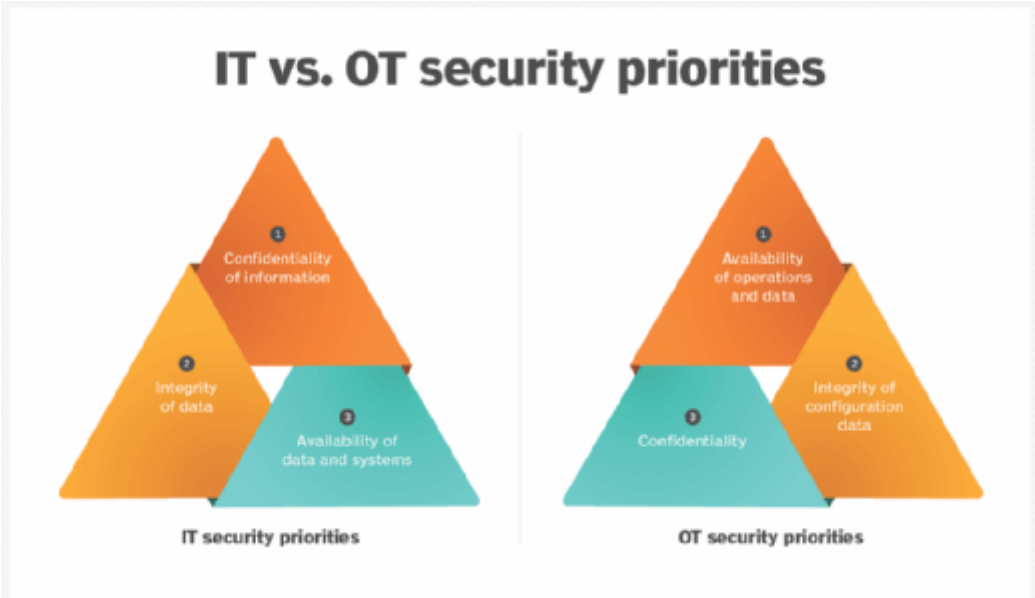
Norberto Limon
5/15/2023
GBA 6230

ICS Cybersecurity: Signature and Behavior-Based Intrusion Detection Methods

Analysis and Critique of: "Advanced Intrusion Detection Combining Signature-Based and Anomaly-Based Detection Methods,"; Department of Electrical and Computer Engineering, Inha University, Incheon 22212, Korea; 2022
https://www.researchgate.net/publication/359132130_Advanced_Intrusion_Detection_Combining_Signature-Based_and_Behavior-Based_Detection_Methods

Introduction to ICS

ICS Cybersecurity refers to a subdomain of cybersecurity that deals primarily with Industrial Cyber-Physical Systems, or Industrial Control Systems (ICS). These systems include the infrastructure like Programmable Logic controllers (PLC), Remote Terminal Units (RTU), Engineering Workstations (EW), or Supervisory Control And Data Acquisition (SCADA), used to support chemical plants, manufacturing plants, water treatment facilities, power grids, nuclear plants, satellites, and other large industrial facilities. The nature of these systems significantly impacts the way that ICS, or Operational Technology (OT) Cybersecurity prioritizes its operations, which in some ways differs significantly from IT Cybersecurity. For instance, the risk associated with a successful ICS cyberattack can take the form of espionage, sabotage, societal disruption or even loss of life. This can be caused by manipulated actuators, compromised data or malware installed to cause physical destruction, which all contribute to the priorities of ICS cybersecurity developing somewhat differently from the financial and privacy concerns of IT, putting greater emphasis on things like availability and visibility.



As ICS/OT environments continue to become increasingly connected to external communications networks to support improvements in convenience and capabilities, the challenges and associated risks also increase proportionately. Furthermore, in IT Security it is common practice to solve a system or software vulnerability by patching and rebooting the device. However, with ICS/OT Security, most of these systems cannot afford to be taken down, even for the relatively short period of time needed to receive the patch. Because of this, patches can only be administered a few times a year in most cases, diminishing the layer of protection afforded by the manufacturer updating its security posture against the ever-changing threat landscape. In an in-depth security model, this would be indicative of a softening of one of the rings of protection offered by the layer of the OEM.

In any case, for such a sensitive industry a defense-in-depth approach remains an important strategy to employ for both its IT and OT systems alike. This approach has many similarities with the zero-trust security model which tries to ensure security through visibility, authentication, and redundancy. Zero-trust in particular puts added emphasis on the importance of ID for secure privileged access management, making it an important point of consideration for modern ICS/OT security. Until the recent introduction of cloud, the way in which these sensitive resources were kept secure was through the implementation of “airgaps” on offline-only devices that protected against malicious traffic from the internet by just eliminating all communication and exposure to external networks. This is no longer the case due to advances in IIoT technology and demands of the industry, which increase capability but also increase vulnerability to many new attack vectors on very sensitive and outdated device software.

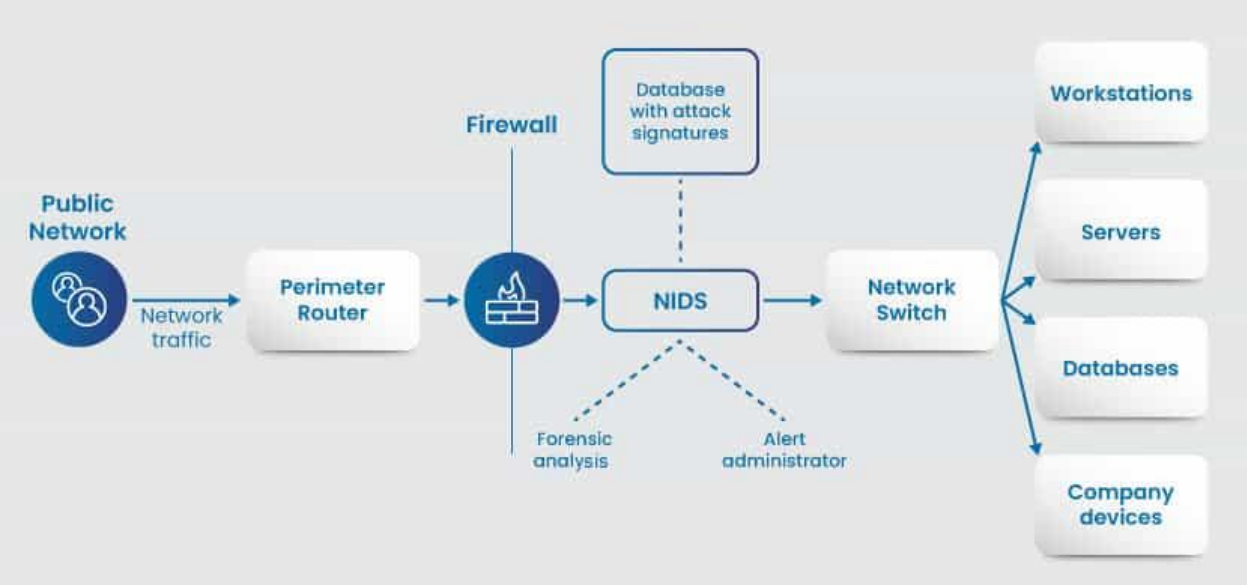
The sensitivity of these facilities keeps them high on the target list for many criminal groups, nation states and other threat actors. SCADA systems for instance, being fit to purpose in the context of industrial infrastructure, are critical elements for many ICS networks to maintain operations and are targets of cyberattacks by threat actors for the exact same reason. This means that rather than rely on the practice of updating their points of contact to the external world, ICS security demands greater emphasis on monitoring and agile, light touch detection and response capabilities to contain and eradicate threats without interrupting normal operations. In some cases, when a threat is detected, and the risk of remediation is greater than the risk of acceptance, eradication efforts may be foregone for containment efforts or even acceptance. Making these determinations, however, requires asset management and visibility.

Introduction to IDS

Intrusion Detection Systems (IDS) are historically some of the most successful defensive cybersecurity tools in the fields of IT and OT cybersecurity that remains a key cornerstone of detection and response. Deployed in conjunction with a Security Incident and Event Management (SIEM) tool, which is used for comprehensive infrastructure monitoring, an IDS can be a powerful tool for threat detection and vulnerability management. Using insights gained from the visibility offered by the SIEM as well as any domain knowledge or threat intelligence

acquired through previous education or from any of the standard TI platforms, an incident responder in the context of ICS Cybersecurity could use an IDS to flag relevant attack patterns. This is typically done using signature-based detection methods that dealt in the realm of explicit rule definition, detection and in the case of an IDS/IPS, event handling. Such systems can leverage unique features of the environment that have been marked and classified by the community as being applicable to the direct defense from the attack vector in question. An IDS can also be anomaly-based as opposed to signature based, with its own handful of accompanying pros and cons that will be touched upon shortly. Before getting to Anomaly-based IDS (AIDS) however, we will first take note of one other dimension of distinction.

In addition to signature and anomaly approaches to IDS these systems can also classified as one of two types: the network intrusion detection system (NIDS), and the host-based intrusion detection system (HIDS). The way in which they operate is similar but they go about their monitoring activities according to different intents and assumptions. A HIDS focuses on monitoring individual host machines and can perhaps gather more detailed information from the host it is installed and configured on. This comes at the cost however of additional overhead in terms of resource utilization and file size. In contrast, a NIDS is typically installed on something like a gateway to apply its monitoring and detection functionality to all network traffic coming in and out of a network instead of monitoring each device on an individual basis with an agent. Both HIDS and NIDS can be signature-based, anomaly-based or both with anomaly-based detection being conducted with either traditional static feature detection or via the use of machine Learning. Although, for some IDS platforms such as Snort, there exist community ruleset to guide both the signature-based and anomaly-based detection, in the context of IDS for ICS, the author proposes the use of machine learning to conduct anomaly detection. Due to the limited resource constraints on actuators found in ICS environments, the authors for this study focus on NIDS as the most viable approach to hybrid AIDS in the context of ICS.



Challenges posed by Industry 4.0

Industry 4.0, also known as the 4th Industrial Revolution, focuses on the proliferation of next generation technology like Cloud Computing, AI, Edge Computing, IoT, and IIoT. It refers to the era of pseudo-autonomous driving, VR and instantaneous feedback and scalability stemming from a vast increase in internet connected devices and sensors. In addition to breakthroughs in science and technology for medical or other purposes, this has also brought on the emergence of LLMs like OpenAI's ChatGPT, greatly enabling your average person's ability to learn and be productive, including your average threat actor. This evolution in capability comes at a time in which ICS cybersecurity is grappling with the task of updating the Perdue Model, and industrial security architecture that relied so heavily on airgaps to maintain secure operations. Keeping in mind that many workstations in ICS environments are running outdated software it presents what could be described as a cybersecurity pandora's box. With so many changes inherent in emergent technologies, the attack surface area is increasing, individual static security controls are becoming less effective and successful attacks much more impactful. With these changes comes a need for dynamic detection and response capability.

Proposed Solution: CAE over LTSM

With the rise in data science, in recent years, IDS research has centered around the consideration of a new approach to Intrusion Detection Systems that incorporates machine learning to perform anomaly detection. Although a bit slower than traditional signature-based approaches, an AIDS automatically detects malicious behavior, even if it has not been explicitly detected or flagged as such in the past through the findings of manual threat hunting efforts. This would enable the ICS network to detect novel attacks that have yet to be diagnosed, and potentially respond to and contain them in real time. To do this our paper explores the use of a composite autoencoder to fuse the capabilities of Signature-Based and Behavior-Based detection methods for improvements in anomaly detection over either of the two methods individually.

	Network-Based	Host-Based
Signature-Based	X	
Anomaly-Based	X	

X = Domain of Proposed Solution

The authors propose the use of a Composite Auto-Encoder (CAE), which is an unsupervised machine learning model used to extract latent variables and perform dimensionality reduction. This approach can produce poor reconstruction outputs when the input data is significantly different from the training data that was used to inform its capabilities. The training data is generally only comprised of "normal" data, to the extent that malicious activity is distinguishable from the recognized normal. When an autoencoder model is introduced to data that differs from the training dataset, the extent to which it differs is

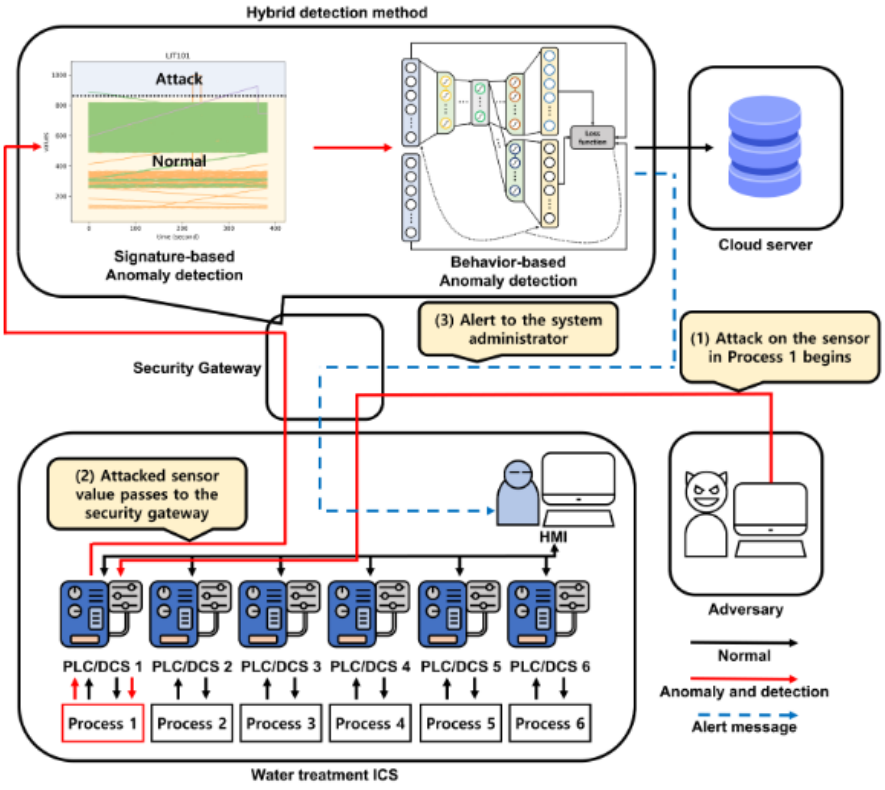
recorded, allowing for a way to set alert thresholds. This allows for a model to determine the extent to which newly received data is anomalous relative to the behavior recorded in the data it was trained on.

- Signature-Based: N Standard Deviations from 'normal' data
- Behavior-Based: Anomaly Detection from AE

Autoencoders are a type of representative unsupervised learning architecture. These models typically encode data into lower dimensional vectors. The decoder then in turn decodes latent variables and restores the original data.

In the context of this study, the statistical filter element of the composite autoencoder first classifies data as attack or normal according to its composition relative to an acceptable range of deviation from the composition of the data it was trained on. If the difference in the data at a given time exceeds a particular threshold, then it is labeled as an attack and labeled normal if the deviation is within a predetermined acceptable range. Any data flagged as normal in this first component of the hybrid detection model, in turn passes it along to the second component, the CAE, which is used as a second inspector to reduce false negatives. The CAE consists of a single encoder and two decoders which work to encode, then reproduce the data and determine whether it is malicious by learning to predict the data for the next window with the second decoder.

In line with the authors' interest in the NIDS model, the proposed hybrid detection model was contained within the security gateway of the network as demonstrated in the architectural diagram in the figure below.



Finally, in the context of LTSM, CAE gains some noteworthy properties that can be understood by first understanding Long Short-Term Memory (LTSM). An LSTM is a variant of a gated recurrent neural network. It uses a set of cells operating as a hidden vector of nodes with a weight and state intersecting a sequence of vectors. Within these cells exist 3 different gates reflecting the state of that cell: the input gate, the forget gate and the output gate. The state of one vector is represented in the gates of the hidden node in its adjacent vector to provide a means to predict future consecutive values in the sequence. This sequence prediction is what enables the CAE to reduce false negatives by predicting the state of the next iteration(s) of data.

Sample Dataset: SWaT

The SWaT dataset contains data collected on a secure water treatment system in a test bench environment which consisted of a series of PLCs that were in contact with a range of industrial processes. The metrics gathered consisted of time series data on the sensors and actuators from the PLCs and were recorded across a total of 11 days, with 7 days of normal operations and 4 days of malicious activity during which a total of 36 different cyberattacks were conducted.

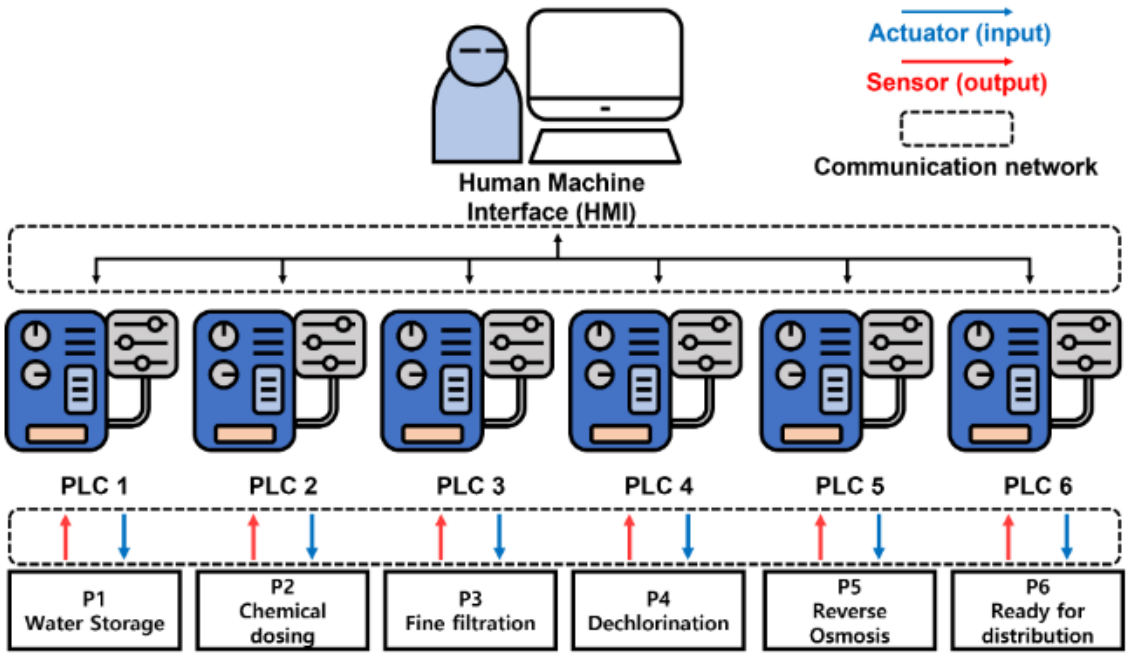


Figure 2. An example of the feedback loops in an ICS (the SWaT testbed example).

Statistical Analysis

Regarding the statistical element of the proposed methodology, in the signature-based anomaly detection process, the input is filtered based on statistical features. This is represented in the figure above by the first segment on the left-hand side of the proposed hybrid detection

method. It is used to set a benchmark for detection, filtering for malicious data involved using the standard deviation of the normal data, comprised of a time series sequence where each element in the sequence represents individual actuators is sampled at one-second intervals. In this sequence, each value d represents an actuator, and its index represents the time the recording was taken. In addition to these, however it is important to also consider the time series in the context of defined time window sizes. By adding the window size to the time index, we can mark the start of each new time window with each entry in the sequence.

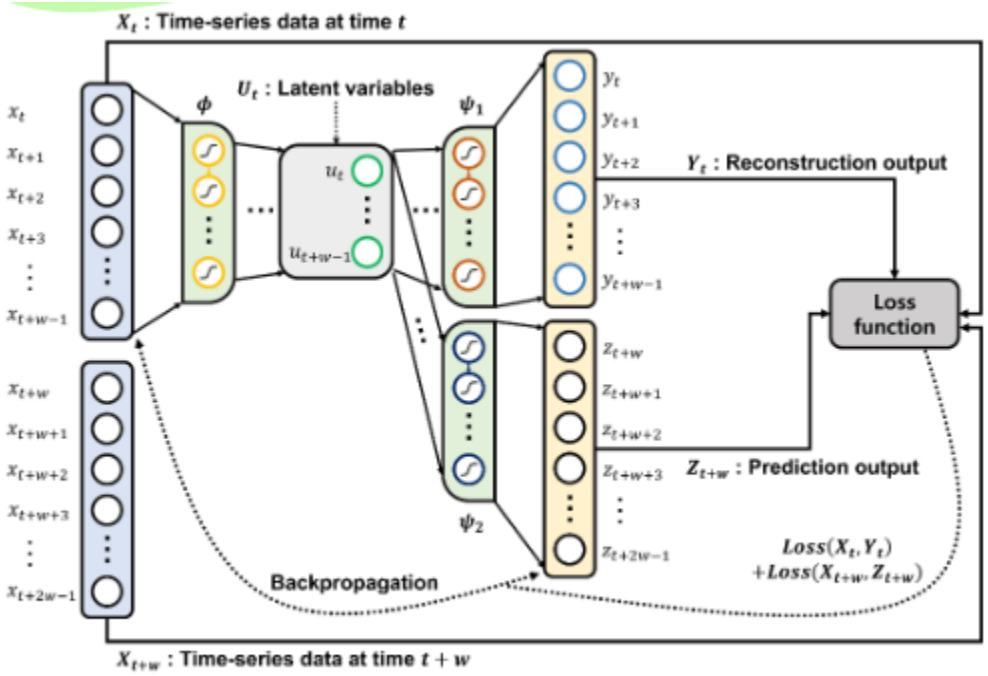


Figure 3. An example of a composite autoencoder.

Critique

The authors assume that normal and attack data are always reflected differently. However as in the case of potential insider threat, dual-use technologies like PowerShell, or lateral movement which requires the use of legitimate credentials to successfully execute, this might not always be representative of reality. This approach to classification may thus be subject to a vulnerability that will need to be addressed independently. One potential approach to mitigation would be to employ the use of manual threat hunting around particular points of concern using detection rules and an EDR/XDR agent deployed on the vulnerable endpoint. Another might be to create dummy “honey-token” accounts in any Identity Infrastructure present on a protected device such that any use of them whatsoever would alert you to the presence of a bad actor that had gone undetected by utilizing legitimate avenues. The second component of our hybrid detection model, the CAE, does mitigate against the issue of malicious activity over protected channels to an extent by picking up on the hidden associations generated by a machine learning model. However, the purpose of the CAE is described to be to reduce false negatives only, making no mention of false positives that might be significantly

different from the baseline for “normal” without necessarily having to be a cyberattack, yet this data would immediately raise an alert without being subjected to further analysis.

MITRE ATT&CK Navigator (ICS)

Conclusions

We began with a question as to whether machine learning could be used to offer any improvements to the existing processes behind IDS/IPS in the context of ICS/OT. Based on the findings from previous studies, and the research conducted in the Composite Auto-Encoder study discussed in detail herein, it seems that machine learning can in fact provide performance gains over traditional methods but with some considerations, caveats, and room for improvement.

Furthermore, the study demonstrated that statistics can be applied with and/or without machine learning to produce efficiency gains. The first example of that is the signature-based detection component of the hybrid detection method which estimated the features of the ‘normal’ training data and using its standard deviation, determined a cutoff threshold for what is to be deemed too much variation from the established baseline. When this is determined to be the case for a segment of the data it is flagged as “attack” data. The “normal” data is then sent to the second example of statistical augmentation in this study, which is the CAE, a machine learning model that was used to help reduce false negatives hiding in the “normal” data. This provides additional protection against Type II errors but not Type I errors.