Operational Technology Threat Detection and Intrusion Response: AI-Enhanced Cybersecurity Frameworks for U.S. Manufacturing Infrastructure Protection

Dr. Cristina Mateos, Professor of Human-Computer Interaction, Universidad Politécnica de Madrid (UPM), Spain

1. Introduction to AI-Enhanced Cybersecurity in U.S. Manufacturing, As part of a ten-year plan to bolster U.S.-owned manufacturing infrastructures, there has been an emphasis on advancing cybersecurity capabilities. Manufacturers have the ability to create increasingly complex products at high-reliability scales, consequences of which can lead to a greater potential for systemic failure due to complexity. This presents a cybersecurity vulnerability ripe for exploiting by bad actors, recent events of which have shown the disruptive capabilities of even non-state sponsored adversaries. Trusted AI has been proposed as an enabling technology whose capabilities would facilitate cyber resiliency in high-stakes manufacturing ecosystems. Applied malicious AI examples in critical infrastructure often focus on denial of service (DoS) attacks, but there is a notable trend of compromised hardware within supply chains that place data exfiltration at a higher success rate than a direct cyber-physical attack. This aligns with the current use scenario, enforcing tamper evidence on equipment requiring IIoT collection in U.S. manufacturing supply chains based on a proof-of-concept implementation to meet domain-specific functionality.

The chief motivation behind protecting data and the devices that integrate a production conglomerate revolves around the mitigation of financial loss that occurs when digital assets are stolen. Recovery from financial insolvency hampers restorative resources that help in rebuilding reputations and customers' support. Though of lesser consideration, recovering from a reputation death spiral is strategically harder for small- and medium-size manufacturers versus their larger counterparts. AI-enhanced cybersecurity solutions presented in this document are designed with these foundational considerations and consequences in mind. They are designed to present cyber-in-a-box recommendations for deployment onto mainly brownfield manufacturing deployments, capable of being installed with medium engineering investments. The system purposefully is an endpoint security solution to cover a million previously unconnected devices deployed across U.S. industrial infrastructures and meeting the prioritized

functional constraints in application in providing just-in-time cybersecurity programming capabilities. The resulting specifications are presented to ensure that the AI-enhanced cybersecurity solution meets the existing guideline requirements for major stakeholders. Implications due to the main system functionality and the derived security requirements are two-fold, one dedicated to the techniques that emerge and another bound to the real-world application.

1.1. Overview of Cybersecurity Threats in Manufacturing

Over the past decade, there has been rapid digitalization and automation of many processes in the manufacturing industry across multiple critical infrastructures. It is expected that this trend will continue and that cyber threats will also steadily increase in frequency and complexity. The U.S. manufacturing sector is prime for cyber-attack, particularly given increasing reliance on the Internet and interconnected systems. In addition, the industrial control systems (ICS) commonly used in manufacturing are a fairly weak link in terms of cybersecurity barriers. These systems were never designed to be secure and often have limited computing capacity. Both of these aspects predispose the manufacturing sector to aggressive cyber threats. The physicality of well-publicized cyber attacks that have taken place in Ukraine (2015, 2016, etc.) are similar in nature to the threats looming in U.S. manufacturing theaters. This wide-reaching cyber attack took down tens of thousands of Ukrainian IT systems and disrupted all forms of electronic communication. In 2017, the United States and European Union coalition counseled that this attack threat model was not only plausible but very likely, strengthening the need for more secure and robust network architectures.

The above threats, and existing control structures to attempt to mediate them, lead to a need for active and adaptive cyber defense. Our work focuses on active, adaptive defense using new, emerging technologies including AI and big data. In particular, we focus on deep learning (DL) as it is now the state-of-the-art in AI cyber threat assessments and decision-making. The remainder of this document discusses methods and procedures to marry and harness DL and big data for AI-enhanced cybersecurity protecting U.S. manufacturing cyber infrastructure.

2. Fundamentals of Artificial Intelligence in Cybersecurity

The adoption of artificial intelligence (AI) technologies to fill gaps in endpoint, network, and perimeter security has become mainstream, yet the application of these technologies

within the manufacturing sector has received little attention. The use of AI in cybersecurity necessitates an understanding of the principles of AI itself. As such, the field of AI has many different subsets, often referred to as weak or strong AI, and further still, as narrow or general AI. Weak AI is a term that refers to a computer's demonstration of "intelligent" activity, such as that demonstrated by computer programming, yet these activities are little more than imitations of human intelligence.

On the other hand, strong AI refers to a system of operations that demonstrate evidence of non-simulated human cognition. Narrow AI and general AI further subdivide these categories: narrow AI, which is widely used in cybersecurity practices, is designed to work within a specific parameter, while general AI is designed to act inside a wide array of tasks not explicitly programmed. Intelligence is an enigmatic entity that is not easily defined, and this eludes the use of the term AI to describe any or all of it. In cybersecurity, artificial intelligence and machine learning are sometimes used interchangeably; however, this is inaccurate. Many AI theories have been used to guide the architecture of intelligent agents. Understanding and implementing heuristics, such as the development of binding (team formation), learning and memory, goal reasoning, planning, acting, and integrating - amongst others - is critical in countermeasure logic development for cybersecurity undertakings. The computational models listed can also be used to provide a functional description of the I3P model.

2.1. Machine Learning and Deep Learning Algorithms

Machine learning (ML) is a field of artificial intelligence (AI) that uses sound mathematical and statistical principles to develop models in a predictive way, which are then used for decision making. When ML models are trained using large amounts of data, they can resemble a person who "learns" information from interactions with datasets. Machine learning can be broken down into different categories (e.g., supervised, unsupervised, semi-supervised, reinforcement learning, and so on). There are various algorithms such as Naive Bayes, k-nearest neighbors (k-NN), decision trees, random forest, gradient boosting, support vector machines (SVM), principal component analysis (PCA), expectation-maximization (EM), k-means, and so on that can be used for each category of machine learning.

Deep learning (DL) is a subset of ML and can be defined as a network with many "layers". Deep learning algorithms are designed to automatically feature-extract

information directly from the data that feed into them. Common deep learning methods include convolutional neural networks (CNNs) for imaging data, recurrent neural networks (RNNs) for temporal/spatial data, long short-term memory (LSTM) units, and autoencoders. Deep learning contrasts with "classic" ML that uses human experts to hand-craft features from raw data in order to facilitate learning. As of today, artificial intelligence (AI) can be associated with security systems to process a large amount of data (e.g., network traffic patterns for intrusion detection) and perform tasks that are typically done by cybersecurity analysts (e.g., alert generations, log analysis, etc.). While historically AI has been associated with heavy computational and storage needs, integrated cloud platforms can now take advantage of AI technologies with large datasets as well as compute power.

3. Integration of AI in Manufacturing Infrastructure Protection

The nations' manufacturing infrastructure is becoming more accessible and interconnected because of technological advancement. This infrastructure needs to be protected from cyber and physical threats. The current protective measures are not able to prevent all these types of threats. One technique widely used for protection of cyber threats is to maintain a consistently stringent, core rule-set that defines what is allowed or denied. These are effective to the extent of one because rule-sets have been developed to be more tolerant towards false positive alarms. The false negative alarms have been increased in this process. Secondly, these rule-sets are not able to detect novel threats. To overcome these limitations, our team is currently working to enhance those rule-sets using artificial intelligence in general and machine learning in particular.

In this manuscript, we specifically explain how we can use AI to enhance cybersecurity measures to protect manufacturing infrastructure. The concepts and real-world applications discussed in this paper are not general and are specific to enhancing cybersecurity in the manufacturing environment and not in neuro inputs/synapses. The important contribution and keynote of this paper includes the use of AI for protecting manufacturing assets using the certificates issued by the original manufacturer of the hardware and the original software installed in the manufacturing facility. We will also discuss the integration of double fail-safe cybersecurity measures.

3.1. AI-Driven Intrusion Detection Systems

In our daily life, we interact with physical entities in various possible ways. Cyber-physical systems (CPSs) blend the essence of the cyber world with the operations in the physical world. In today's age, data-driven techniques and methodologies are employed to sustain operational efficiency, which draws our attention towards refinement in manufacturing, supply chain, customer fulfillment, and financial cryptography. However, the allegorical numerical journey during digital processing is one of the biggest reasons due to which intrusion occurrences in such operational economic infrastructures have recently become not only probable but much easier to achieve. Intrusion Detection Systems (IDS) using AI technologies for logical and physical security enhancements (AI-IDS), such as cloud, research software, and computing infrastructure, are currently serving as the prime choice of academics and researchers in the cyber-physical world.

An IDS with AI-enabled infrastructure is intended to support concerned stakeholders by exploring the areas to identify and take appropriate actions to confront possible invasions. The chain-range in an AI-IDS confines or confines integrated operations and moves to the domain of the cyber-physical environment. Focusing on the real-time domain of AI-IDS that we pursue, as intrusion detectors, our primary focal point is on two functionalities of AI, i.e., the exploitation of the behavioral training data of underlying operations and the inspection of underlying interactions via experiences' outcomes to a very high abstraction level beyond the limited frame of hardware and machine learning. Thus, detailing and researching core components of AI-IDS as a solution furthermore will make us provide real-time solutions supporting automation paradigms.

4. Real-World Applications of AI in U.S. Manufacturing

When charting the surge in U.S. manufacturing infrastructure breaches, it becomes clear that a preventative cybersecurity approach is needed. Incorporating AI technology can help fortify manufacturing security measures. Some of these solutions use AI in real-life manufacturing contexts. BrainChip has demonstrated how an autonomous AI cybersecurity posture could help protect U.S. manufacturing infrastructure and applications. Their AI has shown 98% capability in detecting unauthorized cyber activity in scanned server system data across a wide variety of U.S. metropolitan sites servicing

manufacturing. They have used data files from a large server with a normal request loading of over 300,000 requests per minute, with regular maintenance-driven pauses. AI correctly detected less than 2% legitimate site servicing requests that are largely attributable to staff and subject-matter expert vendor support activities.

The challenge in finding an adaptive cybersecurity solution for U.S. manufacturers is to use AI without getting them caught in the middle of AI tools and security that will not work in a U.S. manufacturing environment. Forward-thinking cybersecurity vendors are already working to help facilitate this industry adaptation. A major U.S. manufacturer is already piloting a new AI cybersecurity solution for some of their new equipment. Specifically, semiconductor chip machines are being protected by autonomous AI software platforms from a small consortium of AI and cybersecurity vendors. These solutions detect, analyze, and adapt to device-level security threats without the need for prior programming of the assessment algorithms. To effectively mitigate the wide variety of macro- and micro-level behaviors by threats to U.S. manufacturers, this effort is exploring how manufacturers can exploit AI to move away from manual, human-run security systems with a binary response to threats.

4.1. Case Studies of Successful Implementations

Example 1: CISA 'CYgN' geographical diverse facility case study. In this case, during the startup phase, CYgN conducted MITRE ATT&CK exercises on diverse test cases of the above facility. To date, because of using AI for cybersecurity, CYgN has stopped 43 targeted attacks and 3 algorithms have generated action responses with all valid test results. Although this technology has been reduced to practice in a technical pilot, no cost data is available, as the phase II award and the resulting SAA is relatively new.

Example 2: University of Illinois Urbana-Campaign/UDRI/Engility CyberLab and the Arms/SofS use case. The CyberLab at Illinois has been working with a diverse group including utilities and manufacturers conducting various research. One manufacturer, Arms Control, has implemented the Honeywell/UDRI/FORGE research in their facilities linking a Digital Twin capability to their operations network in support of active defense. This move from academia to operations provided capabilities that can be leveraged to provide advance warning of threats similar to those involved in significant hacking events. Specifically, the Arms use case.

Example 3: MIT LL 'Red and Blue' defense exercise. This offline Lab evaluation is a CISA/MWRD contract looking to leverage GURTEK as a surrogate for state-of-the-practice BtoB and BtoC Manufacturing Environments. Based upon our current validation of the YK-PP playing field in the QCT, our evaluation will focus on both the Red Team 'Hacking to Break' and the Blue Team 'detect-to-protect' experiences on the GUR-TEK HIL framework. In this exercise, AI will be used in the form of Digital Twin and Gamalization to train participants on how to identify potential threats as they build and deploy frameworks for active response.

Example 4: Secretary of the Air Force (SAF/USF) Priorities Received. With safe alignment also a priority for the digital thread, this initially standalone activity could have promising acquisition opportunities with Digital Twin application potential in commercial industries.

5. Challenges and Limitations of AI-Enhanced Cybersecurity in Manufacturing

This section provides a comprehensive insight to see viable solutions and their challenges. Throughout our discussions, provide evidence to underscore that combining the two technologies: AI and cybersecurity may result in addressing security issues in the manufacturing industry. This discussion provides an overview of some of the challenges and limitations, which need to be addressed. They assert that AI cybersecurity solutions are currently not feasible as there are no considerable cybersecurity databases existing. AI-enhanced cybersecurity in manufacturing also has its own challenges and limitations.

One of the biggest challenges to implementing AI-enhanced cybersecurity solutions in the context of U.S. manufacturing infrastructure is that there is a lack of a large and diverse dataset required to develop data-hungry AI applications. Especially in settings like manufacturing facilities, the volume of network traffic and user interactions may not be sufficient to discern regular activity from really anomalous behavior. Masquerading attacks are when users on these networks show distinct behavioral patterns depending on the terminal device they use, and these need to be identified in the data. Extreme or rare events like emergency shutoffs, production speed changes, and Sony-style malware attacks that may only occur once in five years still need to be captured in the training dataset in order to anticipate them in live data.

5.1. Data Privacy and Ethical Concerns

Any form of cybersecurity solution raises privacy concerns. Likewise, AI is associated with philosophical, ethical, and technical challenges. Since such a duel for AI-enhanced cybersecurity systems is not much discussed, we address some of the typical data privacy and ethical concerns.

When the data used by AI come from rote learning, this work could be inconspicuously training on the proprietary data of a user. Sharing the data between two different companies could be looked on with disfavour from sharing personal information, and submission of a machine for training could be considered a malicious act towards a partner user if the data from such a training were to be leaked.

More practical, an AI system that uses user data can never be rendered fully functional if the user is granted the request to erase this data for the purpose of training, since this would require starting anew. This is further compounded by the question of whether the training process, data sharing and the method specifications are considered proprietary information. Furthermore, aside from these proprietary concerns, own enterprise data needs to be segregated.

Generic compliance and integrity of the system may be affected. Considering that many users do not trust a meal that is not cooked at home, the solution may not be desired by the users. The effectiveness from an ethical standpoint closely shadows the first ethical principle.

On the one hand, because adversarial systems are still under construction, security issues might escape, potentially leading to system destruction. Consequently, there is a fear of the "Frankenstein scenario". On the other hand, if these issues could be taken care of, the system might be too secure, making it difficult for intrusion tests. If someone wanted to verify if the firewall could be compromised, they may not be able to confirm this as a fact. For example, securiCAD, Cyberbit, AGT International, Borwell, and Comforte produce security systems that learn passwords and tokens.

6. Future Trends and Innovations in AI-Enhanced Cybersecurity for Manufacturing

Current trends in cybersecurity for industrial control systems in the energy field will continue to become more common throughout all industrial control systems. At the same time, the sophistication of hackers will continue to evolve. This is particularly true

in attacks targeting AI systems. As AI continues to grow and become more central to protecting manufacturing infrastructure, advancements in more robust, secure, explainable, and scalable AI systems that can learn online from data at scale will change the face of AI-enhanced cybersecurity. Research is already underway to deal with these problems. One approach to hardening AI-enhanced security appliances against attacker AI is transformative adversarial training, which adds complex and realistic complications to decision-making problems so that they cannot be solved exactly with simple functions.

Scalable, secure, and explainable AI has been identified as the cornerstone for the future of AI-enhanced cybersecurity. As AI becomes prevalent in securing critical infrastructure, including U.S. manufacturing, innovations in applying AI to cybersecurity will continue to push the envelope on its capabilities. From learning more from less, to identifying uncertainty in data with unmodeled dynamics, new AI techniques will collectively advance the landscape of AI-enhanced cybersecurity.

6.1. Advancements in Explainable AI

Explainable AI, which has been growing in significance in applications ranging from cybersecurity to blockchain, has seen a large wave of research promoting explainability to the next level. While methods currently exist to provide post-hoc explanations for black-box models, new frontiers in XAI have emerged that provide model transparency and interpretability from the ground up. A major investor in the XAI area, ONR, has established a goal of creating a system that provides "full transparency" in deep learning. With a commitment of over $1 million to this goal, we foresee a traction of studies that contribute to groundbreaking innovations in transparency to our future cybersecurity systems.

Future Developments in Transparency and Interpretability of AI

Recent research studies have been proposing developments that revolutionize the capability of AI in transparency and interpretability. In the near future, advancements in making AI models more transparent can be expected, where semantic adjectives, e.g., invariant and equivariant, are used in model design, training, optimization, and evaluation. The contemporary models capable of untangling causal, structural, and nomic relations, hence providing interpretable understanding, housing these properties

will be developed. These will preserve "minimally designed" causal structure and evolve into building blocks for semi-transparent high-level models. A human cognition-inspired AI framework that innately captures causality, intent, invariance, and adherence to, while being articulate, amicable, malleable, and responsive will be constructed. Moreover, future AI possesses the ability to plan and strategize its intervention and manipulations for empowerment within the cyber-world.