate the application of the work, various case studies and methodologies that are meaningful to critical infrastructure organizations were presented to improve the cybersecurity risk management practice. This application of work is essential to my study's research method, where real-world work applications align with the OODA theoretical framework.

A study by Kure and Islam (2019) used CTI information to provide TI methods and sharing techniques to reduce the risk to critical infrastructure organizations. This cyber security risk management (CSRM) model minimizes the risk from multivectored cyberattacks where CTI can provide appropriate controls from strategic, tactical, and operational perspectives. Incorporated in the CSRM model, concepts of CTI are used so that the actor's threat profile and attack tactics and techniques can be used to calculate the risk. These risks are scored (i.e., weighted) and evaluated against the organization’s vulnerabilities, such as CPS systems (i.e., critical infrastructure assets).

CPS are real-time management of independent systems with high-performance requirements in critical infrastructure (Kure & Islam, 2019). Due to the complexity of

CPS systems, the interdependencies of technological components and evaluating the risk to the system can be difficult. The digitalization of the main grid and the decentralization of the electric sector increases the exposure of CPS systems to network vulnerabilities in microgrids (Kraemer et al., 2022). The risk management of an integrated cybersecurity system depends on the effectiveness of the control (Kure & Islam, 2019). The effectiveness depends on the system's deployment of information and communication technology. This system's complex relationship between information and communication technology and CPS risk depends on deploying cyber and physical approaches. A risk-based methodology is applied to security management systems in critical infrastructure.

A risk-based methodology used in the DisCos power holding company case demonstrated its applicability (Kure & Islam, 2019). The DisCos power security management system, which uses a SCADA infrastructure, provides reliable service and is one of Nigeria's largest critical infrastructures. It distributes electricity to 30,000 customers and offers last-mile services. DisCos power needs continuous monitoring and power protection in the electricity supply chain.

The DisCos power company encountered a spear-phishing email in a recent event (Kure & Islam, 2019). This email contained a malicious Microsoft file attachment that infected the network and operating systems and targeted the SCADA system. The resulting download of the malicious file compromised was the hashed credentials of the server message block (SMB) that identified the organization's network. This compromise allowed data about the sensitive SCADA system to be leaked along with credentials, network assets, control systems, and outlook web access servers that contained email

accounts. The compromised entry point was the virtual private network (VPN) used to connect the organization to users from other networks. Remote users enter the organization's network through the VPN giving access to the networks, systems, and database. When the organization's active directory was compromised using password cracking techniques, it revealed the identity of the administrators and computers. As a result, attackers gained full access to the system that turned off the electrical power to 30,000 customers (2 days up to 2 weeks).

Using the DisCos case study (Kure & Islam, 2019), the data were collected from internal sources by reviewing documents, cyberattacks, risk registers, assets inventory, and internal audits. An internal audit of collected data was analyzed using qualitative and quantitative variables. These variables included the weight of vulnerability, assets criticality, TTP, risk level, and control effectiveness, where the threat actor, assets, vulnerabilities, and threats to determine asset weight were considered. The asset weight evaluated measures effectiveness using a formula, control rating, and metamodel.

In this scenario, the threat actors' profiles were assessed using TTP (Kure & Islam, 2019). Identified from the TTP method are related patterns associated with threat actors. When threat actors perform an attack, the attack progress performed used the TTP methodology by Mitre to categorize attacks. These attackers used cyberattack tools that can be identified and classified using the Mitre methodology. In this example, eleven tactics and techniques provided by Mitre identified these threats. Once these threats were identified and categorized using TTP from the STIX model, stakeholders can better manage and share this threat information. This threat identification is essential to my

study because threat identification is observed in the first stages of Boyd's OODA framework process.

The CTI concepts within the TTP pattern are indicators used to detect suspicious behaviors from cyberattacks. This contextual information associated with the CTI value known as IoC has been commonly found in three categories (Kure & Islam, 2019). These three categories are network indicators (e.g., URL and domain names), host-based indicators (e.g., analysis of infected computers), and email indicators (e.g., email services used for social engineering). These indicators are essential to my study's theoretical framework associated with the OODA process, where orient leads to the second stage to decide and act.

Kure and Islam (2019) used an asset formula to determine the level of risk due to a cyberattack. The potential risk of financial loss, reputational damage, privacy violation, and noncompliance were some factors for understanding the attack's motivation. The following steps included a high, medium, and low-risk score.

Step 1: Estimate the likelihood of the attack.

$$L = (TAF + VF)/n \quad (1)$$

Where

L = Likehood of attack, TAF = Threat Actor Factors, VF = Vulnerability Factors, n = total number of factors.

Step 2: Estimate the overall impact of a successful attack.

$$I = AF/n \quad (2)$$

I = impact of the attack on the organization, AF = Asset Factors, and n = total number of factors.

Step 3: Estimates the likelihood of combined impact and overall severity of a risk.

$$R = L \times I \qquad (3)$$

R = the risk level, L = the likelihood of attack within the timeframe, I = the attack's impact on the organization. This associated risk formula is helpful in my study's theoretical framework in determining the level of risk during the second stage of the OODA process, where a decision is needed.

The overall summation of five parameters determines the effectiveness of the attack. These parameters determine the effectiveness of the individual controls that can be used for improvement (Kure & Islam, 2019).

$$OE = R + S + C + I + T \qquad (4)$$

Where

OE = Overall Effectiveness, R = Relevance, S = Strength, C = Coverage, I = Integration and T = Traceability. The control formula for overall effectiveness is useful to my study's theoretical framework, where the impact of a decision during the OODA process affects the overall action. Kure and Islam (2019) used this risk formula to measure overall effectiveness (OE) from the CTI concepts within the TTP indicator patterns. These elements provide an overall score, where the analyst can make informed decisions when applied to the OODA theoretical framework.

Transition and Summary

This study utilized the OODA theoretical framework methodologies. The cybersecurity roadmap of methodologies that led up to CTI was based on highlights related to the relationship between cybersecurity attacks and mitigation (Andrade et al., 2022). In the first section of academic literature, the review examines the theoretical framework, Boyd's OODA model relative to cyberattacks, research variables, and correlations. In the second half of the academic literature, the application to the applied IT Problem, a quantitative research approach is applied to the IT problem. An in-depth look at the methodologies aim to understand better the strength of a correlational relationship between variables (e.g., CTI attack identification and critical infrastructure industry assets). A measurement of meaningful and effective formulas in case studies is applied to the IT problem.

In section one, a model that supports this decision-making process is Boyd's OODA for dynamic environments (Thulfiqar & Manmeet, 2022). In the OODA model, the observe domain is the fusion of information building situational awareness (Andrade et al., 2022), and the orient domain provides the security analyst practical knowledge. In the second half of the OODA model, the decide and action refer to cybersecurity mitigation and risk. The decision domain is the analyst's knowledge of all the hypotheses from CTI, and the action domain is from the analytical reasoning and the moment to take active defense. An active defense network can be directed to dynamically adjust configurations based on attacker behavior (AlQadheeb et al., 2022). The security system's

active defense is enumerated by looking at the independent and dependent variables from CTI, attack indicators, and critical infrastructure assets.

In the second section of the academic literature, the application to the applied IT problem, a quantitative research approach is applied to the IT problem. This quantitative research approach aims to understand the correlational strength relationship between variables of CTI attack identification and critical infrastructure industry assets. The explored methodologies are MITRE ATT&CK which allows analysts to better understand the event correlations through a standard explanation of cyberattack variables (Jabar & Singh, 2022). Some of these variables include Mitre ICS (Industrial control system) related to critical infrastructure assets, a dependent variable in the research question. This study's independent and dependent variables exemplify various methodologies in the applicability of work from case studies (e.g., CTI datasets, sharable data TIPs, data enumerated attack indicators, and methodologies that measured incident response times and overall effectiveness).

Section 2: The Project

The purpose of this quantitative study was to examine the relationship between CTI's attack identification and critical infrastructure assets in the critical infrastructure industry. The strength of the correlation between these variables (independent and dependent) involves the ex post facto comparative study of CTI attack indicators (e.g., IP address) relative to the critical infrastructure assets. The research question this project aimed to answer addressed the effectiveness of known CTI data feeds and the correlation to infrastructure assets (e.g., ICS devices, OT devices) within the critical infrastructure industry in the United States. A similarity-based approach (Lu et al., 2021) is one method to analyze the alerts correlated from CTI data feeds, forming a causal link to infrastructure assets and quantifying the number of meaningful (Albasheer et al., 2022), actionable intelligence feeds. Measured are the overall results using Boyd's theoretical framework (OODA) to evaluate the experiment's effectiveness using the effectiveness formula.

Purpose Statement

When some SOC analysts receive TI about potential cyberattacks (e.g., attack indicators), a decision is needed to take mitigating action. A SOC analyst may be overwhelmed with the number of incidents from CTI indicators, and deciding which ones may be meaningful to their organization can be stressful. For the SOC professional to promptly process a large quantity of CTI feeds requires understanding what is meaningful to the organization. In the case of a critical infrastructure organization, the SOC operator needs to correlate actionable to the organization's critical infrastructure assets (e.g.,

industrial control systems, OT devices). This quantitative research problem aims to understand better the strength of a correlational relationship between CTI attack identification and critical infrastructure industry assets (e.g., electric generators, SCADA equipment; Kayan et al., 2022). The result of this study can evaluate how effective CTI is to a particular industry, resulting in meaningful mitigating actions to prevent disruptions to critical business operations. As a result, the increased resilience to critical infrastructure will have a social impact allowing customers to access essential resources (e.g., electricity, water, gas).

Role of the Researcher

My role as the researcher in this quantitative study in the data collection process was to find comprehensive CTI data sources and to determine what actionable intelligence is relative to the organization's goals. This relationship between these data sources in TI and actionable cybersecurity intelligence in critical infrastructure relates to my experience in the Department of Defense (DoD). My 8 years of experience as a cyber network analyst, including 3 years as a red team operator, added to my role as the researcher in determining which threats are actionable. Added to real-world experience, I have had a security plus certification from CompTIA and a master’s certification in computer networking and security from Johns Hopkins University. With over 10 years of experience in cybersecurity and IT, I was not biased in choosing CTI data because data points were from independent sources (e.g., CTI vendors, government websites, and universities). I chose high-risk targets to test the strength of these correlations quantitatively. Guidance from government policies such as FIPS 199 [15] and NIST SP

800-60 [25] applied to industrial control systems specific assets identified as high-risk targets (NIST, 2015).

Because the Belmont Report (Department of Health, Education, and Welfare, 1979) protocol involves the use of human subjects, the research focused on the quantitative data process that did not include human subjects. The methods used in the OODA model did not involve a human-in-the-loop for this experiment (Andrade & Yoo, 2019). Instead, the research focused on collecting, processing, and analyzing high-volume CTI data feed.

Research Method and Design

The specific research method I used in this study was quantitative. A quantitative methodology was used to test the relationship between two variables and determine if the correlation is significant. The known variables were the datasets from CTI and critical infrastructure assets. In an ex post facto design experiment, the researcher can set up a TIP using a SIEM instrument (de Melo e Silva et al., 2020). On this platform, data analysis using the instrument R will find the strength of the correlation between research variables using correlational statistics (Abbasnasab Sardareh et al., 2021). The resulting data can be displayed as a graphic visualization (Fife, 2022). The goal is that the data analysis will make it possible to understand the relationships between the independent and dependent variables.

Research Method

The research method chosen was quantitative because the plan was to test the relationship between two variables and to analyze the data using an unbiased design. For

a researcher to meet these objectives, quantitative research is conducted to answer a specific research question. According to Mohajan (2020), a null hypothesis can test the relationships between the dependent and independent variables. Using a statistical approach, the researcher can measure the relationship's strength using a null hypothesis test. A null hypothesis assumes that no relationship exists between the variables, and the results can indicate whether the null hypothesis is accepted or rejected (Mohajan, 2020). As opposed to a qualitative method that typically uses questionnaires, the researcher would not be able to determine the strength of this relationship effectively through nonmathematical means (Kong et al., 2023).

Qualitative and mixed methods, a combination of qualitative and quantitative research methods, were not considered in my research study for the reasons of bias. Collecting data (e.g., interviews, open-ended questionnaires) in a qualitative approach may introduce human bias without controls (Tennant & Ross-Hellauer, 2020). The general definition of qualitative research uses words to describe the research topic, unlike a quantitative approach, which uses numeric values to describe the research topic (Tennant & Ross-Hellauer, 2020). Using words in users' responses from questionnaires may be subjective and imprecise (Jabar & Singh, 2022). To answer the research question precisely, users needed to have high technical knowledge of cybersecurity and critical infrastructure to prevent bias.

Research Design

In this research design, the type of quantitative method was correlational. This correlational research design investigated the strength of the relationship between

variables using correlational statistics (Mohajan, 2020). The relationship in this study was between these two variables, cyberattacks and critical infrastructure targets. If these two variables are related, a positive relationship will show a correlation in both variables. In contrast, a negative relationship will show an increase in one variable and a decrease in the other variable. An unrelated relationship is when one variable increases while the other variable remains unchanged. In this study, an example of an unrelated relationship is that any cyberattack increase does not necessarily correlate with the critical infrastructure target.

Whether or not a variable is significant requires data analysis. One example by Thom et al. (2021) used a correlation approach by quantifying IP addresses from the CTI dataset and correlated cyberattack events through real-world honeypots. In another research design of how liked an application is, researchers analyzed the correlation (ANOVA) between the application and variables such as usability, age, and quality to find which variable was statistically significant (Yan Guo et al., 2022).

The type of quantitative method, correlational, was selected because it fit the research question by determining the degree, strength, and type of relationship between variables but did not determine the cause and effect (Mohajan, 2020). The other research design types, descriptive, quasi-experimental, and experimental quantitative design, were also considered. A descriptive quantitative design is used to describe a phenomenon, typically with large populations, and does not involve manipulating variables (Mohajan, 2020). An example of descriptive design is a study by Naseem et al. (2021), which used word cloud analysis to filter Tweets' hashtags (e.g., #COVID19, #coronavirus) to

represent topics. A quasi-experimental study involves examining causal relationships and asking about the effect of one variable on the other (Mohajan, 2020). An example of a quasi-experimental causal relationship by Lu et al. (2021) used a similarity-based method for correlation analysis based upon alerts from CTI data feeds and causal links to asset variables related to alerts. An experimental design examines the causal relationship of the dependent variable manipulating the independent variable (Mohajan, 2020).

Population and Sampling

For data collection, I used the CTI source feeds in a rich text format such as STIX. STIX is the principal CTI data collection format using a rich, extensive markup language (XML) format. At the same time, really simple syndication (RSS), a lightweight XML format, designed for news items (Ramsdale et al., 2020) was collected. The CTI vendor selected was AlienVault, using the STIX format on a TI OTX platform. STIX, a serialization format by OASIS, is standardized across CTI exchanges and TIP platforms (Chen et al., 2024). In a study by Chen, 232,370 pieces of intelligence from OSINT sources were evaluated creating IoC and generating STIX data bundles. According to Ramsdale et al. (2020), over 275 data feeds from CTI vendors were selected for analysis, and of those 275 data streams, 125 analyzed contained the STIX2.x format sharing protocol. This STIX format is a critical protocol feature (e.g., SSH, FTP, IMAP) for sharing timely, actionable TI amongst communities and organizations. The STIX structured language uses the TAXII TIP for sharing by allowing timely inputs of threat data by the community. This data criterion in a STIX format better represents a wide range of cyberattacks with fidelity relevant to my research question.

Another criterion for dataset selection was the vendor's support of a wide range of cyberattacks from phishing to ransomware attacks that included variables related to my research question (e.g., IoT, OT devices). Other vendors were considered, such as IBM's X-Force Threat intelligence dataset (IBM, 2022) with indicators in manufacturing, malware, extortion, identity, phishing, IoT, cloud, and ransomware (Alaeifar et al., 2024). According to the X-Force Threat Intelligence Index 2022 report (IBM, 2022), a 2,204% increase in reconnaissance of OT devices targeting SCADA Modbus devices from the Internet was reported between January and September 2021. Cybercriminals target SCADA Modbus devices because these are critical infrastructure assets (e.g., IoT, OT devices) that energy companies use. This IBM X-Force dataset was relevant to my study's dependent variable of critical infrastructure assets, where the model contains a similar relationship between IP addresses and TTPs (Noor et al., 2019). The expected volume of dataset size, according to Liu et al. (2022), had an actionable CTI population of over 29k cybersecurity reports from 2000 to 2021, and using their trigger-enhanced TriCTI system, they found 113,543 reports of meaningful, actionable TI. Although IBM X-Force was considered but not used in this study, AlienVault provided an economical and open platform for the community to access and submit TI data.

The three datasets were sampled from three different sources: AlienVault, CISA, and Temple University. I collected the first and second datasets from October 28 to November 14, 2023, from AlienVault and CISA, respectively, and the third dataset from Temple University was collected on December 15, 2023. The first dataset (CTI) represented a daily active collection of the independent variables (IP, domains,

hostnames). At the same time, the second dataset represented the dependent variables (ICS advisories) from the government agency CISA of daily RSS feeds. The third dataset from Temple University was used for triangulation, representing a static collection of real-world critical infrastructure ransomware attacks collected at the same period during the first and second datasets (AlienVault, CISA).

Ethical Research

The ethical research rule Common Rule 45 CFR, Subpart A (HHS.gov, n.d.) of the basic HHS policy protects human research subjects from administrative harm. The Walden IRB process was completed before any data collection with the approval of 05-02-23-0662736.

Data Collection

Instruments

The primary CTI data feed collection instruments consisted of Microsoft Azure Logic Apps. The advantage of using this Azure cloud approach was the automated gathering of dynamic data feeds from CTI data sources. Using Logic Apps Designer, I implemented a script for dynamic data collection that recurred at a certain time daily as the first module in the process scheme. The second module provided services that perform HTTP requests (i.e., authentication) to the TAXII servers hosted by the CTI vendor AlienVault. In this request, I specified the types of indicator feeds (IP, domains, and hostnames). A sample of data collection from AlienVault is shown in Appendix A in a table format; Appendix E contains raw output, and Appendix I is the Logic Apps code used for CTI data collection.

Similarly, the same instrument, Microsoft Azure Logic Apps, was used for the second data collection set from CISA. A script was written in Logic Apps Designer to pull the RSS feed of ICS Advisories from the CISA website. A sample of this data collection (CISA) in table format is shown in Appendix B and Appendix E in the raw output format, while Appendix J contains the code for the RSS data collection. While the instrument used for the first and second data collection was automated, collecting data dynamically, the third instrument used for data collection was static, using a Microsoft Excel spreadsheet.

The third data collection was static data collected by Temple University using Microsoft Excel. The spreadsheet instrument data contained the text of critical infrastructure attacks (ransomware) from 2013 to 2023.

Data Collection Technique

The data collection technique used involved externally sourced observables provided by vendors. Ramsdale et al. (2022) identified 22 vendors of TIPs using different formats such as text, CSV/RSS, JSON/XML, STIX 1.x, STIX 2.x, MISP, IDS, and DNS. In a comparative analysis of the 22 surveyed using TI sharing platforms (i.e., TIPs) by Ramsdale et al. (2020), the CTI data format most adopted by the industry is the STIX protocol and TAXII sharing system. The vendors identified in the survey (Ramsdale et al., 2020) include abuse.ch, AbuseIPDB, Bambenek Consulting, blocklist.de, botvrij.eu, C1fApp, Censys, CINS Army (SIEM), cybercrime-tracker, Dshield (SANS), FreeTAXII, Green Snow, HAIL A TAXII, Limo, Malc0de database, Malware Domain List, MISP

(CIRCL), PickUpSTIX (NC4/Soltra), Spamhaus, TAXIIstand, ÜberTAXII, and Xavier.Mertens.consulting.

In contrast, the unstructured CTI formats do not follow a uniform protocol structure but organically introduce semantics such as blogs (e.g., Twitter), forums, vendors (e.g., dark web, Onion services), and mailing lists (Ramsdale et al., 2020). These unstructured data formats are collected from a community of external open-source (e.g., OSINT) sources (Basheer & Alkhatib, 2021). These OSINT sources include news feeds, public advisories, search technologies from automation engines (e.g., Shodan), antivirus vendors (e.g., malware threats), communications monitoring (e.g., Twitter, Slack, IRC), and Dark Web intelligence (e.g., forums from the criminal underworld, Onion services; Moraliyage et al., 2022). Analyzed from the unstructured datasets are the categories or domains that compare two probability distributions (e.g., Jensen Shannon divergence model; Suryotrisongko et al., 2022). This function helps calculate the target domain's name distance to the dataset (e.g., Botnet). The unstructured CTI format is a practical strategic technique at a high level, whereas a structured CTI format is a realistic view at a low level with actionable intelligence.

From the surveyed CTI sources, feeds, and formats, a structured CTI approach was selected for this research study's data collection technique. The STIX/STIX 2.0 protocol was selected because it is standardized and enumerated across various datasets. This standardization allows the taxonomical analysis of variables to be credible (Ammi et al., 2022). CTI-based ontologies will provide a common language that can interconnect the various taxonomies, enhancing their effectiveness (Mkuzangwe & Khan, 2020).

Another reason for selecting the STIX protocol is to reduce bias (Chen et al., 2024). With the standardization of protocol format, enumeration of observables, and common structured sharing platform, systemic bias is reduced in the intelligence system using externally sourced observables with open standards (Meland et al., 2021). The STIX/TAXII protocol defines a data collection system of a structured dataset by a committee across vendors (Ramsdale et al., 2020). This standardization has led to many interoperable cyber information-sharing architectures. For instance, a data connector is a mechanism to connect the data feed to the instruments needed to process the data. The TI sharing architecture is data from the TI vendor (e.g., AlienVault, IBM X-Force) to the TAXII server (i.e., sharing server) to the data connector (e.g., TAXII connector) on the TIP (e.g., MS Azure SIEM). This technique and data collection process of adding threat indicators from CTI vendors to the TIP using the TAXII data connector is the architecture used in Microsoft's Azure cloud service SIEM (de Melo e Silva et al., 2020).

Data Organization Techniques

The data organization of the structured type (e.g., AlienVault, IBM X-Force) uses the STIX protocol, where the data is categorized into various cyberattack groups (e.g., ransomware, denial of service, phishing). The data in STIX is organized by answering "what" TI is available for sharing, while the data in TAXII answers "how" that information is shared and relayed (Ramsdale et al., 2020). The machine-readable technique is used in STIX and TAXII and can be easily automated.

Data Analysis Technique

The MITRE ATT&CK framework can best describe these categories as threat vectors characterizing the cyberattack into smaller attack components. The MITRE ATT&CK framework categorizes the threat vectors into cyberattack groups and uses common descriptive language to enumerate the threat landscape (Borges Amaro et al., 2022). The Mitre framework for cyberattacks can match various threat vectors to CTI data feeds (Jabar & Singh, 2022). The CTI data feeds are organized in the STIX2.x protocol that contains the IoC. These categories can be correlated between the threat vectors and CTI data feeds. The data feeds are live streams and stored according to each vendor, and the data feed received by the client can be stored accordingly and displayed with graphical visualizations (Borges Amaro et al., 2022). These data visualizations can be performed by SIEM systems (González-Granadillo et al., 2021).

Cyber security analysts analyze the CTI data feeds using tools to extract critical indicators from the datasets. One of these tools includes TIMiner by Zhao et al. (2020) to extract the IoC within the data feeds to produce an efficient domain recognizer. This data mining of a word-based approach is used by some related works focusing on automating the extraction of threat entities (Ren et al., 2022). The cybersecurity expression of the threat knowledge is displayed as an IoC graph structure using NLP for retrieval to extract threat actions. These threat actions by the SIEM can be correlated to data points where the analyst can identify similar threats (e.g., attack events, targets, threat actors, infrastructure). By observing this relationship between the attackers and critical infrastructure assets (e.g., electrical generators, water reservoirs, gas pipeline valves), the

analysts can orient and decide to act (e.g., blocking specific IP addresses; Rahman et al., 2023). This relationship between defenders and attackers of a critical infrastructure system that utilizes SCADA industrial frameworks is an area researchers can explore active defense (Cifranic et al., 2020).

SIEM provides critical infrastructure security solutions to analyze and correlate critical infrastructure assets' incidents (González-Granadillo et al., 2021). Since critical infrastructures rely on monitoring industrial control devices such as SCADA systems over a network of control systems (NCS), SIEM solutions provide real-time analysis. A security orchestration automation and response (SOAR) system is employed when a SIEM provides TI and automated response. This automated response is a correlation between a database of TI and endpoint detection and response (EDR), also known as endpoint threat detection and response (ETDR) (Karantzas & Patsakis, 2021).

A SIEM service is used to correlate big data from CTI datasets (e.g., AlienVault OSSIM) and perform analysis using correlation rules for critical infrastructure SCADA systems (Radoglou-Grammatikis et al., 2021). One technique using the cloud (e.g., MS Azure) for data analysis is the workbook (Wescott, 2020; de Melo e Silva et al., 2020) is programmed to gather large datasets from data-sharing platforms (e.g., TI) and pick out meaningful target-centric observables to store in smaller datasets for storage, alerts, or additional analysis. In workbooks, the analyst can write scripts to call upon instruments. The workbook instrument "R" is a statical tool like IBM's "SPSS" statical tool (Abbasnasab Sardareh et al., 2021) that the researcher can use to calculate the correlation strength between variables. Workbooks allow stakeholders to customize the TIP for data

analysis to reveal vulnerabilities, threats, countermeasures, cyberattacks, risks, and critical network assets. Critical network asset security controls can be mapped to MITRE ATT&CK Framework for analysis (Center for Threat Informed Defense, n.d.). Researchers have used this modeling technique of real-world simulations to digitally model the electric power ecosystem using an experimental platform to test for reliability and resilience (Salvi et al., 2022).

Data Analysis Validity

The threats describe the study's validity to external and internal data systems. A threat intelligence management platform (TIMP), the cloud platform offers scalability, reliability, and technologies effectively under a manageable platform (Zibak et al., 2021). To find the threats to internal validity, the integrity of the data feed ingested and the infrastructure used (e.g., data connectors) needs monitoring. External to the infrastructure, the data source threats are the quality of the data feeds (e.g., uniqueness) and data source types (e.g., IoC) meaningful for tactical intelligence (e.g., IP addresses). The analyst can monitor the health of these internal validity indicators using the data connectors in the Microsoft SIEM service (de Melo e Silva et al., 2020). By monitoring the health of the data connector, the volume of data feeds can be monitored for internal validity to determine if the data feeds are functioning at standard capacity. The analyst can monitor the quality of data feeds received from the data connector by the type of data feeds valid for tactical intelligence (e.g., IP addresses) using the SIEM service workbook. The data feeds can be compared to real-life events for external validity (see Appendix D).

An external dataset of cyberattacks by data type can be compared to internal datasets for external validity.

A statistical method using linear regression analysis can determine the strength between the variables with variations. These variations between independent and dependent variables of the actionable intelligence and dependent real-world values can tell the magnitude of the effect. The magnitude of the effect using a regression equation is the Pearson correlation coefficient (Martineau et al., 2023). The Pearson coefficient can determine the strength between the data source (CTI) attack identification and the critical infrastructure assets relating to IT critical infrastructure attacks (Lu, 2020).

Reliability and Validity

Reliability

One of the most popular instruments for researchers to test for statistical correlation is IBM SPSS. This popular software analytic tool in academia focuses on statistical investigation, emphasizing data analysis (Abbasnasab Sardareh et al., 2021). SPSS has significant quantitative data analysis capabilities to perform descriptive statistics and test hypothesis significance (e.g., t-tests, ANOVA, correlation, ordinary least squares regression). Another popular statistical analytic tool, R (R Core Team, 2021), is free (open-source) compared to SPSS, which is low cost per client with site license (e.g., University) purchased. R is open-source software that can do a Pearson correlation test using the function "cor( )," resulting in a Pearson correlation coefficient value that will give the strength between two numeric vectors of the same length (Sancho et al., 2020).

Validity

The threat to this correlation test is that some data values that lay outside the normal distribution or abnormal distance from other values, called outliers, may need to be scrubbed from the dataset (Fife, 2022). These outliers represent an abnormal event determined by the analyst. Researchers have other techniques using R to represent data models graphically (e.g., bivariate, multivariate) using functions to create meaningful visual differences (e.g., ggplot2, flexplot). This method can be used to test the external validity. A dataset for external validity can be cross-correlated with real-world data datasets (Temple University, n.d.) and triangulated with the results from experimental data. Across these various datasets, the enumerations of variables are aligned and verified using a common language or meta-language (Rantos et al., 2020). The measurement of the stakeholder's perceived data from the experiment is a combination of the internal and external validity of the TIP, CTI data quality, and external usefulness of the data to the organization (Zibak et al., 2021).

Transition and Summary

Internal and external validity convergence is based on several factors, such as data quality, content, and platform. The measure of validity from a survey by Zibak et al. (2021) of 152 security professionals about TI sharing models revealed the importance of effectiveness. Some of the success factors from Zibak's study of the TIMP are the quality of the TI content and the trust permissions of the TIMP platform. This measurement model assessment by Zibak et al. (2021) confirms the convergent validity, discriminant validity, and construct reliability to evaluate the inner structural model of TIMP

platforms. Busquim e Silva et al. (2021) approach verification and validation using a simulation model conducted in parallel with conceptual model validation and computerized model verification to assess the experiment referencing subsystems of a nuclear power plant. Overall, the empirical validation of this experiment will help advance the theoretical and practical understanding.

Section 3: Application to Professional Practice and Implications for Change

Overview of Study

This study was a quantitative correlational study about the relationship between CTI's attack identification and critical infrastructure assets that some SOC analysts use to decide whether to act in a cybersecurity attack. A SOC analyst uses CTI's threat indicators called IoC to identify potential cyberattacks and other types of TI such as TTP to build security information and reports (Schlette et al., 2021c).

These findings are presented in two parts: an experiment and an application. The experiment answers the research question, while the application answers the theoretical model. The experiment consists of independent and dependent variables used to find if there is a meaningful relationship. Data were collected from various sources to create datasets using manual and automated data collection techniques. The experiment's results proved a similar-based correlation between cyberattacks and CTI in critical infrastructure. The application representing a SOC environment demonstrated the effectiveness of Boyd's theoretical model (OODA) in practice, highlighting the option for decision-making automation in deciding whether to act in a cyberattack.

Presentation of the Findings

Description of Tests, Variable, Purpose, Hypotheses

The similar-based correlation method I used in this study was based on matching indicators between two datasets, finding a similar link between TI and real-world cyberattacks on critical infrastructure. A direct correlation of 55 matching indicators from the experiment and 138 matching indicators in the application answered the research

question of links between CTI and critical infrastructure assets to an organization (Dekker & Alevizos, 2024).

A quantitative statistical approach quantified this relationship between two objects (Alert, CTI). A Jaccard similarity index was used to find the number of matching indicators (Hamroun et al., 2024) from the intersection of two data objects, A and B. Object A represents a pool of alerts, and Object B is a pool of CTI indicators. The strength of similarity between the objects was tested using the Jaccard equation, resulting in a value between 0 and 1. This measurement indicated the degree of similarity between two objects that share matching indicators. The correlation is represented as the intersection between two data sets, similar to a Venn diagram.

The purpose of these statistical tests impacting a SOC analyst is to measure the similarity between the independent variable CTI and the dependent variable Alerts. The hypothesis is that actionable intelligence correlates statistically with CTI attack identification and critical infrastructure assets. The impact for an overwhelmed SOC analyst is trying to identify which attack is actionable to their industry (Liu et al., 2022).

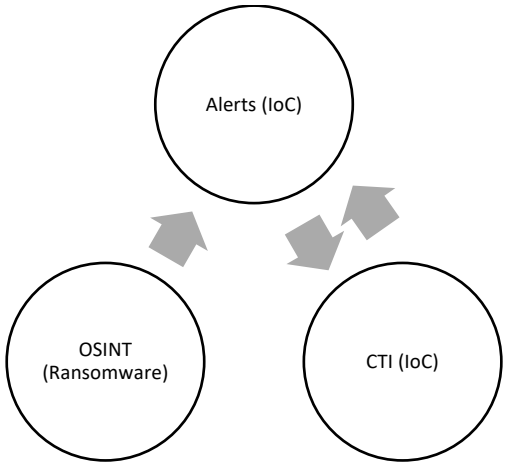
Data Collection Methods

The method for data collection consisted of gathering indicators (i.e., IoC) using causal links from security information sources and reports (Schlette et al., 2021c). The data collection method used to create the Alert dataset of IoC is finding data sources that use the STIX format (Irshad & Siddiqui, 2023). This STIX format provided a familiar syntax for the Alert dataset, linking IoC between datasets. The dependent variable of Alerts and the independent variable of CTI are directly linked (Figure 1). These linked

elements of OSINT, Alert, and CTI are represented in a data model with the direction of data flow. The data flow is a one-way direction from OSINT to Alerts, while data flow is bidirectional between Alerts and CTI (Figure 1).

Figure 1

Data Model of Dataset Relationship



Note. This data model represents the relationship between datasets and the data flow for this CTI experiment. The nature of the data flow between datasets has a common link to the indicator of compromise (IoC) in Alerts.

The relationship between independent and dependent variables can be described as the intersection of indicators. A data model (Figure 1) was created to explain the dataset relationship, the taxonomy of CTI indicators (IoC), and data flow. In the data model, the indicators in the alert dataset are common to the two datasets (CTI and OSINT). In a similar-based link approach, the correlation of two groups, A and C, (A –

B) and (B – C) represents a Venn diagram with B as the intersection. B represents a matching variable to link datasets A and C. An analyst (i.e., cyber professional) can use this correlation information in a SOC to gain insight into cyberattacks by correlating indicators using a causal inference (Sen et al., 2022).

Statistical Methods

A different descriptive dataset evaluation was needed to match the statistical method. Traditional methods to correlate two objects using Pearson's statistical analysis (Martineau et al., 2023) assume a linear regression of the datasets. Due to the nature of the dataset volume generated, I used automated techniques utilizing a data connector on the cloud, which allowed CTI data collection to be ingested at a high rate from the CTI source to cloud computing applications (Marinho & Holanda, 2023). This rapid increase in cyber data volume generated from different devices and sources is like the issue of big-data analytics (Rawat et al., 2021). Relevant to my experiment and application findings, descriptive statistics is a similarity-based link between two objects (Kotenko et al., 2022). Any two matches of fingerprints (e.g., hash values, IPv4, domain) of IoC between the two objects provide a unified presentation of correlated threat data from different sources (Gao et al., 2020).

Jaccard Coefficient

An example of this similarity can be evaluated using the Jaccard coefficient (Schaberreiter et al., 2019). The Jaccard coefficient is a similarity algorithm used to answer how similar or dissimilar two datasets are proportional to the intersection size of the union of these datasets. This statistical measurement was used by Schaberreiter

(2019) to quantify the CTI attributes of quality or trust in data sources. A Jaccard and cosine similarity was used to discover correlations between IoCs and APT groups behaviors (Chen, Vaitheeshwari et al., 2024). The relationship between two groups using the Jaccard coefficient (J) or the Jaccard index (Koeneman & Cavanaugh, 2024) is a useful statistical tool where the union and intersection signify a minimal value (0) and a maximum value (1), indicating minimum to maximum range of similarity. Roemsri and Hewett (2021) used the Jaccard coefficient to show device types in network traffic logs as a crucial step in identifying IoT devices as part of the security review. I will use the Jaccard coefficient to match vulnerabilities identified in the Alert data set against network CTI data feeds.

The Jaccard coefficient was selected instead of the Pearson correlation coefficient (PCC) for statistical measurement because PCC requires a traditional linear regression dataset for analysis. The Jaccard index provides the ability to determine significant feature matching for finding nonlinear correlations. (Kamiura & Sekine, 2023).

A spatial analysis of the experimental data (CTI) resulted in a nonlinear graph using the IBM SPSS statistics modeler instrument. A traditional Pearson correlation analysis using the linear regression technique was unsuitable due to the nature of the TI data (CTI). In contrast, the Jaccard coefficient does not require linear regression to understand how similar the two data sets are. Instead, the Jaccard coefficient measures the intersection area between the two data sets. This method is better suited to the datasets where A is a partial intersection of B, resulting in a unary rating of the data for quantification (Berahmand et al., 2022).

$$J(A,B) = \frac{|\,A \cap B\,|}{|\,A \cup B\,|} = \frac{|\,A \cap B\,|}{|\,A| + |B| - |A \cap B\,|}$$

The Jaccard similarity index, $J(A,B)$ equals the intersection (∩) of the objects (A, B) divided by the size of the union (∪) of the objects of A and B (Verma & Aggarwal, 2020). The resulting value of the Jaccard index (J) is between 0 and 1 or $0 \leq J(A,B) \leq 1$. The greater the area of the intersection over the union (IoU), the greater the similarity between the two objects. The IoU can be expressed as the intersection between two objects, A and B, and the attributes within these objects can be represented as either 0 or 1 (nonmatching, matching). The total number of binary attributes (n) must be in one of these categories (M) as represented in the matrix ($M_{11}, M_{01}, M_{10}, M_{00}$).

$$M_{11} + M_{01} + M_{10} + M_{00} = n.$$

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$

In the Jaccard equation (J), M_{11} represents the total number of attributes with 1 in both objects A and B (Table 1). $M_{01}$is the number of attributes that are 0 for A but 1 for B, and M_{10} is the number of attributes that are 1 for A but 0 for B. The M_{00} represents the total number of attributes with 0 in A and B. Using *n* binary attributes in the Jaccard equation (J) is an asymmetric binary similarity measurement (Koeneman & Cavanaugh, 2022) between two data objects (A, B). In the experiment shown in the matrix (Table 1), the intersection M_{11} equals 55, representing the total matching attributes (i.e., indicators) in A (i.e., Alerts) and B (i.e., CTI). The M_{10} is 119, representing the number of nonmatching attributes in A, and M_{01} is 0, representing the nonmatching attributes in B. M_{00} equals 64 because there are matching attributes of value 00 in both A and B. The

number n represents the sum of the attributes equals 238. The Jaccard index equation, the intersection (M_{11}) is 55 divided by the union ($M_{01} + M_{10} + M_{11}$) is 174, where J equals 0.316. The Jaccard index value (J) is between 0 and 1 and is significant if J is greater than 0.5. Since the Jaccard index of 0.316 is less than 0.5 then the relationship between A (Alerts) and B (CTI) is insignificant. Another more accurate calculation than the Jaccard index is a vector-based approach called the cosine similarity (Sarwar et al., 2021).

Cosine Similarity

The same data were applied to a different similarity index using the angle to measure the distance between objects, called the cosine similarity. Cosine similarity has been used as a computing distance algorithm for grouping CTI threat reports, articles, social media, and forums (Rahman et al., 2023). A survey of similarity indexes (Verma & Aggarwal, 2020) showed that the cosine similarity index was more accurate than the Jaccard index (Sarwar et al., 2021). In place of the Jaccard index with the same dataset, the cosine similarity was used to calculate the similarity between CTI and critical infrastructure assets using the same instrument, IBM SPSS. The results with a value between -1 and 1 showed a cosine similarity equaling 0.562 in the experiment. Since the cosine similarity of 0.562 is not greater than 0.5, the relationship was insignificant between CTI and critical infrastructure assets. In the following application portion, this CTI control variable was used to find if there were any changes to the dependent variable of Alert IoCs.

Table 1

Matrix of Jaccard Index for the Experiment

n	Experient		Application	
	Object A			
	0	1	0	1
Object B				
0	$M_{00} = 64$	$M_{10} = 119$	$M_{00} = 64$	$M_{10} = 36$
1	$M_{01} = 0$	$M_{11} = 55$	$M_{01} = 0$	$M_{11} = 138$

Note. In the experiment, the intersection M_{11} equals 55, representing the matching attributes in A and B. The M_{10} is 119, representing the number of nonmatching attributes in B, and M_{01} is 0, representing the nonmatching attributes in A. M_{00} is 64 with 00 matching attributes in A and B. In the application, the intersection M_{11} equals 138, representing the matching attributes in A and B. The M_{10} is 36, representing the number of nonmatching attributes in B, and M_{01} is 0, representing the nonmatching attributes in A. M_{00} is 64 with 00 matching attributes in A and B.

Jaccard and Cosine Similarity in Application

The same methodology was used in the application with the Jaccard equation (J) for an asymmetric binary similarity measurement between two data objects (A, B). Shown in the matrix (Table 1), the intersection M_{11} equals 138, representing the total matching attributes (i.e., indicators) in A (i.e., Alerts) and B (i.e., CTI). The M_{10} is 36, representing the number of nonmatching attributes in B, and M_{01} is 0, representing the nonmatching attributes in A. Again, M_{00} equals 64 because there are 00 matching attributes in both A and B. The number n represents the sum of the attributes is 238. In

the Jaccard index equation, the intersection (M_{11}) is 138 divided by the union ($M_{01} + M_{10} + M_{11}$) is 174, where J equals 0.793. Next, the cosine similarity was calculated using the same instrument, IBM SPSS, that was used to calculate the Jaccard index.

The results of the cosine similarity for the application with a value between -1 and 1 showed a similarity equaling 0.891. Since the cosine similarity of 0.891 is greater than 0.5, the relationship was significant between CTI and Alerts of critical infrastructure assets. In both calculations (Table 1), the Jaccard index between [0, 1], J = 0.316, and cosine similarity of 0.562, the results were insignificant with a value not greater than 0.5. In the following application portion, this CTI control variable was increased to 327.8K to find if there was any significant change in the Alerts (M_{11}) representing critical infrastructure assets on the network. When the CTI data increased from 8K to 328.7K IoCs, the number of matching Alert IoCs rose significantly from 55 to 138 (Table 1). This high Jaccard index of 0.793 and cosine similarity of 0.891 indicated substantial similarity between CTI data and Alert IoCs at the larger CTI data volume rate of 328.7K. These matching Alert IoCs provides early detection in a SIEM/SOAR application, Sentinel, integrated with the CTI TIP platform. Together with EDR or ETDR, it provides a comprehensive cybersecurity system in critical infrastructure, providing network protection (González-Granadillo et al., 2021).

Descriptive Statistics

The Jaccard index was used to quantify the similarity between alerts and CTI, which resulted in a meaningful correlation when applied to the experiment and application (Table 1). When the sample size was increased by 4,008% from 8K to 328.7K

during the experiment to the application phase, the dependent variable (Alert) increased by 151% from 55 to 138, matching more IoCs. The resulting Jaccard index in the application shows less similarity with the 8K CTI data size measurement of $J = 0.316$ compared to more similarity with the 328.7K CTI data size measurement of $J = 0.793$. A greater than the average of 0.5 similarity means a greater relationship between the objects (A, B). Since the Jaccard index is between 0 and 1, where $J > 0.5$ is considered statistically significant, a SOC analyst can describe a range of similarities or dissimilarities.

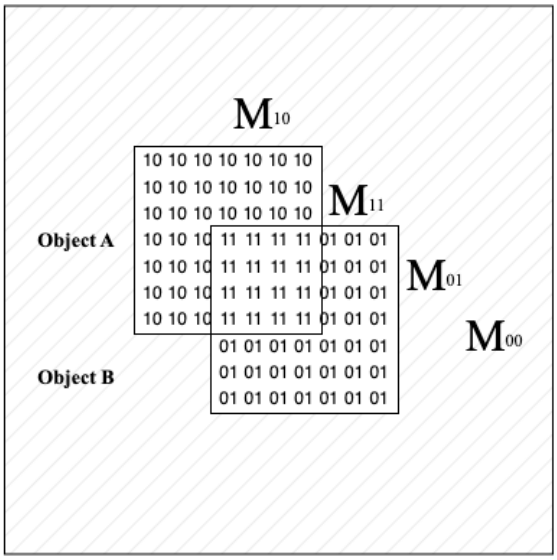
Jaccard Dissimilarity

The relative distance between the objects (A, B) can be used to find the relative dissimilarity. This Jaccard distance ($1 - J$) can be used to measure the dissimilarity between the two objects (Alerts, CTI). In this case, 1 - 0.316 equals 0.684 for the experiment and 1 - 0.793 equals 0.207 for the application. Even though this Jaccard distance value approaches 1 for dissimilar and 0 for similar, the Jaccard distance helps characterize a data set. The Jaccard index and distance are quantified between 0 and 1, with a similarity index in the experiment of 0.316 and a dissimilarity index of 0.684. A similarity index can be increased or decreased when the intersection (M_{11}) of the two objects (A, B) increases or decreases. As the number of indicators (M_{11}) that match between the two datasets (Alert, CTI) increases, the two objects become more similar (A, B). This visual representation of the Jaccard index model (Figure 2) is shown with the intersection (M_{11}) of indicator matches (1,1) and the union of objects (A, B) represented by the two squares of datasets, Alert (0,1) and CTI (1,0). Since no (0,0) matches

represented (M_{00}), the Jaccard distance is represented as the space outside the two squares.

Figure 2

Jaccard Index Model With the Intersection of Objects



Note. This data model represents the relationship between datasets using the Jaccard index matrix with emphasis on the intersection (1, 1). The union of objects (A, B) is represented by the two squares of datasets, Alert (0,1) and CTI (1,0), while the Jaccard distance is represented as the space outside the two squares.

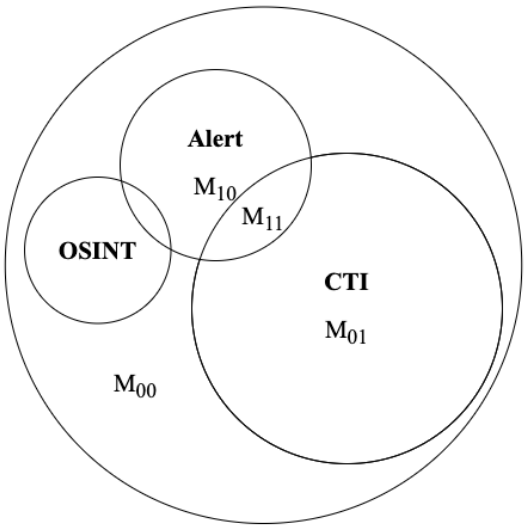
Evaluation of Statistical Assumptions

The convergence of the Jaccard distance is observed in the Jaccard matrix, where the similarity is greater in the experiment than in the application. As a result of this discovery, the elements that contributed to the quality of Alerts needed to be included in a

new data model to describe the relationship between OSINT, Alerts, and CTI. This representation, like a Venn diagram (Figure 3), shows the intersection between the three elements (OSINT, Alert, CTI) along with the Jaccard matrix attributes ($M_{01} + M_{10} + M_{11}$) and the space outside these elements with attribute (M_{00}; Verma & Aggarwal, 2020). This representation of these elements can inform about new outcomes with this method of incremental validity from real-world OSINT to Alerts forming indicators (i.e., predictors) to outcomes of cyberattacks from forensic evidence. According to Rohrer (2024), this causal inference of knowledge is improved by the knowledge of the outcome and the shared reality that is vital in measuring success. The assumption is that causal inference knowledge, where the intersection between two realities (OSINT and Forensic) should not be dismissed as a predictive measurement when rigorous reasoning is applied.

Figure 3

Jaccard Index Model With the Intersection of Objects



Note. This data model represents the relationship between datasets using the Jaccard index matrix. The nature of the data correlation between datasets is the intersection shown as a common link to the indicator of compromise (IoC) between Alert and CTI objects.

Reports of Inferential Statistical Analysis of the Experiment

A sample of the CTI dataset was dynamically collected (Appendix A), representing independent variables of interchangeable data feeds between November 05, 2023, and January 04, 2024 (60 days). A script to ingest CTI data automatically from the data source (AlienVault) into the cloud Microsoft Azure using the instrument Microsoft Logic Apps. This script (Appendix I) specified the daily feeds and the categories of IoC: domain, email, FileHash-MD5, FileHash-SHA1, FileHash-SHA256, IPv4, URL, and URI. A sample of total CTI indicators (Appendix E) collected is 8,938 IoC, although ten days recorded no data (11/13/2023, 11/27/2023, 12/03/2023, 12/18/2023, 12/19/2023, 12/24/2023, 12/26/2023, 01/01/2024, 01/02/2024, 01/03/2024) primarily during holidays. This experiment collected a small sample of data feeds to examine meaningful links between variables, CTI data feeds, and cyberattacks in critical infrastructure.

Experiment Variables

The dependent variable represents a static collection of cyberattacks meaningful to organizations in critical infrastructure (Zibak et al., 2022a). Using OSINT, Temple University collected victims of ransomware cyberattacks in various sectors of this industry from November 6, 2013, to November 27, 2023 (Rege, 2023). A sample of this

OSINT dataset from November 5 to November 27, 2023 (Appendix B), contained a total of 17 IoC data points of ransomware cyberattacks (Malik et al., 2023) filtered to eight different ransomware strains: RansomExx/defray, BlackCat/ALPHV, LockBit, Medusa, Snatch, Meow, NoEscape, Rhysida, Hunters International Ransomware Team. Based on these ransomware strains, a new alert dataset was created.

An alert dataset was created using a similar-based correlation link approach (Rohrer, 2024). IoCs were gathered based on ransomware attacks from the OSINT dataset (Appendix B) to make the alert dataset (Appendix C). This alert dataset mainly comprises IoCs from STIX packages manually web-scraped from CISA's Internet site of advisories. Other data sources of alerts, reports, and advisories include the FBI and AlienVault. This alert dataset contains a total of 448 IoC with ransomware strains quantified: RansomExx/defray (56), BlackCat/ALPHV (39), LockBit (18), Medusa (69), Snatch (80), Meow (18), NoEscape (54), Rhysida (83), Hunters International Ransomware Team (31).

Once these datasets (CTI, OSINT, Alert) were created, the research variables (independent, dependent) were analyzed. The analysis results show that the independent and dependent research variables correlated when the IoCs matched. The IoCs matched when these indicators were found in both the alert and CTI datasets. This relationship between CTI and OSINT datasets were described in a table of research variable correlation (Appendix D) within the similar-based links column. This column links the dataset Temple University (Source) to CISA (Source) with LockBit (similar-based link) and AlienVault (Source) with IoC (Similar-based link). Similar-based links show 448

correlations of ransomware indicators (IoC) matching CTI indicators (IoC) from the raw data feed (Appendix E).

A flowchart represents the data correlation for this CTI experiment (Appendix F). The hierarchical nature of datasets starts at the top with the ransomware dataset and then the process of creating the alert dataset generated from OSINT. The OSINT decision point shows the group/strain of ransomware gathered from the OSINT dataset and from OSINT sources (e.g., CISA) with IoCs, then adds this information to the alert dataset. The next decision point is whether the IoCs match the CTI dataset. If the IoC from the alert dataset matches the IoC from the CTI dataset, then quantify this IoC dataset. The IoC dataset results of matching IoCs from the alert dataset quantified two months of CTI feeds: RansomExx/defray (0), BlackCat/ALPHV (0), LockBit (14), Medusa (3), Snatch (14), Meow (0), NoEscape (6), Rhysida (14), Hunters International Ransomware Team (4).

Data Model

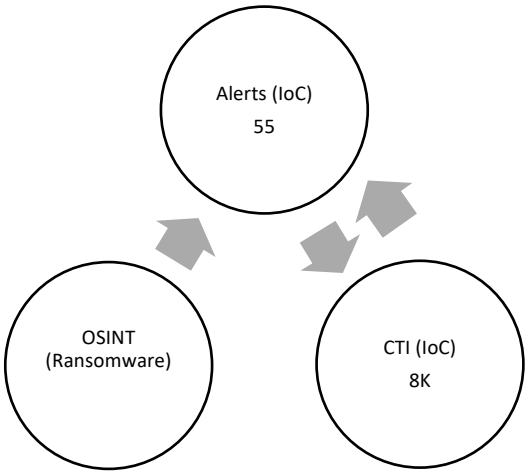
A data model of this relationship between datasets (CTI, Alerts, OSINT) was conceived (Figure 4). The data flow is one-way from the OSINT (Ransomware) to the Alerts (IoC) dataset, and it is two-way from Alerts (Ioc) to CTI (IoC). This relationship between datasets was an unexpected finding due to the nature of IoC curated from OSINT sources. OSINT sources played a significant role in finding similar-based links from data sources such as CISA ICS advisories. Some of these advisories of cyberattacks contained IoC information that helps defend against future attacks. This convergence of

intelligence of cyberattacks and defense impacts data quality quantifiable in matching indicators (IoC).

From the ransomware node with nine different indicator strains to the alert node with 55 IoCs and to the CTI node with 408 IoCs, three datasets (ransomware, alert, CTI) observed a nonlinear distribution of information (Figure 4). The one-way data flow from the ransomware to the alert node of indicators was filtered, and duplicate indicators were discarded. The two-way data flow between the alert and CTI nodes displayed an increase in IoC indicators from the alert node (55) to the CTI node (138). Since the CTI node is unfiltered, the same indicator (e.g., IPv4, file-hash, domain) could be submitted multiple times into the data feed from the data source (AlienVault), increasing the CTI node count. New indicators from the CTI node can be added to the alert node for a two-way data flow.

Figure 4

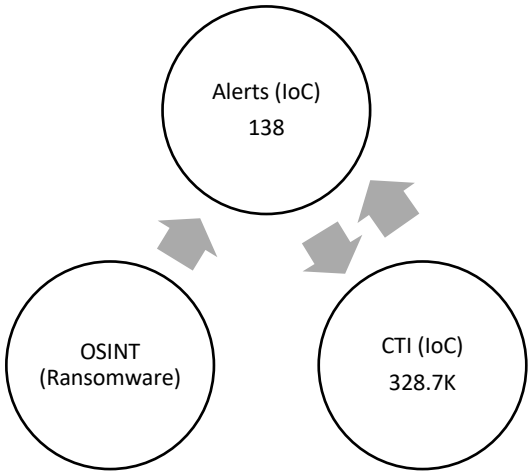
Data Model of Dataset Relationship in Experiment



Note. This data model represents the relationship between datasets and the data flow for this CTI experiment. The nature of the data flow between datasets has a common link to the indicator of compromise (IoC) in Alerts. 55 Alerts matched out of an 8K CTI dataset. The Jaccard index (J) of similarity between the two datasets (Alert, CTI) is 0.316.

Figure 5

Data Model of Dataset Relationship in Application



Note. This data model represents the relationship between datasets and the data flow for this CTI experiment. The nature of the data flow between datasets has a common link to the indicator of compromise (IoC) in Alerts. 138 Alerts matched out of a 328.7K dataset. The Jaccard index (J) of similarity between the two datasets (Alert, CTI) is 0.793.

Observed from November 5 to 27, 2023, a sample of this OSINT dataset was collected to correspond to the timeline of the alert dataset. In the OSINT dataset, meaningful categories related to the research question affected by cyberattacks include countries, organizations, and critical infrastructure sectors. During this period, quantified

meaningful to the organizations of ransomware cyberattacks include countries: France (1), USA (4), CA (1), OH (1), TN (1), TX (1), UT (1), MI (1), NJ (1), China (1), Australia (1), Canada (1), Japan (1). Other categories include critical infrastructure sectors and organizations: information technology (3), financial services (6), transportation systems (1), food and agriculture (1), government facilities (1), critical manufacturing (2), healthcare and public health (5), education facilities (1), energy (1), and organizations affected (17). The resulting ransomware paid during that month in their local currency is 16.6 million or approximately 16 million dollars (Rege, 2023).

Application

This experiment's data model was applied to simulate a real-world application in a SOC environment using the instrument Microsoft Sentinel (SIEM/SOAR). A different data collection mechanism was used for the construction of the SOC. Using the TAXII data connector, data feeds from data source AlienVault (Alaeifar et al., 2024) streamed CTI indicators near real-time between January 5 and March 3, 2024. Collecting more CTI data using this method increased the sample size significantly. When the sample size of the CTI indicators was increased from 8K to 328.7K, the number of IoC correlations rose from 55 to 138 due to a larger sample size (41x) and a more extended collection period from two to three months. Using the alert dataset, the resulting IoC matches using Sentinel TI quantified three months of ransomware strains: RansomExx/defray (10), BlackCat/ALPHV (4), LockBit (5), Medusa (1), Snatch (44), Meow (18), NoEscape (6), Rhysida (46), Hunters International Ransomware Team (4).

The application reflected upon the evaluated experiment applying the theoretical framework, Boyd's OODA (Ryder & Downs, 2022). This OODA framework evaluated the effectiveness of the process from observation of the CTI live data feeds to orienting these feeds where alerts were used to decide whether to act (Appendix G). The instrument used for this framework is a TIP (Sentinel) simulating a SOC environment. A data connector is used to ingest the data feed, a watchlist to alert on malicious IoCs, and a workspace to take mitigating action. On this platform, a mitigating action by a human or machine-in-the-loop can automate a workflow to take an action like a SIEM/SOAR (Copeland, 2021). The Jaccard similarity index can be used to support this decision to act.

A playbook can automate responses from alerts using customizable rules to act upon a decision to mitigate (Schlette et al., 2021b). An analyst can quickly identify threats from the CTI data ingestion data (Appendix H) and tag potential threats using threat management (e.g., TI, threat hunting, MITRE ATT&CK). Workflows (e.g., Workbooks, notebooks, workspace) can be constructed to take mitigating action to block malware on edge devices (e.g., firewall, IDS) using the alert dataset created in the experiment implemented with the watchlist and automation tool (Alaeifar et al., 2024). The OODA theoretical model with a feedback loop on an SIEM/SOAR application with TI management demonstrates the model's effectiveness.

Description of Convergence

The Jaccard index can quantify the convergence of indicators (Chen et al., 2024) with causal links to OSINT and similar-based links to Alerts and CTI. The taxonomy of

the data with similar-based semantics contributed to the Alert dataset that intersects OSINT and CTI (Schlette et al., 2021b). Quantifying the semantic alignment using the CTI TAXII format for quality indicators (IoCs) impacted the correlation (Appendix F). A data flow relationship model of these three datasets (OSINT, Alert, and CTI) is illustrated (Figure 1). The measure of validity includes a similar-based diagram using the Jaccard index and matrix (Appendix M). The analysis used instruments such as Microsoft Excel and IBM SPSS statistics tool, and data collection was automated using cloud-based instruments such as Microsoft Azure Logic Apps and Sentinel.

Analysis and Interpretation of the Findings

In the experiment, the causal link from the Alert dataset can be traced back to the OSINT dataset (Rege, 2023) to the type of ransomware strain. In this real-world OSINT dataset, the "Critical Infrastructure Ransomware Attacks (CIRA) Dataset" contained the ransomware strain BlackCat/ALPHV with the category MITRE ATT&CK containing S1068. S1068 is an identifier of the type of cyberattack in the MITRE ATT&CK framework. The MITRE ATT&CK website describes S1068 of the category software associated with ALPHV and type malware (Chen, Vaitheeshwari et al., 2024). Describing adversarial TTP provides actionable intelligence with behavioral analytics that helps identify a defensive gap in a SOC (Georgiadou et al., 2021). This correlation of cyberattacks impacts the analyst in the SOC because the risk is not assessing these defensive gaps in organizational assets (e.g., application security software, data security, privacy, hardware, information, network, asset management, personnel security, physical security). In the experiment, the similar-based correlation of cyberattack type

ransomware BlackCat/ALPHV is linked to real-world attacks on critical infrastructure. From the OSINT table (Appendix B), this attack occurred approximately on November 11, 2023, using malware software of ransomware strain, BlackCat/ALPHV, described by MITRE ATT&CK as S1068 affecting the critical infrastructure sector of Information Technology in the United States reported by Dragos corporation. In the Alert table (Appendix C), on April 22, 2022, my data source CISA/FBI reported in an FBI Flash CU-000167-MW that BlackCat/ALPV was analyzed and the IoC of file-hash type is d241df7b9d2ec0b8194751cd5ce153e27cc40fa4. The analyst validated this IoC using the method developed in the Dataset Correlation Flowchart (Appendix F), and the SOC cyber professional using the method developed in the OODA Model Flowchart (Appendix G). This IoC of type file hash was detected four times in the application data stream but not in the experiment dataset. This discrepancy may be due to the bandwidth of data streams and sampling of the data in the experiment (Logic Apps) compared to the live stream, which has a more considerable bandwidth in the experiment (Sentinel).

Summarizes Answers to Research Questions

The study's research problem question asks what the relationship is between CTI attack identification and critical infrastructure assets in IT critical infrastructure attacks. The result of the experiment and its application is a direct correlation between indicator (IoC) matches and CTI indicators that are meaningful to organizations in the critical infrastructure industry.

In the experiment, the intersection M_{11} is the number of matching indicators from objects A and B. Object A represents a pool of alerts, and object B is a pool of CTI

indicators. The nonmatching indicators are M_{10} and M_{01} (Alert and CTI). The number of nonmatching indicators from the Alert dataset M_{10} is 119, and from the CTI dataset M_{01} is 0. Any duplicates of the same attributes in the set were not counted again. The experiment shows a direct correlation of 55 matching indicators (M_{11}) from the two objects of the Jaccard index matrix (Table 2), representing the application with 138 matching indicators (M_{11}).

Table 2

Matrix of Jaccard Index

M_{11}	Matching Indicators	
	Experiment	Application
Ransomware		
BlackCat	0	4
LockBit	14	5
Medusa	3	1
Snatch	14	44
NoExcape	6	6
Rysida	14	46
Hunters	4	46
RansomExx	0	10
Meow	0	18
Total	55	138

Note. In the experiment, the intersection M_{11} represents the matching attributes in A and B. The total matching indicators between objects A and B in the experiment is 55, and in the application is 138.

The findings showed in the null Hypothesis (H0) that when the CTI attack identification and critical infrastructure assets have no significant relationship with IT critical infrastructure attacks, the value is 0 when using the Jaccard index. A Jaccard

index greater than 0.5 and less than or equal to 1 shows similarity between data sets (Li et al., 2024). The alternative Hypothesis (H1) is when the CTI attack identification and critical infrastructure assets have a relationship with IT critical infrastructure attacks when the Jaccard index is greater than 0.5. In the experiment, the Jaccard index had a value of 0.316; in the application, the value was 0.793. The index increased by 151% between the experiment and application is because the number of assets and network transactions increased to approximately 328.7K from 8.9K. This method aims to enhance the decision-making process by quantifying the effectiveness and efficiency of the security measures in the experiment and application. Compared to a qualitative approach, quantitative approaches aim to compare the results to certain aspects of the relationship between CTI and Alerts (Dekker & Alevizos, 2024). The Jaccard index was used to show similarity of the Alert data set and the experiment and application data set showing matches of vulnerability types.

Applications to Professional Practice

This last decade has seen an increase in the volume of data generated by different cyber security solutions for CTI (Choo et al., 2019). When collected with the voracity of volume, big data becomes increasingly difficult for traditional cyber investigation methods. For a cyber professional to investigate CTI streams over a large scale of threat de-identification to find the causality of the cyberattack from a single IoC data point is difficult. This problem of causal inference (Rohrer, 2024) of observational analysis, when applied to a smaller realm of real-world data, can predict outcomes (Verma & Aggarwal, 2020).

The process for monitoring and responding to security events at a critical infrastructure facility is time-critical. This process starts with suspicious circumstances triggering an alert (Li et al., 2024). The cyber professional investigates the alert to determine if it is credible and performs a triage. This triage process involves a lengthy examination of details of the incident, such as the traffic source IP, group of events, traffic protocol, payload information, file hashes, and attack vectors. The targeted asset and the geographic location data (i.e., IPv4, domain) are associated with the originating country/city. TTP (Kure & Islam, 2019) can assess the threat profile of the incident using related patterns associated with groups.

Constructing an experiment to simulate real-world CTI applications and transferring these techniques and procedures to the application was successful. A simulated SOC environment was possible using a SIEM/SOAR application hosted on a Microsoft Sentential cloud platform. What was transferred from the experiment using causal inference knowledge is the alerts. These alerts intersect two realities (OSINT and Forensic), resulting in an evidence-based dataset. When these elements combine (Appendix M) OSINT, Alerts, and CTI, a similar-based index between two objects (Alerts, CTI) can be used as a correlation measurement of indicator matches (Schaberreiter et al., 2019). Using this method, the Jaccard index, a statistical method (Verma & Aggarwal, 2020), can inform the SOC analyst about the incremental validity of the alerts. When rigorous reasoning is applied, this method can shorten the OODA loop (Ryder & Downs, 2022) to gain better performance in deciding and acting. This causal inference of knowledge (Rohrer, 2024) between real-world and CTI improves

knowledge, increasing the likelihood of success for a SOC analyst with high confidence that the alert data will be meaningful to their organization.

The application to professional practice is the capability to automate alerts and prevent the cyberattack before it happens using the OODA model flowchart (Appendix G). A watchlist of alerts derived from the experiments' dataset correlation flowchart (Appendix F) containing cyber IoC indicators (e.g., IPv4, domain, file hashes, CVE) is uploaded to the SIEM/SOAR application (Sentinel). These alerts are used to trigger a workflow of mitigating actions (Tuyishime et al., 2023), such as alerting a supervisor of a cyber incident and blocking malicious events by updating the firewall rules to block specific IPs or files from entering these edge devices (e.g., firewalls, IDS, IoT). These IoT edge devices may include industrial CPS devices (Radanliev et al., 2020) such as Siemens S7 controllers (Malik et al., 2023) monitored by the SIEM/SOAR architecture (Sentinel), wherein correlation rules can be applied with the alerts, automating a response to prevent critical infrastructure attacks.

Implications for Social Change

In 2018, there were more than two million cyberattacks of ransomware incidents, costing more than $45 billion worldwide in losses (González-Granadillo et al., 2021). In February 2020, a ransomware attack resulted in a two-day shutdown of the U.S. natural gas pipeline (González-Granadillo et al., 2021). In my research from November 5 to 27, 2023, seventeen critical infrastructure organizations worldwide were affected by ransomware attacks (Malik et al., 2023) in the sectors of information technology (3), financial services (6), transportation systems (1), food and agriculture (1), government

facilities (1), critical manufacturing (2), healthcare and public health (5), education facilities (1), and energy (1). The resulting ransomware paid during that month in their local currency is 16.6 million or approximately 16 million dollars (Rege, 2023).

Ransomware attacks can affect classical IT networks (e.g., enterprise), which focus on confidentiality and integrity, and critical infrastructure networks (e.g., SCADA), which focus on operations availability (Malik et al., 2023). Malicious entities can work laterally from corporate networks into critical infrastructure networks through gaps in network-segmented firewalls. While some cyber professionals cannot detect all corporate IT and industrial OT network attacks (Sen, Ö et al., 2022), one solution is to use an SIEM/SOAR application to collect large amounts of data to identify potential threats. Advancing these SIEM systems by aligning them with mitigation strategies using data analytics to perform correlations using TIPs helps cope with these challenges. The impact of integrating CTI into the SIEM/SOAR architecture provides actionable intelligence to organizations affected by cyberattacks. These improvements in cybersecurity may provide customers of organizations affected by critical infrastructure attacks with uninterrupted services such as electric, gas, water, financial, and medical.

Recommendations for Action

In my research, workflows (Appendix F, G) optimized the data collection of CTI in the experiment and the application. The recommendation for action is to automate using new technologies using cloud applications (e.g., Logic Apps, Sentinel) that permit CTI data ingestion directly from data sources (e.g., AlienVault). This data collection process impacts data quality by preventing disruptions in data flow. In the experiment, I

wrote scripts to automate the CTI ingestion process, providing attributes such as interval of collection times, data storage location, data type (IoC), and data format (STIX/TAXII). In the application, I used a data connector using a SIEM/SOAR platform (Sentinel) that ingested CTI information from the data server (AlienVault) using the STIX/TAXII (2.0 or 2.1) format.

By automating the data collection, the SOC analyst can focus on important analytic tasks, such as identifying which CTI's threat indicators correlate to potential cyberattacks that affect the organization (Schlette et al., 2021c). The SOC analyst can focus on the decide and action portion of the OODA model (Ryder & Downs, 2022) and apply the attack vector knowledge learned from the experiment. This knowledge in the form of alerts is uploaded to the SIEM/SOAR application of Microsoft Sentinel (Tuyishime et al., 2023) using the Jaccard index to quantify two sets of data objects (CTI, Alerts) (Schaberreiter et al., 2019). The SOC analyst ranking the confidence level creates workflows triggered by the alerts and scripted to take mitigating action, allowing some automation in the event of a cyberattack.

In early May of 2021, the resulting cyberattack on the U.S. critical infrastructure shut down the gas fuel supply to the East Coast for several days, affecting millions of people (Smith, 2021). According to the U.S. Department of Homeland Security (DHS), analysts and practitioners face new challenges in a rapidly shifting threat landscape to keep up with cyberattacks from state and nonstate actors (Black et al., 2024). To better prepare for these high-tech threats through education and training, DHS identified six core competencies that every DHS intelligence analyst should have for counter-terrorism,

and one of those competencies is a greater understanding of technology. According to Black (2024), leadership needs to consider integrating these new technologies that can help optimize intelligence flows, automate mundane tasks, and augment analysts by automating some types of analysis.

The results of this study will be disseminated via literature, training, and conferences. The literature via publication through Walden University, training through educational institutions, papers at conferences, and research grants.

Recommendations for Further Study

In the past decade, big data has increased in volume, generating different cybersecurity solutions such as TI and big-data analytics techniques (Choo et al., 2019). Although technology plays a crucial role in cybersecurity with applications in big data, statistical methods, and ML in the SOC, humans are the key players in their roles of attacker and defender. This knowledge gap in cognitive security by the security analyst must focus on the organization during the detection and response of cybersecurity incidents (Andrade et al., 2022). During the experiment, there was a need for pre-processing small data. When the alerts were uploaded to the SIEM/SOAR application of Microsoft Sentinel (Tuyishime et al., 2023) using the Jaccard index to quantify the similarity between two sets of data objects (CTI, Alerts), I saw the value of characterization of the data.

Future work will address this OSINT process of generating alerts in cybersecurity with applications in big data, statistical methods, and AI techniques (Browne et al., 2024). This CTI system extracts descriptive or static CTI data from unstructured text,

determines their semantic relationships gathering IoC, and identifies their target critical infrastructure sectors. A framework (e.g., Vulcan) extracts unstructured text and analyzes CTI indicators into categories meaningful to organizations (Jo et al., 2022). While AI techniques such as ML are used to learn the contextual information of words, they require a large amount of labeled information for deep neural networks. A better approach is large language models (LLMs) because the broad spectrum of CTI indicators can assist the SOC analyst in the correlation between CTI and critical infrastructure assets, such as a classifier of alerts (Bayer et al., 2023). These generative AI models using LLMs can be the next step for cybersecurity using pre-trained models to answer meaningful questions like ChatGPT (Amine et al., 2024). A method for automated extraction of TI using NLP technology to identify cyberattack groups, tools, locations, affected organizations, industries, and sectors and a fine-tuned language model using a bidirectional encoder representation from transformers (BERT) technique to extract reports (Ren et al., 2023) and advisories to identify relevant alerts in TI is future work.

Reflections

The experience gained from this study and the progress made was challenging during the data collection phase of the dissertation. During this phase, I assumed the dataset would align with the Pearson correlation statistical model, but I did not realize that this model required linear regression of two datasets. Determining which statistical model to use before collecting the data was difficult. This preconceived notion was that the two datasets (alerts, CTI) would follow a linear regression on pattern-based advisories, and CTI feeds did not match accordingly. Instead, the pattern reflected data

fluctuations like big data. After this discovery, my study shifted from the pattern's nature to the real-world effects of cyberattacks and the indicators (IoC). From this perspective, the impact on data collection improved by aligning the taxonomy and semantics across various datasets (OSINT, alerts, CTI), describing the nature of data using the Jaccard index (Hamroun et al., 2024), a statistical model for big data and similar-based links described the data environment and the relationship between datasets.

The alert dataset was the most labor-intensive and complex, gathering OSINT and CTI indicators from various institutions. Once this alert dataset was created using the Dataset Correlation Flowchart (Appendix F), the OSINT process of collection processing, exploitation, and production (Browne et al., 2024) was easily transitioned from experiment to application. Upon reflection of the theoretical OODA in the application, the TIP (Sentinel) and application (SIEM/SOAR) fulfilled all elements (Observe, Orient, Decide, Act) with significant efficacy. In a SOC environment, the cyber operator can focus on the decision and act elements of the OODA Model Flowchart (Appendix G), which is meaningful for automating tasks to prevent cyberattacks.

Summary and Study Conclusions

The study's results correlated CTI's attack identification with assets meaningful to critical infrastructure organizations. In a SOC environment using the OODA theoretical framework, the CTI data feeds are monitored and oriented by selecting indicator (IoC) attributes. With high confidence, some SOC analysts can use these OSINT methods to create alerts and decide whether to act in a cybersecurity attack. The impact of integrating CTI into the SIEM/SOAR architecture is that it improves the cybersecurity process with

active intelligence and mitigating responses. Automating some OODA processes, such as deciding and acting with fast-response mitigating action, may prevent cyberattacks. Preventing these cyberattacks with actionable CTI intelligence affects customers with uninterrupted electric, gas, water, financial, and medical services.

References

Abbasnasab Sardareh, S., Brown, G. T., & Denny, P. (2021). Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. *Teaching Statistics, 43*, S157-S172. https://doi.org/10.1111/test.12274

Admass, W. S., Munaye, Y. Y., & Diro, A. (2023). Cyber security: State of the art, challenges and future directions. *Cyber Security and Applications*, 100031. https://doi.org/10.1016/j.csa.2023.100031

Agyepong, E., Cherdantseva, Y., Reinecke, P., & Burnap, P. (2020). Challenges and performance metrics for security operations center analysts: a systematic review. *Journal of Cyber Security Technology, 4*(3), 125-152. https://doi.org/10.1080/23742917.2019.1698178

Alaeifar, P., Pal, S., Jadidi, Z., Hussain, M., & Foo, E. (2024). Current approaches and future directions for Cyber Threat Intelligence sharing: A survey. *Journal of Information Security and Applications, 83*, 103786. https://doi.org/10.1016/j.jisa.2024.103786

Albasheer, H., Md Siraj, M., Mubarakali, A., Elsier Tayfour, O., Salih, S., Hamdan, M., & Kamarudeen, S. (2022). Cyber-attack prediction based on network intrusion detection systems for alert correlation techniques: a survey. *Sensors, 22*(4), 1494. https://doi.org/10.3390/s22041494

Aldauiji, F., Batarfi, O., & Bayousef, M. (2022). Utilizing cyber threat hunting techniques to find ransomware attacks: A survey of the state of the art. *IEEE*

Access, 10, 61695-61706. https://doi.org/10.1109/access.2022.3181278

Aldhaheri, A., Alwahedi, F., Ferrag, M. A., & Battah, A. (2023). Deep learning for cyber threat detection in IoT networks: A review. *Internet of Things and Cyber-Physical Systems*. https://doi.org/10.1016/j.iotcps.2023.09.003

AlQadheeb, A., Bhattacharyya, S., & Perl, S. (2022). Enhancing cybersecurity by generating user-specific security policy through the formal modeling of user behavior. *Array, 14*. https://doi.org/10.1016/j.array.2022.100146

Alves, F., Bettini, A., Ferreira, P. M., & Bessani, A. (2021). Processing tweets for cybersecurity threat awareness. *Information Systems, 95*, Article 101586. https://doi.org/10.1016/j.is.2020.101586

Amine Ferrag, M., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., & Tihanyi, N. (2024). Generative AI and large language models for cyber security: All insights you need. *arXiv e-prints, arXiv-2405*. https://doi.org/10.2139/ssrn.4853709

Ammi, M., Adedugbe, O., & Benkhelifa, E. (2022). Taxonomical challenges for cyber incident response threat intelligence: A review. *International Journal of Cloud Applications and Computing, 12*(1), 1-14. https://doi.org/10.4018/IJCAC.300770

Andrade, R. O., Fuertes, W., Cazares, M., Ortiz-Garcés, I., & Navas, G. (2022). An exploratory study of cognitive sciences applied to cybersecurity. *Electronics, 11*(11), 1692. https://doi.org/10.3390/electronics11111692

Andrade, R. O., & Yoo, S. G. (2019). Cognitive security: A comprehensive study of cognitive science in cybersecurity. *Journal of Information Security and Applications, 48*, Article 102352. https://doi.org/10.1016/j.jisa.2019.06.008

Ashley, T. D., Kwon, R., Gourisetti, S. N. G., Katsis, C., Bonebrake, C. A., & Boyd, P. A. (2022). Gamification of cybersecurity for workforce development in critical infrastructure. *IEEE Access, 10*, Article 112487. https://doi.org/10.1109/ACCESS.2022.3216711

Atkins, S., & Lawson, C. (2021). An improvised patchwork: Success and failure in cybersecurity policy for critical infrastructure. *Public Administration Review, 81*(5), 847-861. https://doi.org/10.1111/puar.13322

Basheer, R., & Alkhatib, B. (2021). Threats from the dark: a review over dark web investigation research for cyber threat intelligence. *Journal of Computer Networks and Communications, 2021*(1), 1302999. https://doi.org/10.1155/2021/1302999

Bayer, M., Frey, T., & Reuter, C. (2023). Multi-level fine-tuning, data augmentation, and few-shot learning for specialized cyber threat intelligence. *Computers & Security, 134*, Article 103430. https://doi.org/10.1016/j.cose.2023.103430

Benyahya, M., Collen, A., Kechagia, S., & Nijdam, N. A. (2022). Automated city shuttles: Mapping the key challenges in cybersecurity, privacy and standards to future developments. *Computers & Security, 122*. https://doi.org/10.1016/j.cose.2022.102904

Berahmand, K., Haghani, S., Rostami, M., & Li, Y. (2022). A new attributed graph clustering by using label propagation in complex networks. *Journal of King Saud University-Computer and Information Sciences, 34*(5), 1869-1883. https://doi.org/10.1016/j.jksuci.2020.08.013

Black, M., Obradovic, L., & House, D. (2024). Behind the curve: Technology challenges

facing the homeland intelligence and counterterrorism workforce. *Journal of Cybersecurity, 10*(1), Article tyae002. https://doi.org/10.1093/cybsec/tyae002

Borges Amaro, L. J., Percilio Azevedo, B. W., Lopes de Mendonca, F. L., Giozza, W. F., Albuquerque, R. D. O., & García Villalba, L. J. (2022). Methodological framework to collect, process, analyze and visualize cyber threat intelligence data. *Applied Sciences, 12*(3), Article 1205. https://doi.org/10.3390/app12031205

Boyd, J. R. (1987). *Destruction and creation*. U.S. Army Command and General Staff College. https://goalsys.com/books/documents/DESTRUCTION_AND_CREATION.pdf

Browne, T. O., Abedin, M., & Chowdhury, M. J. M. (2024). A systematic review on research utilizing artificial intelligence for open source intelligence (OSINT) applications. *International Journal of Information Security*. https://doi.org/10.1007/s10207-024-0086

Bryan, L. L. (2020). Effective information security strategies for small business. *International Journal of Cyber Criminology, 14*(1), 341-360. https://doi.org/10.5281/zenodo.3760328

Busquim e Silva, R. A., Piqueira, J. R. C., Cruz, J. J., & Marques, R. P. (2021). Cybersecurity assessment framework for digital interface between safety and security at nuclear power plants. *International Journal of Critical Infrastructure Protection, 34*. https://doi.org/10.1016/j.ijcip.2021.100453

Cascavilla, G., Tamburri, D. A., & Van Den Heuvel, W. J. (2021). Cybercrime threat intelligence: A systematic multi-vocal literature review. *Computers &*

Security, 105, 102258. https://doi.org/10.1016/j.cose.2021.102258

Center for Threat Informed Defense. (n.d.). *Microsoft Azure Security Control Mappings to MITRE ATT&CK*. https://center-for-threat-informed-defense.github.io/security-stack-mappings/Azure/README.html#controls

Chen, S. S., Hwang, R. H., Ali, A., Lin, Y. D., Wei, Y. C., & Pai, T. W. (2024). Improving quality of indicators of compromise using STIX graphs. *Computers & Security, 103972*. https://doi.org/10.1016/j.cose.2024.103972

Chen, Z. S., Vaitheeshwari, R., Wu, E. H. K., Lin, Y. D., Hwang, R. H., Lin, P. C., & Ali, A. (2024). Clustering APT Groups through Cyber Threat Intelligence by Weighted Similarity Measurement. *IEEE Access*. https://doi.org/10.1109/access.2024.3469552

Choo, K. K. R., Conti, M., & Dehghantanha, A. (2019). Special issue on big data applications in cyber security and threat intelligence–part 1. *IEEE Transactions on Big Data, 5*(3), 279-281. https://doi.org/10.1109/tbdata.2019.2933039

Chowdhury, N. H., Adam, M. T., & Teubner, T. (2020). Time pressure in human cybersecurity behavior: Theoretical framework and countermeasures. *Computers & Security, 97*, 101931. https://doi.org/10.1016/j.cose.2020.101931

Cifranic, N., Hallman, R. A., Romero-Mariona, J., Souza, B., Calton, T., & Coca, G. (2020). Decepti-SCADA: A cyber deception framework for active defense of networked critical infrastructures. *Internet of Things, 12*. https://doi.org/10.1016/j.iot.2020.100320

Copeland, M. (2021). Threat Intelligence. In *Cloud Defense Strategies with Azure*

Sentinel (pp. 145-161). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-7132-2_5

Corallo, A., Lazoi, M., & Lezzi, M. (2020). Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts. *Computers in Industry, 114*. https://doi.org/10.1016/j.compind.2019.103165

Cybersecurity and Infrastructure Security Agency (CISA). (2021). Alert (AA21-131A) *DarkSide Ransomware: Best Practices for Preventing Business Disruption from Ransomware Attacks*. https://us-cert.cisa.gov/ncas/alerts/aa21-1'1a

D'Amico, A., Whitley, K., Tesone,'D., O'Brien, B., & Roth, E. (2005). Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 229-233. https://doi.org/10.1177/154193120504900304

Dawson, M., Bacius, R., Gouveia, L. B., & Vassilakos, A. (2021). Understanding the challenge of cybersecurity in critical infrastructure sectors. *Revista Academiei Fortelor Terestre, 26*(1), 69–75. https://doi.org/10.2478/raft-2021-0011

Dekker, M., & Alevizos, L. (2024). A threat-intelligence driven methodology to incorporate uncertainty in cyber risk analysis and enhance decision-making. *Security and Privacy, 7*(1), e333. https://doi.org/10.1002/spy2.333

Department of Health, Education, and Welfare. (1979). *The Belmont Report*. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html

Dykstra, J., Gordon, L. A., Loeb, M. P., & Zhou, L. (2023). Maximizing the benefits

from sharing cyber threat intelligence by government agencies and departments. *Journal of Cybersecurity, 9*(1), tyad003. https://doi.org/10.1093/cybsec/tyad003

Fang, Y., Gao, J., Liu, Z., & Huang, C. (2020). Detecting cyber threat event from twitter using IDCNN and BiLSTM. *Applied Sciences, 10*(17), 5922. https://doi.org/10.3390/app10175922

Fife, D. (2022). Flexplot: Graphically-based data analysis. *Psychological Methods, 27*(4), 477–496. https://doi.org/10.1037/met0000424

Gao, Y., Li, X., Peng, H., Fang, B., & Philip, S. Y. (2020). Hincti: A cyber threat intelligence modeling and identification system based on heterogeneous information network. *IEEE Transactions on Knowledge and Data Engineering*, *34*(2), 708-722. https://doi.org/10.1109/tkde.2020.2987019

Gage, N.L. (1963). *Handbook of Research on Teaching*. (pp. 1-76) Chicago: Rand McNally.

Georgiadou, A., Mouzakitis, S., & Askounis, D. (2021). Assessing Mitre Att&ck risk using a cyber-security culture framework. *Sensors, 21*(9), 3267. http://dx.doi.org/10.3390/s21093267

González-Granadillo, G., González-Zarzosa, S., & Diaz, R. (2021). Security information and event management (SIEM): analysis, trends, and usage in critical infrastructures. *Sensors, 21*(14), 4759. http://dx.doi.org/10.3390/s21144759

Hamroun, C., Amamou, A., Haddadou, K., Haroun, H., & Pujolle, G. (2024). A review on lexical based malicious domain name detection methods. *Annals of*

Telecommunications, 1-17. https://doi.org/10.1109/csnet56116.2022.9955618

HHS.gov. (n.d.). *45 CFR 46*. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html#:~:text=Subpart%20A%2C%20also%20known%20as,provides%20requirements%20for%20IRB%20registration.

Huang, L., Zhao, G., Li, L., & Zhang, F. (2022). Host identification based on self-similarity of network activity. *Computer Communications, 191*, 467–476. https://doi.org/10.1016/j.comcom.2022.05.017

Husák, M., Sadlek, L., Špaček, S., Laštovička, M., Javorník, M., & Komárková, J. (2022). CRUSOE: A toolset for cyber situational awareness and decision support in incident handling. *Computers & Security, 115*, 102609. https://doi.org/10.1016/j.cose.2022.102609

IBM. (2022). *X-Force Threat Intelligence Index 2022*. https://www.ibm.com/downloads/cas/ADLMYLAZ

Ibrahim, A., Thiruvady, D., Schneider, J. G., & Abdelrazek, M. (2020). The challenges of leveraging threat intelligence to stop data breaches. *Frontiers in Computer Science, 2*, 36. https://doi.org/10.3389/fcomp.2020.00036

Irshad, E., & Siddiqui, A. B. (2023). Cyber threat attribution using unstructured reports in cyber threat intelligence. *Egyptian Informatics Journal, 24*(1), 43-59. https://doi.org/10.1016/j.eij.2022.11.001

Jabar, T., & Mahinderjit Singh, M. (2022). Exploration of mobile device behavior for mitigating advanced persistent threats (APT): a systematic literature review and

conceptual framework. *Sensors, 22*(13), 4662. https://doi.org/10.3390/s22134662

Jeyaraj, A., Zadeh, A., & Sethi, V. (2021). Cybersecurity threats and organisational response: Textual analysis and panel regression. *Journal of Business Analytics, 4*(1), 26–39. https://doi.org/10.1080/2573234X.2020.1863750

Jo, H., Lee, Y., & Shin, S. (2022). Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text. *Computers & Security, 120*. https://doi.org/10.1016/j.cose.2022.102763

Kamiura, M., & Sekine, R. (2023). Jaccard matrix for nonlinear filter statistics. SICE *Journal of Control, Measurement, and System Integration, 16*(1), 152-163. https://doi.org/10.1080/18824889.2023.2194169

Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion, 97*, 101804. https://doi.org/10.1016/j.inffus.2023.101804

Kayan, H., Nunes, M., Rana, O., Burnap, P., & Perera, C. (2022). Cybersecurity of industrial cyber-physical systems: A Review. *ACM Computing Surveys, 54*(11s), 1–35. https://doi.org/10.1145/3510410

Koeneman, S. H., & Cavanaugh, J. E. (2022). An improved asymptotic test for the Jaccard similarity index for binary data. *Statistics & probability letters, 184*, 109375. https://doi.org/10.1016/j.spl.2022.109375

Kong, J., Zhang, C., & Simonovic, S. P. (2023). Resilience and risk-based restoration strategies for critical infrastructure under uncertain disaster scenarios. *Sustainable Cities and Society, 92*, 104510. https://doi.org/10.1016/j.scs.2023.104510

Kotenko, I., Gaifulina, D., & Zelichenok, I. (2022). Systematic literature review of security event correlation methods. *IEEE Access, 10*, 43387-43420. https://doi.org/10.1109/access.2022.3168976

Kotecha, K., Verma, R., Rao, P. V., Prasad, P., Mishra, V. K., Badal, T., Jain, D., Garg, D., & Sharma, S. (2021). Enhanced network intrusion detection system. *Sensors (Basel, Switzerland), 21*(23). https://doi.org/10.3390/s21237835

Kotsias, J., Ahmad, A., & Scheepers, R. (2023). Adopting and integrating cyber-threat intelligence in a commercial organisation. *European Journal of Information Systems, 32*(1), 35-51. https://doi.org/10.1080/0960085X.2022.2088414

Kraemer, R. A. S., Dias, D. P., da Silva, A. C., Martins, M. A. I., & Ludwig, M. A. (2022). Cost and cybersecurity challenges in the commissioning of microgrids in critical infrastructure: COGE Case Study. *Energies* (19961073), 15(8), 2860. https://doi.org/10.3390/en15082860

Kure, H., & Islam, S. (2019). Cyber threat intelligence for improving cybersecurity and risk management in critical infrastructure. *Journal of Universal Computer Science, 25*(11), 1478-1502. https://doi.org/10.3217/jucs-025-11-1478

Li, J., Liu, J., & Zhang, R. (2024). Advanced persistent threat group correlation analysis via attack behavior patterns and rough sets. *Electronics, 13*(6), 1106. https://doi.org/10.3390/electronics13061106

Liu, J., Yan, J., Jiang, J., He, Y., Wang, X., Jiang, Z., ... & Li, N. (2022). TriCTI: an actionable cyber threat intelligence discovery system via trigger-enhanced neural network. *Cybersecurity, 5*(1), 1-16. http://dx.doi.org/10.1186/s42400-022-00110-

3

Lu, J. (2020). Data analytics research-informed teaching in a digital technologies curriculum. *INFORMS Transactions on Education, 20*(2), 57-72. http://dx.doi.org/10.1287/ited.2019.0215

Lu, P., Hu, T., Wang, H., Zhang, R., & Wu, G. (2021). G-CAS: Greedy algorithm-based security event correlation system for critical infrastructure network. *Security and Communication Networks*, 2021. http://dx.doi.org/10.1155/2021/3566360

Maccarone, L. T., & Cole, D. G. (2022). Bayesian games for the cybersecurity of nuclear power plants. *International Journal of Critical Infrastructure Protection, 37*. https://doi.org/10.1016/j.ijcip.2021.100493

Malik, M. I., Ibrahim, A., Hannay, P., & Sikos, L. F. (2023). Developing resilient cyber-physical systems: a review of state-of-the-art malware detection approaches, gaps, and future directions. *Computers, 12*(4), 79. https://doi.org/10.3390/computers12040079

Maglaras, L., Janicke, H., & Ferrag, M. A. (2022). Cybersecurity of critical infrastructures: Challenges and solutions. *Sensors (14248220), 22*(14), N.PAG. https://doi.org/10.3390/s22145105

Makrakis, G. M., Kolias, C., Kambourakis, G., Rieger, C., & Benjamin, J. (2021). Industrial and critical infrastructure security: Technical analysis of real-life security incidents. *IEEE Access, Access, IEEE, 9*, 165295–165325. https://doi.org/10.1109/ACCESS.2021.3133348

Malatji, M., Marnewick, A. L., & Von Solms, S. (2021). Cybersecurity capabilities for

critical infrastructure resilience. *Information & Computer Security, 30*(2), 255–279. https://doi.org/10.1108/ICS-06-2021-0091

Marinho, R., & Holanda, R. (2023). Automated emerging cyber threat identification and profiling based on natural language processing. *IEEE Access, 11*, 58915-58936. https://doi.org/10.1109/access.2023.3260020

Martineau, M., Spiridon, E., & Aiken, M. (2023). A comprehensive framework for cyber behavioral analysis based on a systematic review of cyber profiling literature. *Forensic Sciences, 3*(3), 452-477. https://doi.org/10.3390/forensicsci3030032

Meland, P. H., Tokas, S., Erdogan, G., Bernsmed, K., & Omerovic, A. (2021). A systematic mapping study on cyber security indicator data. *Electronics*, *10*(9), 1092. http://dx.doi.org/10.3390/electronics10091092

Miranda-Calle, J. D., Reddy C., V., Dhawan, P., & Churi, P. (2021). Exploratory data analysis for cybersecurity. *World Journal of Engineering, 18*(5), 734–749. https://doi.org/10.1108/WJE-11-2020-0560

Mkuzangwe, N. N., & Khan, Z. C. (2020). Cyber-threat information-sharing standards: A review of evaluation literature. *The African Journal of Information and Communication, 25*, 1-12. https://doi.org/10.23962/10539/19784

Mohajan, H. K. (2020). Quantitative research: A successful investigation in natural and social sciences. *Journal of Economic Development, Environment and People, 9*(4), 50-79. https://doi.org/10.26458/jedep.v9i4.679

Moraliyage, H., Sumanasena, V., De Silva, D., Nawaratne, R., Sun, L., & Alahakoon, D. (2022). Multimodal classification of snion services for proactive cyber threat

intelligence using explainable deep learning. *IEEE Access, Access, IEEE, 10*, 56044–56056. https://doi.org/10.1109/ACCESS.2022.3176965

Mottahedi, A., Sereshki, F., Ataei, M., Nouri Qarahasanlou, A., & Barabadi, A. (2021). The resilience of critical infrastructure systems: A systematic literature review. *Energies, 14*(6), 1571. https://doi.org/10.3390/en14061571

Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., & Kim, J. (2021). COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE Transactions on Computational Social Systems, 8*(4), 1003-1015. https://doi.org/10.1109/TCSS.2021.3051189

National Infrastructure Advisory Council. (2009). *Critical Infrastructure Resilience Final Report and Recommendations*, U.S. Department of Homeland Security, Washington, DC, 2009.

National Institute of Standards and Technology (NIST). (2015, May). *SP 800-82 Rev. 2 Guide to Industrial Control Systems (ICS) Security.* https://csrc.nist.gov/publications/detail/sp/800-82/rev-2/final

Noor, U., Anwar, Z., Malik, A. W., Khan, S., & Saleem, S. (2019). A machine learning framework for investigating data breaches based on semantic analysis of adversary's attack patterns in threat intelligence repositories. *Future Generation Computer Systems, 95*, 467–487. https://doi.org/10.1016/j.future.2019.01.022

Nweke, L. O., & Wolthusen, S. (2020). Legal issues related to cyber threat information sharing among private entities for critical infrastructure protection. In *2020 12th International Conference on Cyber Conflict (CyCon)* (Vol. 1300, pp. 63-78).

IEEE. https://doi.org/10.23919/CyCon49761.2020.9131721

Petit, F., Bassett, G. W., Buehring, W. A., Collins, M. J., Dickinson, D. C., Haffenden, R. A., Peerenboom, J. P. (2013). *Protective Measures Index and Vulnerability Index: indicators of critical infrastructure protection and vulnerability* (No. ANL/DIS-13/04). Argonne National Lab.(ANL), Argonne, IL (United States).

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Radanliev, P., De Roure, D., Page, K., Nurse, J. R., Mantilla Montalvo, R., Santos, O., ... & Burnap, P. (2020). Cyber risk at the edge: current and future trends on cyber risk analytics and artificial intelligence in the industrial internet of things and industry 4.0 supply chains. *Cybersecurity, 3*, 1-21. https://doi.org/10.1186/s42400-020-00052-8

Radoglou-Grammatikis, P., Sarigiannidis, P., Iturbe, E., Rios, E., Martinez, S., Sarigiannidis, A., Eftathopoulos, G., Spyridis, Y., Sesis, A., Vakakis, N., Tzovaras, D., Kafetzakis, E., Giannoulakis, I., Tzifas, M., Giannakoulias, A., Angelopoulos, M., & Ramos, F. (2021). SPEAR SIEM: A security information and event management system for the smart grid. *Computer Networks, 193*. https://doi.org/10.1016/j.comnet.2021.108008

Rahman, M. R., Hezaveh, R. M., & Williams, L. (2023). What are the attackers doing now? Automating cyberthreat intelligence extraction from text on pace with the changing threat landscape: A survey. *ACM Computing Surveys, 55*(12), 1-36.

https://doi.org/10.1145/3571726

Rakas, S. V. B., Stojanović, M. D., & Marković-Petrović, J. D. (2020). A review of research work on network-based scada intrusion detection systems. *IEEE Access, 8*, 93083-93108. https://doi.org/10.1109/access.2020.2994961

Ramsdale, A., Shiaeles, S., & Kolokotronis, N. (2020). A comparative analysis of cyber-threat intelligence sources, formats and languages. *Electronics, 9*(5), 824. http://dx.doi.org/10.3390/electronics9050824

Rantos, K., Spyros, A., Papanikolaou, A., Kritsas, A., Ilioudis, C., & Katos, V. (2020). Interoperability challenges in the cybersecurity information sharing ecosystem. *Computers, 9*(1), 18. https://doi.org/10.3390/computers9010018

Rawat, D. B., Doku, R., & Garuba, M. (2021). Cybersecurity in big data era: From securing big data to data-driven security. *IEEE Transactions Services Computing 14* (6), 2055–2072. https://doi.org/10.1109/TSC.2019.2907247.

Rege, A. (2023). "Critical Infrastructure Ransomware Attacks (CIRA) Dataset". Version 12.9. *Temple University*. Online at https://sites.temple.edu/care/cira/. Funded by National Science Foundation CAREER Award #1453040. ORCID: 0000-0002-6396-1066.

Rehak, D., Markuci, J., Hromada, M., & Barcova, K. (2016). Quantitative evaluation of the synergistic effects of failures in a critical infrastructure system. *International Journal of Critical Infrastructure Protection, 14*, 3-17. http://dx.doi.org/10.1016/j.ijcip.2016.06.002

Rehak, D., Senovsky, P., Hromada, M., & Lovecek, T. (2019). Complex approach to

assessing the resilience of critical infrastructure elements. *International Journal of Critical Infrastructure Protection, 25*, 125-138. http://dx.doi.org/10.1016/j.ijcip.2019.03.003

Rehak, D., Slivkova, S., Janeckova, H., Stuberova, D., & Hromada, M. (2022). Strengthening resilience in the energy critical infrastructure: methodological overview. Energies, 15(14), 5276. https://doi.org/10.3390/en15145276

Ren, F., Jiang, Z., Wang, X., & Liu, J. (2020). A DGA domain names detection modeling method based on integrating an attention mechanism and deep neural network. *Cybersecurity, 3*(1), 1-13. http://dx.doi.org/10.1186/s42400-020-00046-6

Ren, Y., Xiao, Y., Zhou, Y., Zhang, Z., & Tian, Z. (2023). CSKG4APT: A Cybersecurity Knowledge Graph for Advanced Persistent Threat Organization Attribution. *IEEE Transactions on Knowledge & Data Engineering, 35(06), 5695-5709*. https://doi.org/10.1109/tkde.2022.3175719

Rendall, K., Mylonas, A., & Vidalis, S. (2022). Toward situational awareness in threat detection. A survey. Wiley Interdisciplinary Reviews: *Forensic Science, 4*(4), e1448. https://doi.org/10.1002/wfs2.1448

Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., ... & Sarwat, A. I. (2023). Impact, vulnerabilities, and mitigation strategies for cyber-secure critical infrastructure. *Sensors, 23*(8), 4060. https://doi.org/10.3390/s23084060

Roemsri, P., & Hewett, R. (2021, October). Device identification for iot security. In *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)* (pp. 866-870). IEEE. https://doi.org/10.1109/icsip52628.2021.9688598

Rohrer, J. M. (2024). Causal inference for psychologists who think that causal inference is not for them. *Social and Personality Psychology Compass*, 18(3), e12948. https://doi.org/10.1111/spc3.12948

Rosado, D. G., Santos-Olmo, A., Sánchez, L. E., Serrano, M. A., Blanco, C., Mouratidis, H., & Fernández-Medina, E. (2022). Managing cybersecurity risks of cyber-physical systems: The MARISMA-CPS pattern. *Computers in Industry, 142.* https://doi.org/10.1016/j.compind.2022.103715

Ryder, M., & Downs, C. (2022). Rethinking reflective practice: John Boyd's OODA loop as an alternative to Kolb. *The International Journal of Management Education*, *20*(3), 100703. http://dx.doi.org/10.1016/j.ijme.2022.100703

Salvi, A., Spagnoletti, P., & Noori, N. S. (2022). Cyber-resilience of critical cyber infrastructures: Integrating digital twins in the electric power ecosystem. *Computers & Security*, *112*, 102507. http://dx.doi.org/10.1016/j.cose.2021.102507

Samtani, S., Yidong Chai, & Hsinchun Chen. (2022). Linking exploits from the dark web to known vulnerabilities for proactive cyber threat intelligence: An attention-based deep structured semantic model. *MIS Quarterly, 46*(2), 911–946. https://doi.org/10.25300/MISQ/2022/15392

Sancho, J. C., Caro, A., Ávila, M., & Bravo, A. (2020). New approach for threat classification and security risk estimations based on security event management. *Future Generation Computer Systems, 113*, 488–505. https://doi.org/10.1016/j.future.2020.07.015

Sathurshan, M., Saja, A., Thamboo, J., Haraguchi, M., & Navaratnam, S. (2022). Resilience of critical infrastructure systems: a systematic literature review of measurement frameworks. *Infrastructures, 7*(5), 67. https://doi.org/10.3390/infrastructures7050067

Schaberreiter, T., Kupfersberger, V., Rantos, K., Spyros, A., Papanikolaou, A., Ilioudis, C., & Quirchmayr, G. (2019). A quantitative evaluation of trust in the quality of cyber threat intelligence sources. In *Proceedings of the 14th International Conference on Availability, Reliability and Security* (pp. 1-10). http://dx.doi.org/10.1145/3339252.3342112

Schlette, D., Böhm, F., Caselli, M., & Pernul, G. (2021a). Measuring and visualizing cyber threat intelligence quality. International Journal of *Information Security, 20*, 21-38. https://doi.org/10.1007/s10207-020-00490-y

Schlette, D., Caselli, M., & Pernul, G. (2021b). A comparative study on cyber threat intelligence: The security incident response perspective. *IEEE Communications Surveys & Tutorials, 23*(4), 2525-2556. https://doi.org/10.1109/comst.2021.3117338

Schlette, D., Vielberth, M., & Pernul, G. (2021c). CTI-SOC2M2 – The quest for mature, intelligence-driven security operations and incident response capabilities. *Computers & Security, 111*. https://doi.org/10.1016/j.cose.2021.102482

Sen, Ö., van der Velde, D., Wehrmeister, K. A., Hacker, I., Henze, M., & Andres, M. (2022). On using contextual correlation to detect multi-stage cyber attacks in smart grids. *Sustainable Energy, Grids and Networks, 32*, 100821.

http://dx.doi.org/10.1016/j.segan.2022.100821

Serketzis, N., Katos, V., Ilioudis, C., Baltatzis, D., & Pangalos, G. J. (2019). Actionable threat intelligence for digital forensics readiness. *Information & Computer Security*. http://dx.doi.org/10.1108/ICS-09-2018-0110

Shin, B., & Lowry, P. B. (2020). A review and theoretical explanation of the 'Cyberthreat-Intelligence (CTI) capability' that needs to be fostered in information security practitioners and how this can be accomplished. *Computers & Security*, *92*, 101761. http://dx.doi.org/10.1016/j.cose.2020.101761

Song, J., Kang, J., Hyun, S., Jee, E., & Bae, D. H. (2022). Continuous verification of system of systems with collaborative MAPE-K pattern and probability model slicing. *Information and Software Technology, 147*, 106904.

Smith, D. C. (2021). Cybersecurity in the energy sector: are we really prepared? *Journal of Energy & Natural Resources Law, 39*(3), 265–270. https://doi.org/10.1080/02646811.2021.1943935

Sreedevi, A. G., Harshitha, T. N., Sugumaran, V., & Shankar, P. (2022). Application of cognitive computing in healthcare, cybersecurity, big data and IoT: A literature review. *Information Processing & Management, 59*(2), 102888. https://doi.org/10.1016/j.ipm.2022.102888

Sun, N., Ding, M., Jiang, J., Xu, W., Mo, X., Tai, Y., & Zhang, J. (2023). Cyber threat intelligence mining for proactive cybersecurity defense: a survey and new perspectives. *IEEE Communications Surveys & Tutorials, 25*(3), 1748-1774. https://doi.org/10.1109/comst.2023.3273282

Suryotrisongko, H., Musashi, Y., Tsuneda, A., & Sugitani, K. (2022). Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing. *IEEE Access, Access, IEEE, 10*, 34613–34624. https://doi.org/10.1109/ACCESS.2022.3162588

Tejada, Z. (2017). *Mastering azure analytics: architecting in the cloud with azure data lake, HDInsight, and Spark*. O'Reilly Media, Inc.

Temple University (n.d.). Dataset: *Critical infrastructure ransomware attacks*. Retrieved May 22, 2021 from https://sites.temple.edu/care/ci-rw-attacks/

Tennant, J. P., & Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research integrity and peer review, 5*(1), 6. https://doi.org/10.1186/s41073-020-00092-1

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 (CCWC)* (pp. 0766-0772). IEEE. http://dx.doi.org/10.1109/CCWC51732.2021.9376038

Thulfiqar Jabar, & Manmeet Mahinderjit Singh. (2022). Exploration of mobile device behavior for mitigating advanced persistent threats (APT): A systematic literature review and conceptual framework. *Sensors, 22*(4662), 4662. https://doi.org/10.3390/s22134662

Tiwari, A., & Tiwari, A. (2020). Automation in decision OODA: Loop, spiral, or fractal. *i-Manager's Journal on Communication Engineering and Systems, 9*(2), 1. https://doi.org/10.26634/jcs.9.2.18116

Tuma, K., Sandberg, C., Thorsson, U., Widman, M., Herpel, T., & Scandariato, R. (2021). Finding security threats that matter: Two industrial case studies. *Journal of Systems and Software, 179*, 111003. https://doi.org/10.1016/j.jss.2021.111003

Tuyishime, E., Balan, T. C., Cotfas, P. A., Cotfas, D. T., & Rekeraho, A. (2023). Enhancing cloud security—Proactive threat monitoring and detection using a SIEM-based approach. *Applied Sciences, 13*(22), 12359. https://doi.org/10.3390/app132212359

Verma, V., & Aggarwal, R. K. (2020). A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective. *Social Network Analysis and Mining, 10*(1), 43. https://doi.org/10.1007/s13278-020-00660-9

Wescott, J. (2020). *Bring your threat intelligence to Azure Sentinel*. Microsoft. https://techcommunity.microsoft.com/t5/azure-sentinel/bring-your-threat-intelligence-to-azure-sentinel/ba-p/1167546

Yadav, A., Kumar, A., & Singh, V. (2023). Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security. Artificial Intelligence Review, 56(11), 12407-12438. https://doi.org/10.1007/s10462-023-10454-y

Yan Guo, Hiu Ching Lam, A., Chiu, D. K. W., & Ho, K. K. W. (2022). Perceived Quality of WhatsApp Reference Service: A Quantitative Study from User Perspectives. *Information Technology & Libraries, 41*(3), 1–17. https://doi.org/10.6017/ital.v41i3.14325

Wagner, T. D., Mahbub, K., Palomar, E., & Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security, 87*, 101589. https://doi.org/10.1016/j.cose.2019.10158

Wallis, T., & Leszczyna, R. (2022). EE-ISAC—practical cybersecurity solution for the energy sector. *Energies, 15*(6), 2170. https://doi.org/10.3390/en15062170

Zibak, A., Sauerwein, C., & Simpson, A. (2021). A success model for cyber threat intelligence management platforms. *Computers & Security, 111*, 102466. https://doi.org/10.1016/j.cose.2021.102466

Zibak, A., Sauerwein, C., & Simpson, A. C. (2022). Threat intelligence quality dimensions for research and practice. *Digital Threats: Research and Practice, 3*(4), 1-22. https://doi.org/10.1145/3484202

Zhao, J., Yan, Q., Li, J., Shao, M., He, Z., & Li, B. (2020). TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data. *Computers & Security, 95*, 101867. https://doi.org/10.1016/j.cose.2020.101867

Zrahia, A. (2018). Threat intelligence sharing between cybersecurity vendors: Network, dyadic, and agent views. *Journal of Cybersecurity, 4*(1), tyy008. https://doi.org/10.1093/cybsec/tyy008

Appendix A: CTI Dataset

Sample of CTI Dataset

Date	Id	Indicator of Compromise (IoC)	Type	Title	Descrip-tion	Source/ Data Connector
Fri, 10 Nov 2023 06:34:37 GMT	3768301872	81.19.135.226	IPv4	null	CC=RU ASN=AS209588 Flyservers S.A.	AlienVault/ LogicApps
Sat, 25 Nov 2023 06:34:41 GMT	3784485387	ED5D694D561C97B4D70EFE934936286FE562ADDF7D6836F795B336D9791A5C44	File Hash-SHA256	null	null	AlienVault/ LogicApps
Thu, 23 Nov 2023 06:34:41 GMT	3784485401	adobe-us-updatefiles.digital	Domain	null	null	AlienVault/ LogicApps
Thu, Jan 5, 2023, 9:59:37 AM CST	N/A	d241df7b9d2ec0b8194751cd5ce153e27cc40fa4	File SHA-1	N/A	HackTool:Win32/ Mimikatz.D / SHA1 of 31eb1de7e840a342fd468e558e5ab627bcb4c542a8fe01aec4d5ba01d539a0fc	OTX-AlienVault/ Sentinel

Note. This table presents a CTI data feed from AlienVault, including the IoC, indicator type, description, and data source. The instruments used to collect this data are Microsoft Azure Logic Apps and Sentinel.

Appendix B: OSINT Dataset

Sample of OSINT Dataset of Ransomware

Date	Ransomware Strain	MITRE ATTACK	Types (Sector)	Organization	Location
11/08/2023	LockBit		Financial Services	The Industrial and Commercial Bank of China (ICBC)	China (ICBC)
11/11/2023	BlackCat/ ALPHV	S1068	Information Technology	Dragos	USA
11/13/2023	Medusa		Financial Services	Moneris Solutions	Canada
11/13/2023	Medusa		Financial Services	Toyota Financial Services	Japan
11/13/2023	Snatch		Food	Tyson Foods	USA
11/19/2023	BlackCat/ALPHV	S1068	Financial Services	Fidelity National Financial	USA
11/22/2023	BlackCat/ ALPHV	S1068	Healthcare	Henry Schein	USA
11/23/2023	Meow		Education Facilities	Vanderbilt University Medical Center	USA, TN
11/24/2023	NoEscape		Healthcare	Granger Medical Clinic	USA, UT
11/24/2023	Rhysida		Energy	China Energy Engineering	China
11/25/2023	Hunters International Ransomware Team		Healthcare	Crystal Lake Health Center	USA, MI

Note. This table presents real-world ransomware critical infrastructure (CI) cyberattacks gathered from OSINT sources. Adapted from Rege, A. (2023). “Critical Infrastructure Ransomware Attacks (CIRA) Dataset”. Version 12.9. Temple University. Online at

https://sites.temple.edu/care/cira/. Funded by National Science Foundation CAREER Award #1453040. ORCID: 0000-0002-6396-1066.

Appendix C: Alert Data

Sample of Alert Data

Date	Source	Alert Code	Reference	Strain	Indicator of Compromise (IoC)
11/21/2023	CISA	AA23-325A	https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-325a	LockBit	81.19.135.226
11/21/2023	CISA	AA23-325A	https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-325a	LockBit	ED5D694D561C97B4D70EFE934936286FE562ADDF7D6836F795B336D9791A5C44
11/21/2023	CISA	AA23-325A	https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-325a	LockBit	adobe-us-updatefiles.digital
04/22/2022	CISA/FBI	FBI Flash CU-000167-MW	https://www.cisa.gov/news-events/alerts/2022/04/22/fbi-releases-iocs-associated-blackcatalphv-ransomware	BlackCat/ALPV	d241df7b9d2ec0b8194751cd5ce153e27cc40fa4

Note. A sample of the alert dataset with IoC was gathered from OSINT sources (e.g., CISA).

Appendix D: Research Variables Correlation

Sample of Research Variables Correlation

Variables	Type	Source/ Mechan-ism	Similar-Based Link	Data Type	Indicator of Compromise (IoC)
Dependent	OSINT	Temple/ Manual	CISA	Ransom-ware	LockBit
Dependent	OSINT	CISA/Web	LockBit	IPv4	81.19.135.226
Independent	CTI	AlienVault/LogicApps	IoC	IPv4	81.19.135.226
Dependent	OSINT	CISA/Web	LockBit	File	ED5D694D561C97B4D70EFE934936286FE562ADDF7D6836F795B336D9791A5C44
Independent	CTI	AlienVault/LogicApps	IoC	File	ED5D694D561C97B4D70EFE934936286FE562ADDF7D6836F795B336D9791A5C44
Dependent	OSINT	CISA/Web	LockBit	Domain	adobe-us-updatefiles.digital
Independent	CTI	AlienVault/LogicApps	IoC	Domain	adobe-us-updatefiles.digital
Dependent	OSINT	Temple/ Manual	CISA	Ransom-ware	BlackCat/ALPV
Dependent	OSINT	CISA/Web	Black Cat/ ALPHV	File	d241df7b9d2ec0b8194751cd5ce153e27cc40fa4
Independent	CTI	AlienVault/Sentinel	IoC	File SHA-1	d241df7b9d2ec0b8194751cd5ce153e27cc40fa4

Note. This table presents relationships of variables with a direct correlation of matching IoC, noting similar-based links. The experiment and application phase used automated collection mechanisms (LogicApps, Sentinel) to collect data from the AlienVault data source.

Appendix E: Sample of STIX Data

CTI STIX Data Format Sample

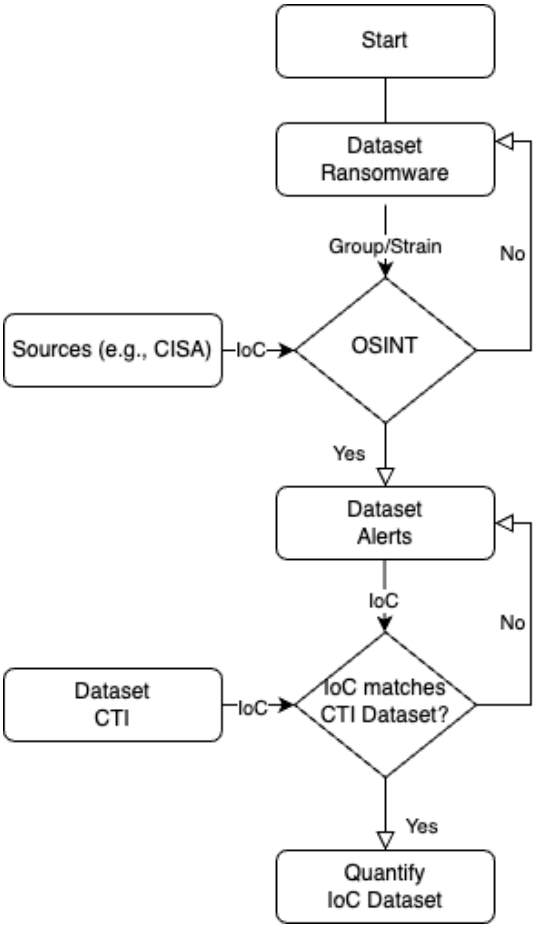

```
{
   "statusCode": 200,
   "headers": {
      "Connection": "keep-alive",
      "Date": "Fri, 10 Nov 2023 06:34:37 GMT",
      "Server": "gunicorn",
      "X-Frame-Options": "SAMEORIGIN",
      "X-Remote-User-Name": "albertysu",
      "X-OTX-ACTIVE": "1",
      "Access-Control-Allow-Origin": "*",
      "Vary": "Accept-Encoding",
      "X-Cache": "Miss from cloudfront",
      "Via": "1.1 060d1d4202fff21bc3e6249da337fee6.cloudfront.net (CloudFront)",
      "X-Amz-Cf-Pop": "ORD52-C3",
      "X-Amz-Cf-Id": "E813XCvt2KpYxD-sRBtJTob9jfWGk7A9Nb-
WaDtwsOWg2XT8dOrS7w==",
      "Content-Type": "application/json"
   },
   "body": {
      "results": [
         {
            "id": 3768301872,
            "indicator": "81.19.135.226",
            "type": "IPv4",
            "title": null,
            "description": "CC=RU ASN=AS209588 Flyservers S.A.",
            "content": ""
         },
…
],
      "count": 432,
      "previous": null,
      "next": null
   }
}
{
   "statusCode": 200,
   "headers": {
      "Connection": "keep-alive",
      "Date": "Thu, 23 Nov 2023 06:34:41 GMT",
```

```
      "Server": "gunicorn",
      "X-Frame-Options": "SAMEORIGIN",
      "X-Remote-User-Name": "albertysu",
      "X-OTX-ACTIVE": "1",
      "Access-Control-Allow-Origin": "*",
      "Vary": "Accept-Encoding",
      "X-Cache": "Miss from cloudfront",
      "Via": "1.1 4d5a0f70d1ac913d9a695c407458cff6.cloudfront.net (CloudFront)",
      "X-Amz-Cf-Pop": "ORD52-C3",
      "X-Amz-Cf-Id":
"twgXPEgRUE94NgiyQx5NnGQmJxoLgoxGqfytFjvkkgaEPFBoxViJgw==",
      "Content-Type": "application/json"
   },
   "body": {
      "results": [
         {
            "id": 3784485387,
            "indicator":
"ed5d694d561c97b4d70efe934936286fe562addf7d6836f795b336d9791a5c44",
            "type": "FileHash-SHA256",
            "title": null,
            "description": "null",
            "content": ""
         },
…
],
      "count": 102,
      "previous": null,
      "next": null
   }
}
{
   "statusCode": 200,
   "headers": {
      "Connection": "keep-alive",
      "Date": "Thu, 23 Nov 2023 06:34:41 GMT",
      "Server": "gunicorn",
      "X-Frame-Options": "SAMEORIGIN",
      "X-Remote-User-Name": "albertysu",
      "X-OTX-ACTIVE": "1",
      "Access-Control-Allow-Origin": "*",
      "Vary": "Accept-Encoding",
      "X-Cache": "Miss from cloudfront",
      "Via": "1.1 4d5a0f70d1ac913d9a695c407458cff6.cloudfront.net (CloudFront)",
```

```
      "X-Amz-Cf-Pop": "ORD52-C3",
      "twgXPEgRUE94NgiyQx5NnGQmJxoLgoxGqfytFjvkkgaEPFBoxViJgw==",
      "Content-Type": "application/json"
   },
   "body": {
      "results": [
         {
            "id": 3784485401,
            "indicator": "adobe-us-updatefiles.digital",
            "type": "domain",
            "title": null,
            "description": "null",
            "content": ""
         },
…
],
      "count": 432,
      "previous": null,
      "next": null
   }
}
```


Note. A sample of CTI data collection in an STIX data format. The instrument used is Microsoft Logic Apps, querying an AlienVault TAXII server.

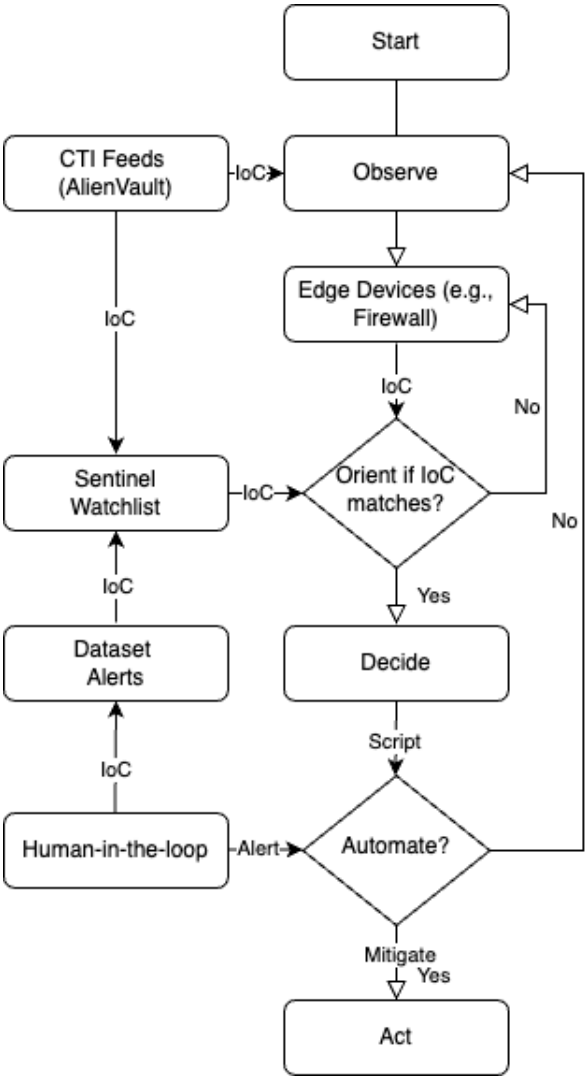
Appendix F: Dataset Correlation Flowchart



Note. This flowchart represents the data correlation for this CTI experiment. The hierarchical nature of datasets top-down is dataset ransomware, alert dataset generated from OSINT, and CTI dataset.

Appendix G: OODA Model Flowchart

OODA Model Flowchart in SOC Environment



Note. This flowchart represents the OODA theoretical model for this CTI experiment in a SOC environment. The instrument used is Microsoft Sentinel (SIEM/SOAR), where an analyst (Human-in-the-loop) can automate the decision-making process to mitigate a cyberattack on the Edge Device using Sentinel Watchlist.

Appendix H: Sample of AlienVault Data

Sample of AlienVault Data on Sentinel

Name
OTX feed=Collection: user_AlienVault, pulse_name=C
Values
hashes.'SHA-1' : d241df7b9d2ec0b8194751cd5ce153e27cc40fa4
Tags
Threat types
Description
HackTool:Win32/Mimikatz.D / SHA1 of 31eb1de7e840a342fd468e558e5ab627bcb4c542a8fe01aec4d5ba01d539a0fc
Name
OTX feed=Collection: user_AlienVault, pulse_name=Cyberespionage Attacks Against Southeast Asian Government Linked to Stately Taurus, Aka Mustang Panda
Revoked
Confidence
--
Source
OTX-AlienVault
Pattern
[file:hashes.'SHA-1' = 'd241df7b9d2ec0b8194751cd5ce153e27cc40fa4']
Kill chains
Created
Thu, Jan 5, 2023, 9:59:37 AM CST
Valid from
Mon, Sep 25, 2023, 1:52:46 PM CDT
Valid until
Modified
Sun, Jan 14, 2024, 9:05:15 AM CST
Created by

Note. A sample of AlienVault data collection using a TAXII data connector from the Microsoft Sentinel application.

Appendix I: CTI Data Collection Code

CTI Data Collection Code using Microsoft Logic Apps


```
{
   "definition": {
      "$schema":
"https://schema.management.azure.com/providers/Microsoft.Logic/schemas/2016-06-
01/workflowdefinition.json#",
      "actions": {
         "HTTP": {
            "inputs": {
               "headers": {
                  "X-OTX-API-KEY":
"3f72665898dd96f0fa524420da66d4d7015b5xxa56f84710371dc530x9bf60pa"
               },
               "method": "GET",
               "queries": {
                  "modified_since": "@{addDays(utcNow(),-1)}",
                  "types": "domain,email,FileHash-MD5,FileHash-SHA1,FileHash-
SHA256,IPv4,URL,URI"
               },
               "uri": "https://otx.alienvault.com/api/v1/indicators/export"
            },
            "runAfter": {},
            "type": "Http"
         }
      },
      "contentVersion": "1.0.0.0",
      "outputs": {},
      "parameters": {},
      "triggers": {
         "Recurrence": {
            "evaluatedRecurrence": {
               "frequency": "Day",
               "interval": 1
            },
            "recurrence": {
               "frequency": "Day",
               "interval": 1
            },
            "type": "Recurrence"
         }
```

```
        }
    },
    "parameters": {}
}
```

Note. This code is used for CTI data collection during the CTI experiment. The instrument used is Microsoft Logic Apps, which allows data feeds from AlienVault to be collected.

Appendix J: RSS Data Collection Code

RSS Data Collection Code using Microsoft Logic Apps


```
{
   "definition": {
      "$schema":
"https://schema.management.azure.com/providers/Microsoft.Logic/schemas/2016-06-
01/workflowdefinition.json#",
      "actions": {
         "List_all_RSS_feed_items": {
            "inputs": {
               "host": {
                  "connection": {
                     "name": "@parameters('$connections')['rss']['connectionId']"
                  }
               },
               "method": "get",
               "path": "/ListFeedItems",
               "queries": {
                  "feedUrl": "https://www.cisa.gov/cybersecurity-advisories/ics-
advisories.xml",
                  "sinceProperty": "UpdatedOn"
               }
            },
            "runAfter": {},
            "type": "ApiConnection"
         }
      },
      "contentVersion": "1.0.0.0",
      "outputs": {},
      "parameters": {
         "$connections": {
            "defaultValue": {},
            "type": "Object"
         }
      },
      "triggers": {
         "Recurrence": {
            "evaluatedRecurrence": {
               "frequency": "Day",
               "interval": 1,
               "startTime": "2023-09-24T16:58:00Z",
```

```
            "timeZone": "Central Standard Time"
          },
          "recurrence": {
            "frequency": "Day",
            "interval": 1,
            "startTime": "2023-09-24T16:58:00Z",
            "timeZone": "Central Standard Time"
          },
          "type": "Recurrence"
        }
      }
    },
    "parameters": {
      "$connections": {
        "value": {
          "rss": {
            "connectionId": "/subscriptions/e3afdd4f-f68e-4eb4-a94c-
24cdc9427d7c/resourceGroups/Vsentinel/providers/Microsoft.Web/connections/rss",
            "connectionName": "rss",
            "id": "/subscriptions/e3afdd4f-f68e-4eb4-a94c-
24cdc9427d7c/providers/Microsoft.Web/locations/centralus/managedApis/rss"
          }
        }
      }
    }
}
```


Note. The code from Microsoft Azure Logic Apps is used to gather RSS data feeds. The data collection code is a workflow that shows daily collection (recurrence) of RSS feed advisories (Industrial Control Systems) from data source (www.cisa.gov).