Probabilistic Attack Sequence Generation and Execution Based on MITRE ATT&CK for ICS Datasets

Seungoh Choi
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
sochoi@nsr.re.kr

Jeong-Han Yun
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
dolgam@nsr.re.kr

Byung-Gil Min
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
bgmin@nsr.re.kr

ABSTRACT

To practically leverage a dataset, various attack situations should be created according to the user's objective and how realistic the generated attack sequence is should be expressed. However, there is a limit to manually generating various attack sequences that can be credible by reflecting the characteristics of the intrusion process for known cyber attacks and the field's constraints. In this paper, we propose an automatic generation method of various attack sequences that satisfy the characteristics of the attack desired by the user based on the tactics and techniques of MITRE ATT&CK, and introduce the application method through a case study. To collect the industrial control system security dataset based on the attack sequence for future work, an attack sequence executor is applied to automatically drive the attack sequence on the HAI testbed.

CCS CONCEPTS

• **Information systems → Test collections**; • **Mathematics of computing → Markov processes**; • **Computer systems organization → Embedded systems**.

KEYWORDS

attack sequence, datasets, hidden Markov model, MITRE ATT&CK

ACM Reference Format:
Seungoh Choi, Jeong-Han Yun, and Byung-Gil Min. 2021. Probabilistic Attack Sequence Generation and Execution Based on MITRE ATT&CK for ICS Datasets. In *Cyber Security Experimentation and Test Workshop (CSET '21), August 9, 2021, Virtual, CA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3474718.3474722

1 INTRODUCTION

When creating a dataset for security research in industrial control systems (ICSs), it is essential to construct and reproduce various attack situations. When leveraging a dataset for anomaly detection, both a normal dataset and an abnormal dataset should be provided for learning and evaluation. Without an abnormal dataset that includes attack-related data, it cannot be evaluated whether it trains properly or detects anomalies. Such a dataset is helpful for research on the following areas: development of data-driven defense technologies, such as machine learning; testing of various attack situations; and performance evaluation according to the desired purpose.

An attack sequence is defined as a series of attacks made by an adversary. The generation and reproduction of a sequence of attacks and their inclusion in the dataset are always a significant concern. In particular, three aspects are mainly considered when creating various attack situations for dataset development:

- **Reproducibility**: To design an attack sequence suitable for the purpose of the dataset and implement it through automation, an integrated representation that can reproduce the attack sequence is required because the representation allows to explain the abnormal dataset information.
- **Diversity**: Some attacks are possible through the use of fragmented information obtained from security threat cases and attack reports. However, it is not appropriate to create a sequence by simply enumerating or reordering attacks because generating a large number of attack sequences with specific requirements (enumerating or reordering) is difficult. Therefore, a model that can reflect more diverse attack sequences is needed.
- **Reality**: How realistic a generated attack sequence is should be expressed. The reality of an attack sequence can be used to check whether it meets the user purpose or to evaluate the performance of the attack detection system.

Although various studies on the method of generating attack sequences have been conducted, the purpose of the study and method has limitations in terms of dataset generation mentioned above. Several studies have also been conducted to envision the attacker's behavior according to its type [8, 9, 34]. However, only limited information is used to generate an attack sequence or evaluate the attacker's behavior at the network level. For adversary emulation, some studies have suggested a method of generating an attack sequence based on the pre-condition and post-condition of the attacker's behavior [20, 36]. Nevertheless, such a method is suitable for reproducing the attack sequence due to its detailed host level, but it lacks a high-level view that can compare similarities with specific attacks that have already occurred or check the possibility of individual attacks.

Hence, for creating a dataset, attack cases were analyzed from the viewpoint of MITRE ATT&CK and an attack sequence was generated based on self-defined rules [31]. However, because this method randomly generates attack scenarios based on rules, there is a scenario reality issue. In addition, another method was also presented to manually infer the attack strategy and method used in the competition dataset based on MITRE ATT&CK through the existing dataset rather than the dataset creation [21]. This method

CSET '21, August 9, 2021, Virtual, CA, USA

ACM ISBN 978-1-4503-9065-1/21/08…$15.00
https://doi.org/10.1145/3474718.3474722

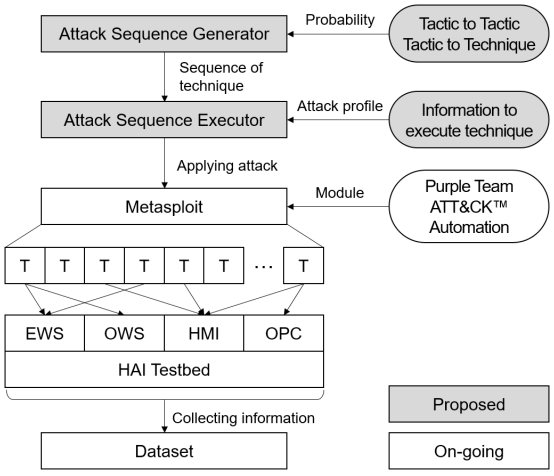

Figure 1: Proposed method to generate and execute attack sequences for the dataset

is a reverse analysis for a technique that infers the attack sequence from the provided dataset. The method for testing MITRE ATT&CK was presented, but the method considered a every single attack module rather than how to generate an attack sequence as multiple attacks for a diverse dataset [35]. The feasibility of these related works is challenging for our purpose: developing diverse dataset to benefit from generating and reproducing attack sequence.

To overcome the three difficulties in developing the ICS dataset, we implemented an attack sequence generation method based on the hidden Markov model (HMM), which is fairly capable of representing MITRE ATT&CK. Fig. 1 shows the structure diagram for proposed method to generate and execute attack sequences for the dataset. First, the attack sequence generator produces the attack sequence based on MITRE ATT&CK using probabilities as input: transition probability between tactics and the emission probability of the technique from the tactic. The characteristics of the proposed attack sequence scheme are as follows:

- **Representable attack sequence**: By adopting tactics and techniques based on MITRE ATT&CK as an attack sequence representation, attack sequence information from related incident reports can be expressed.
- **Probabilistic attack sequence**: Various attack sequences can be automatically generated by expressing the order of occurrence and probabilities of tactics and techniques in the HMM. We introduce methods to generate various attack sequences through three case studies.
- **Practical attack sequence**: We analyzed the existing ICS incident reports to determine the probability, which is a parameter used in HMM. By adjusting the order and probability of each tactic and technique, the probability of occurrence between the generated attack sequences can be compared for the user purpose: an initial tactic, a specific attack inclusion or exclusion, etc.). The transition probability determines the order of the individual attacks constituting the attack sequence. Namely, the order of individual attacks can vary by adjusting this probability.

Lastly, the attack sequence executor performs the sequence generated from the attack profile, of which information is required as a user input to execute each technique. We are developing the attack sequence executor with the Purple Team ATT&CK Automation[1] module, which can automatically emulate MITRE ATT&CK tactics and techniques through *Metasploit*.

We used a HAI testbed that had a target to reproduce the attack sequence and was able to collect data related to it. HAI testbed includes real ICSs widely used in critical infrastructure, operating components such as engineering workstation (EWS) and human-machine interface (HMI), and a log server that collects data generated in the testbed. Accordingly, a dataset can be obtained from the HAI testbed environment [6, 7, 23, 24].

The remainder of this paper is organized as follows: Section 2 describes the MITRE ATT&CK referenced for creating an attack sequence. Section 3 describes the method we proposed for creating an attack sequence. Section 4 introduces the process of attack sequence reproduction. Section 5 presents the preliminary results of the attack sequence generated according to the user's purpose. Finally, Section 6 concludes the work and discusses the scope of future research.

2 BACKGROUND: MITRE ATT&CK FOR ICSS

MITRE ATT&CK for ICSs (hereinafter referred to as ATT&CK for ICSs) compiled security threats related to control systems from the viewpoint of an attacker. ATT&CK for ICSs established components such as tactics, techniques, and data sources by referring to attack cases targeting actual control systems and analysis reports issued by related organizations such as the SANS Institute, Cybersecurity and Infrastructure Security Agency (CISA), and National Institute of Standards and Technology (NIST) [2].

ATT&CK for ICSs[2] constructed a matrix by mapping a total of 96 types (81 types when excluding duplicates) with 11 types of tactics. The tactics of ATT&CK for ICSs are inherited from ATT&CK for enterprise excluding "Privilege Escalation," "Credential Access," and "Exfiltration." [28] New tactics specialized in the control system operating environment, i.e., "Inhibit Response Function" and "Impair Process Control," were added. As shown in Fig. 3 at Appendix A, the techniques of ATT&CK for ICSs cross-reference the 29 techniques of ATT&CK for enterprises, 67 new techniques of ATT&CK for ICSs have been added. In particular, all of the new tactics consist of new techniques.

3 ATTACK SEQUENCE GENERATOR

We define an attack sequence as representing the execution order of techniques. When we create a set of attack sequences required to be differently generated as varying user objectives, ATT&CK for ICSs is feasible to reflect user requirements. ATT&CK for ICSs provides a framework that can express various attacks: existing incident cases and security threats.

[1] https://github.com/praetorian-code/purple-team-attack-automation
[2] We referred to the initial version of ATT&CK for ICSs in which is provided in Appendix A. We leveraged the initial version is ATT&CK for ICSs, which was extensively edited on 29 April 2021 while working on this paper.

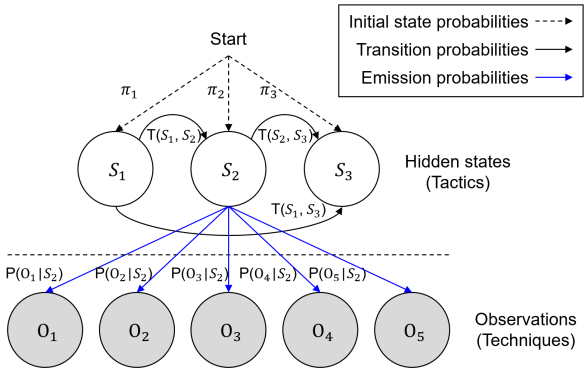

Figure 2: Elements of the first-order HMM

3.1 Proposed Method

We proposed the method of generating attack sequences using the HMM to represent user purpose. We leverage the first-order HMM[3] as shown in Fig. 2, which can explain the characteristics of this adversary behavior based on ATT&CK for ICSs, and the assumptions can be replaced as follows [22]:

- The tactic (x_t) used by the attacker at the current time (t) is only affected by the tactic (x_{t-1}) used by the previous ($t-1$). (i.e., Markovian property)
- The technique (y_t) observed at the current time (t) is affected only by the tactic (x_t) at the current time (t).

When the HMM is applied to ATT&CK for ICSs, a set of tactics is involved (11 states, S), and a set of attack technologies (81 observations, O) can occur from each tactic. The following parameters are used as the input values of the HMM:

- **Initial state probability (π)**: probability of starting each tactic
- **Transition probability (T)**: probability of movement between each tactic
- **Emission probability (E)**: probability of the occurrence of the technique included in each tactic

From the perspective of ATT&CK for ICSs, the adversary executes an attack in the direction of "Impact" from "Initial Access" tactic to compromise a target. When an attack sequence is expressed, the state and observation based on the HMM can be interpreted as follows: First, because the "Impact" state is the final stage targeted by the attacker, transition to another state is not probable. In other words, once the final state is reached, the single attack sequence is considered complete. Second, depending on the failure or reuse of the technique at the current stage, multiple techniques can be used within the same tactic or the same technique can be retried. We represent this situation as a self-transition. Finally, to consider an attack sequence where the final tactic 'Impact' is not reached, the length of the sequence (i.e., the number of state transitions, including the initial state) can be limited. Unless the final state does reach "Impact," it is regarded as an attack sequence in which the attackers failed to achieve their purpose.

[3]In the hidden semi-Markov model, the transition between states is determined by the time spent in the current state. Therefore, the values observed during the time spent in the corresponding state are not suitable in terms of the actual attack sequence.

Table 1: Related materials of ICS incidents

Type	Name (Incident)	Materials
Malware	• Stuxnet (Iran nuclear facilities)	[11, 16]
	• BlackEnergy3, Industroyer (Ukraine power grid)	[3, 15, 18, 29]
	• Triton (Saudi Arabia petrochemical plant)	[5, 18]
	• Duqu	[30]
	• Flame	[25]
	• BlackEnergy (KillDisk)	[12]
	• ACAD/Medre.A	[10]
	• Backdoor.Oldrea (HAVEX)	[18]
	• Conficker	[4]
	• VPNFilter	[17]
	• Bad Rabbit (Ukrainian transportation)	[19]
	• LockerGoga (Norway aluminum company)	[1, 26]
	• NotPetya (Ukrainian organizations)	[32]
	• Ryuk	[13, 26]
	• WannaCry	[14, 26]
PoC	• PLC-Blaster (Worm that runs on Siemens S7 PLC)	[27]
	• SoftPLC	[33]

3.2 Generating Attack Sequence

The HMM can generate various attack sequences with its parameters. To provide HMM parameters, we analyzed not only actual ICS incidents and threats reports but also ICS vulnerabilities papers shown in Table 1. Based on the analysis results, we calculated the frequency of transition between each tactic and the frequency of observed technique according to the tactic as a probability. We aggregated each probability to derive the initial probability, transition probability, and emission probability. In particular, Table 2 show the emission probability calculated from ICS incident materials as mentioned at Table 1. By adjusting the three probabilities of the HMM parameters, it is possible to create an attack sequence reflecting on the user purpose, such as which tactic to start from and which tactic and technique to be included instead of human-generated attack sequences. Moreover, we can also generate a set of different attack sequences without changing value of parameters because the HMM is a probabilistic model. The generated attack sequence is arranged by techniques in ATT&CK for ICSs, and its length depends on the total number of transitions for tactics as the user desired.

4 ATTACK SEQUENCE EXECUTOR

4.1 Attack Profile for Execution

To reproduce attack sequences, we organized an attack profile, of which information required to execute each technique constituting the attack sequence. First, information on the attacker executing the attack technique and the victim targeting the attack is required.

Second, time information is necessary so that the technique is sequentially performed according to the timeline. Finally, various options should be defined in advance, such as commands and file paths to be used when performing the technique. The attack sequence executor can reproduce the attack by sequence with the attack profile for the pre-configuration and scheduling.

Because the attack profile must be configured according to the technique within the attack sequence, a problem may arise, such that the user-defined input range is also increased as the attack sequence lengthens. In this case, the attack execution time information and attacker and victim information, excluding information commonly used for the attack technique, can be assigned in chronological order or automated selection method through the attacker and victim candidates in the test environment.

4.2 Execution Tool

We reproduce the attack based on ATT&CK for ICSs with the Purple Team ATT&CK Automation module. In addition, we are currently developing an attack reproducing the automation tool based on *Metasploit*'s msfrpc to facilitate attack reproduction and create a dataset through various and long-term replays of attacks in the future. The main purpose of this tool is to emulate generated attack sequence for developing a dataset. The execution tool can perform an attack to HAI testbed by referring to the attack profile, which is predefined configuration information. We will leverage the attack profile for dataset label generation when collecting datasets by reproducing attack sequences.

5 CASE STUDY: ATTACK SEQUENCE GENERATION

The proposed attack sequence generation method can be applied according to various use case; First, a realistic attack sequence similar to the actual threat can be created using the parameters derived from the analysis results of the attack case of the existing control system. Second, the user can create an attack sequence while filtering for a specific purpose. Users can exclude specific attack strategies or attack techniques by changing parameters. Similarly, when a user reflects only a specific attack case in a parameter, the sequence similar to a specific attack type can be created. Finally, users can create random attack sequences by randomly setting parameters while complying with HMM probability properties. In the following subsections, we arranged the preliminary results for the attack sequence generated through the HMM according to the case. The total number of attack attempts was set to 20 for all cases.

5.1 Case I: Sequence for All ICS Incidents

Based on the analysis according to Table 1, we calculated the frequency of transition between each tactic and the frequency of the observed technique according to the tactic as a probability. We aggregated each probability to derive the initial probability, transition probability, and emission probability as the HMM parameters.

Fig. 4 at Appendix C.1 represents a graph created by reflecting the HMM parameters of the initial probability, transition probability (black solid line), and emission probability (blue solid line). Applying all emission probabilities to each state is difficult to check because almost all observations are generated, so only the observations emitted from "TA0007 (Collection)" are shown in the graph as an example. In addition, the emission probability according to state is provided in Table 2 at Appendix B.

A sequence similar to an actual attack was created based on the HMM parameters derived from the ICS incident reports. The attack sequence differed from the initial attack strategy according to the initial probability. In particular, it was divided into an attack success sequence and an attack failure sequence due to the limited number of attack attempts as shown in Fig. 5 and Fig. 6 at Appendix C.1.

5.2 Case II: Sequence for Specific ICS Incidents

We can create an attack sequence while filtering for a specific purpose. For this, we adjusted the HMM parameters for Triton malware discovered in Saudi Arabia, which was reported as a various attack techniques as shown in Fig. 7 at Appendix C.2. Although the actual attack sequence of Triton and the attack sequence automatically generated through the HMM are not identically equal, Triton similarly generated the flow of tactic and attack technique performed in each tactic as shown in Fig. 8 at Appendix C.3.

5.3 Case III: Evaluation for Possible Sequence

Using the HMM's forward algorithm addressed in Appendix D, we evaluated the randomly generated attack sequence by calculating the likelihood of a given sequence of techniques. We evaluated the Triton-like attack sequence automatically generated through the HMM and randomly generated attack sequence. Based on the evaluation results, the Triton-like attack sequence was calculated as 2.616e-09, and the random attack sequence was calculated as zero. The likelihood represents the product of the transition probability between attack sequences, which is zero even if one of the transitions has zero probability. Therefore, the probability of a random attack sequence appearing is zero or can be used as an indicator to determine whether a realistic sequence is based on a specific value. To calculate the probability of a realistic attack, the method of calculating the probability of the current attack sequence can also be improved. For example, the probability of individual tactics and techniques may have to reflect changes depending on the situation as increasing the probability for the technique that occurred once.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed an automatic attack sequence generation method based on the HMM to generate a dataset including various attack situations. We determined HMM parameters by referring to an actual case analysis and ATT&CK for ICSs to create a realistic control system attack sequence. An attack sequence was also created by filtering attack strategies and attack techniques for a user-specific purpose through parameter tuning, and an evaluation plan for an arbitrary attack sequence created by a user was also presented. Although the proposed method cannot accurately modeled and manually evolving attack patterns due to probabilistic model, this method allows to create similarly attack sequence from the viewpoint of diverse dataset. In the future, we will supplement the proposed method to generate normal and abnormal scenarios in the HAI testbed when developing a dataset, of which all areas of the ICSs from level 0 to level 2 are covered.

REFERENCES

[1] Alexander Adamov, Anders Carlsson, and Tomasz Surmacz. 2019. An Analysis of LockerGoga Ransomware. In *2019 IEEE East-West Design Test Symposium (EWDTS)*. 1–5. https://doi.org/10.1109/EWDTS.2019.8884472

[2] Otis Alexander, Misha Belisle, Miller, and Jacob Steele. 2020. *MITRE ATT&CK for Industrial Control Systems: Design and Philosophy*. Technical Papers MP01055863. MITRE Corporation.

[3] Michael J. Assante, Robert M. Lee, and Tim Conway. 2017. *ICS Defense Use Case No. 6: Modular ICS Malware*. Technical Report. Retrieved May 1, 2021 from https://www.eisac.com/cartella/Asset/00006542/TLP_WHITE_E-ISAC_SANS_Ukraine_DUC_6_Modular_ICS_Malware_Final.pdf

[4] Carissa Broadbent. 2015. *Simple steps to protect yourself from the Conficker Worm*. Retrieved May 1, 2021 from https://knowledge.broadcom.com/external/article?legacyId=tech93179

[5] Carissa Broadbent. 2020. *Telling the Full Story with the MITRE ATT&CK for ICS Framework*. Retrieved May 1, 2021 from https://cyberx-labs.com/blog/telling-the-full-story-with-the-mitre-attck-for-ics-framework/

[6] Seungoh Choi, Jongwon Choi, Jeong-Han Yun, Byung-Gil Min, and HyoungChun Kim. 2020. Expansion of ICS Testbed for Security Validation based on MITRE ATT&CK Techniques. In *13th USENIX Workshop on Cyber Security Experimentation and Test (CSET 20)*. USENIX Association. https://www.usenix.org/conference/cset20/presentation/choi

[7] Seungoh Choi, Jeong-Han Yun, Byung-Gil Min, and HyoungChun Kim. 2020. POSTER: Expanding a Programmable CPS Testbed for Network Attack Analysis. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security* (Taipei, Taiwan) *(ASIA CCS '20)*. Association for Computing Machinery, New York, NY, USA, 928–930. https://doi.org/10.1145/3320269.3405447

[8] Christopher Deloglos, Carl Elks, and Ashraf Tantawy. 2020. An Attacker Modeling Framework for the Assessment of Cyber-Physical Systems Security. In *Computer Safety, Reliability, and Security*, António Casimiro, Frank Ortmeier, Friedemann Bitsch, and Pedro Ferreira (Eds.). Springer International Publishing, Cham, 150–163.

[9] Martin Drašar, Stephen Moskal, Shanchieh Yang, and Pavol Zat'ko. 2020. Session-Level Adversary Intent-Driven Cyberattack Simulator. In *Proceedings of the IEEE/ACM 24th International Symposium on Distributed Simulation and Real Time Applications* (Prague, Czech Republic) *(DS-RT '20)*. IEEE Press, 7–15.

[10] ESET. 2012. *ACAD/Medre.A: 10000's of AutoCAD Designs Leaked in Suspected Industrial Espionage*. Technical Report. Retrieved May 1, 2021 from https://www.welivesecurity.com/wp-content/uploads/200x/white-papers/ESET_ACAD_Medre_A_whitepaper.pdf

[11] Nicolas Falliere, Liam O. Murchu, and Eric Chien. 2010. *W32. stuxnet dossier*. Technical Report. Retrieved May 1, 2021 from https://www.wired.com/images_blogs/threatlevel/2010/11/w32_stuxnet_dossier.pdf

[12] Center for Research in Cyber Security. 2016. *iTrust-Analysis-001: BlackEnergy - Malware for Cyber-Physical Attacks*. Technical Report. Retrieved May 1, 2021 from https://itrust.sutd.edu.sg/wp-content/uploads/sites/3/2016/10/itrust-analysis-blackenergy.pdf

[13] Alexander Hanel. 2019. *Big Game Hunting with Ryuk: Another Lucrative Targeted Ransomware*. Retrieved May 1, 2021 from https://www.crowdstrike.com/blog/big-game-hunting-with-ryuk-another-lucrative-targeted-ransomware/

[14] Kaspersky ICS-CERT. 2017. *WannaCry on industrial networks: error correction*. Retrieved May 1, 2021 from https://ics-cert.kaspersky.com/reports/2017/06/22/wannacry-on-industrial-networks/

[15] Kaspersky. 2020. *ATT&CK for ICS: Industroyer*. Retrieved May 1, 2021 from https://www.kaspersky.com/enterprise-security/mitre/industroyer

[16] Ralph Langner. 2013. *To Kill a Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve*. Technical Report. Retrieved May 1, 2021 from https://www.langner.com/wp-content/uploads/2017/03/to-kill-a-centrifuge.pdf

[17] William Largent. 2018. *VPNFilter Update - VPNFilter exploits endpoints, targets new devices*. Retrieved May 1, 2021 from https://blog.talosintelligence.com/2018/06/vpnfilter-update.html

[18] Yamila Levalle. 2020. *Understanding ATT&CK for Industrial Control Systems (Part II)*. Retrieved May 1, 2021 from https://dreamlab.net/en/blog/post/understanding-attck-for-industrial-control-systems-part-ii-1/

[19] Orkhan Mamedov, Fedor Sinitsyn, and Fedor Ivanov. 2017. *Bad Rabbit ransomware*. Retrieved May 1, 2021 from https://securelist.com/bad-rabbit-ransomware/82851/

[20] Doug Miller, Ron Alford, Andy Applebaum, Henry Foster, Caleb Little, and Blake E. Strom. 2018. Automated Adversary Emulation : A Case for Planning and Acting with Unknowns.

[21] Nuthan Munaiah, Akond Rahman, Justin Pelletier, Laurie Williams, and Andrew Meneely. 2019. Characterizing Attacker Behavior in a Cybersecurity Penetration Testing Competition. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. 1–6. https://doi.org/10.1109/ESEM.2019.8870147

[22] L. Rabiner and B. Juang. 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine* 3, 1 (1986), 4–16. https://doi.org/10.1109/MASSP.1986.1165342

[23] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and HyoungChun Kim. 2019. Implementation of Programmable CPS Testbed for Anomaly Detection. In *12th USENIX Workshop on Cyber Security Experimentation and Test (CSET 19)*. USENIX Association. https://www.usenix.org/conference/cset19/presentation/shin

[24] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and HyoungChun Kim. 2020. HAI 1.0: HIL-based Augmented ICS Security Dataset. In *13th USENIX Workshop on Cyber Security Experimentation and Test (CSET 20)*. USENIX Association. https://www.usenix.org/conference/cset20/presentation/shin

[25] sKyWIper Analysis Team. 2012. *sKyWIper (a.k.a. Flame a.k.a. Flamer): A complex malware for targeted attacks*. Technical Report. Retrieved May 1, 2021 from https://ioactive.com/wp-content/uploads/2012/06/skywiper.pdf

[26] SophosLabs. 2019. *How Ransomware Attacks*. Technical Report. Retrieved May 1, 2021 from https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-ransomware-behavior-report.pdf

[27] Ralf Spenneberg, Maik Brüggemann, and Hendrik Schwartke. 2016. PLC-blaster: A worm living solely in the PLC. *Black Hat Asia* 16 (2016), 1–16.

[28] Blake E. Strom, Andy Applebaum, Doug P Miller, Kathryn C Nickels, Adam G Pennington, and Cody B Thomas. 2018. *MITRE ATT&CK: Design and philosophy*. Technical Papers MP180360. MITRE Corporation.

[29] Jake Styczynski. 2016. *When the light went out*. Technical Report. Retrieved May 1, 2021 from https://www.boozallen.com/content/dam/boozallen/documents/2016/09/ukraine-report-when-the-lights-went-out.pdf

[30] Symantec. 2011. *W32.Duqu*. Technical Report. Retrieved May 1, 2021 from https://docs.broadcom.com/doc/w32-duqu-11-en

[31] Yusuke Takahashi, Shigeyoshi Shima, Rui Tanabe, and Katsunari Yoshioka. 2020. APTGen: An Approach towards Generating Practical Dataset Labelled with Targeted Attack Sequences. In *13th USENIX Workshop on Cyber Security Experimentation and Test (CSET 20)*. USENIX Association. https://www.usenix.org/conference/cset20/presentation/takahashi

[32] Microsoft Security Team. 2017. *Advanced Threat Analytics security research network technical analysis: NotPetya*. Retrieved May 1, 2021 from https://www.microsoft.com/security/blog/2017/10/03/advanced-threat-analytics-security-research-network-technical-analysis-notpetya/

[33] The Claroty Research Team. 2020. *Security flaws in software-based PLC enable remote code execution on Windows box*. Retrieved May 1, 2021 from https://www.claroty.com/2020/05/14/security-flaws-in-software-based-plc-enable-remote-code-execution-on-windows-box/

[34] Sharif Ullah, Sachin Shetty, Anup Nayak, Amin Hassanzadeh, and Kamrul Hasan. 2019. Cyber Threat Analysis Based on Characterizing Adversarial Behavior for Energy Delivery System. In *Security and Privacy in Communication Networks*, Songqing Chen, Kim-Kwang Raymond Choo, Xinwen Fu, Wenjing Lou, and Aziz Mohaisen (Eds.). Springer International Publishing, Cham, 146–160.

[35] Daniel Wyleczuk-Stern and Matt Southworth. 2019. *Lessons in Purple Teaming with ATT&CK*. Retrieved July 4, 2021 from https://www.slideshare.net/attackcon2018/mitre-attckcon-20-lessons-in-purple-team-testing-with-mitre-attck-daniel-wyleczukstern-praetorian-and-matt-southworth-priceline

[36] Jeong Do Yoo, Eunji Park, Gyungmin Lee, Myung Kil Ahn, Donghwa Kim, Seongyun Seo, and Huy Kang Kim. 2020. Cyber Attack and Defense Emulation Agents. *Applied Sciences* 10, 6 (2020). https://doi.org/10.3390/app10062140

A MITRE ATT&CK MATRIX FOR ICS

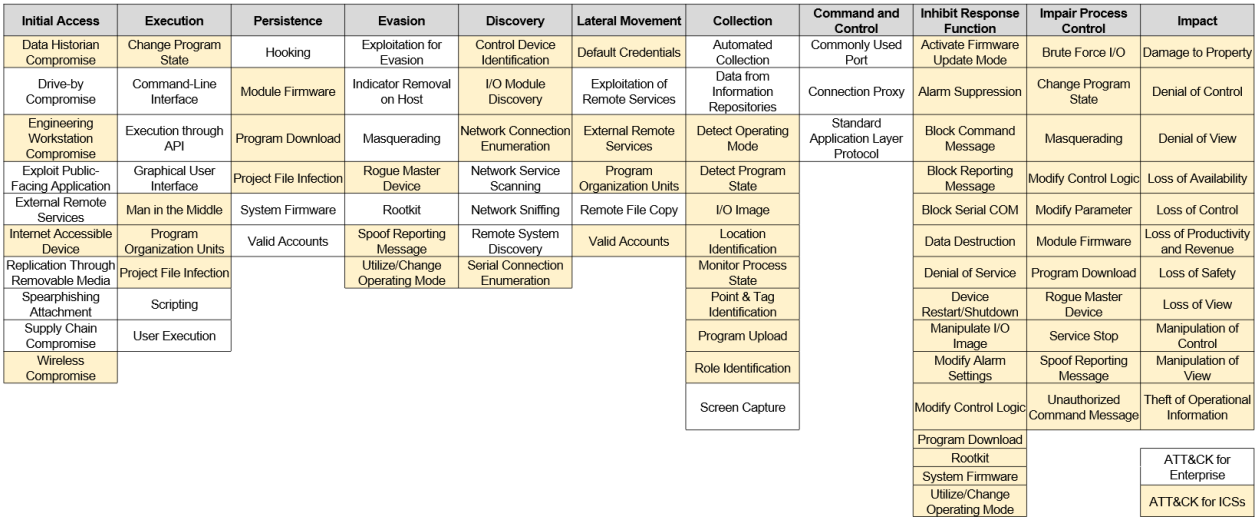

Figure 3: MITRE ATT&CK matrix for ICSs compared with matrix for ATT&CK enterprise

B HMM PARAMETERS

Table 2: Emission probability(E) derived from all ICS incidents

State	Observation		Emission	State	Observation		Emission
ID	ID	Name	Prob.	ID	ID	Name	Prob.
TA0001	T810	Data Historian Compromise	0.0769	TA0007	T802	Automated Collection	0.2500
	T817	Drive-by Compromise	0.0769		T811	Data from Information Repositories	0.1667
	T818	Engineering Workstation Compromise	0.1539		T868	Detect Operating Mode	0.0833
	T819	Exploit Public-Facing Application	0.0769		T870	Detect Program State	0.0833
	T822	External Remote Services	0.1539		T877	I/O Image	0.0000
	T847	Replication Through Removable Media	0.0769		T825	Location Identification	0.1667
	T865	Spearphishing Attachment	0.2308		T850	Role Identification	0.2500
	T862	Supply Chain Compromise	0.1538	TA0008	T885	Commonly Used Port	0.6667
TA0002	T875	Change Program State	0.1111		T869	Standard Application Layer Protocol	0.3333
	T807	Command-Line Interface	0.1111	TA0009	T800	Activate Firmware Update Mode	0.1111
	T871	Execution through API	0.3334		T803	Block Command Message	0.0555
	T853	Scripting	0.1111		T804	Block Reporting Message	0.0555
	T863	User Execution	0.3333		T805	Block Serial COM	0.0555
TA0003	T857	System Firmware	0.6667		T809	Data Destruction	0.1111
	T859	Valid Accounts	0.3333		T814	Denial of Service	0.2778
TA0004	T820	Exploitation for Evasion	0.2500		T816	Device Restart/Shutdown	0.1111
	T872	Indicator Removal on Host	0.2500		T835	Manipulate I/O Image	0.0556
	T849	Masquerading	0.2500		T833	Modify Control Logic	0.0556
	T851	Rootkit	0.1250		T843	Program Download	0.0556
	T858	Utilize/Change Operating Mode	0.1250		T857	System Firmware	0.0556
TA0005	T808	Control Device Identification	0.4286	TA0011	T813	Denial of Control	0.0526
	T840	Network Connection Enumeration	0.0714		T815	Denial of View	0.0526
	T842	Network Sniffing	0.0714		T826	Loss of Availability	0.0526
	T846	Remote System Discovery	0.3572		T827	Loss of Control	0.1053
	T854	Serial Connection Enumeration	0.0714		T828	Loss of Productivity and Revenue	0.2632
TA0006	T866	Exploitation of Remote Services	0.3750		T880	Loss of Safety	0.1053
	T822	External Remote Services	0.1250		T829	Loss of View	0.1579
	T844	Program Organization Units	0.1250		T831	Manipulation of Control	0.0526
	T867	Remote File Copy	0.3750		T832	Manipulation of View	0.0526
					T882	Theft of Operational Information	0.1053

C CASE STUDY

C.1 Attack Sequences generated by the HMM for All Incidents

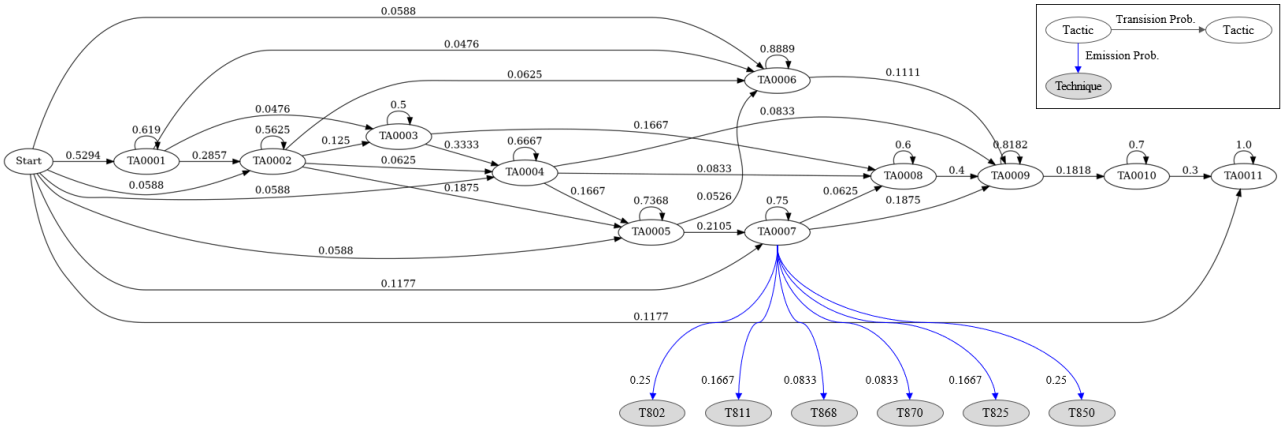

Figure 4: State and observation graph with the HMM parameters(π, T, E) for attack sequence generation

Figure 5: Success case of attack sequence generated by the HMM for all ICS incidents

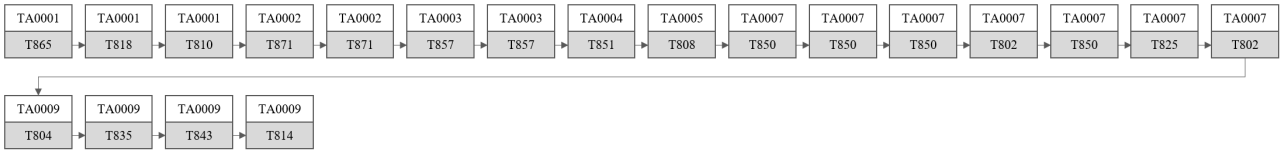

Figure 6: Failure case of attack sequence generated by the HMM for all ICS incidents

C.2 HMM for generating Triton-like Attack Sequence

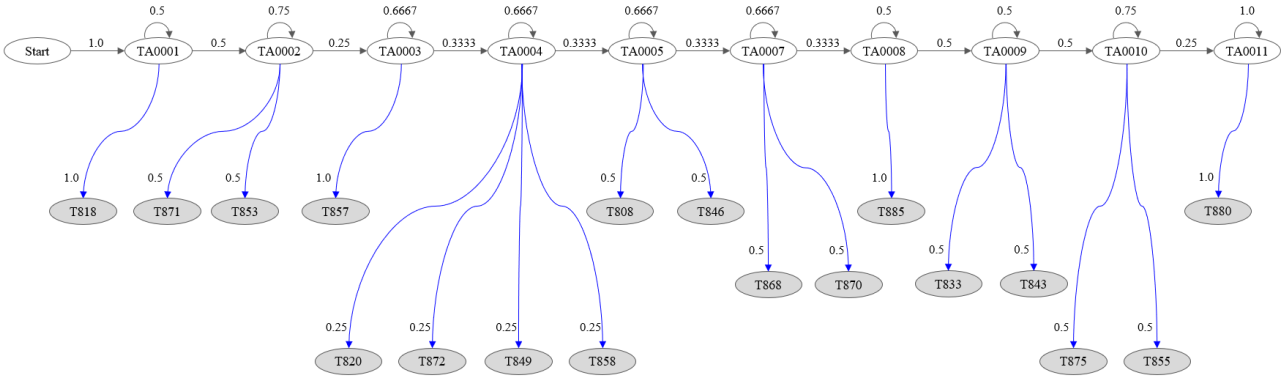

Figure 7: The HMM for generating a Triton-like attack sequence

C.3 Attack Sequence generated by the HMM for Triton

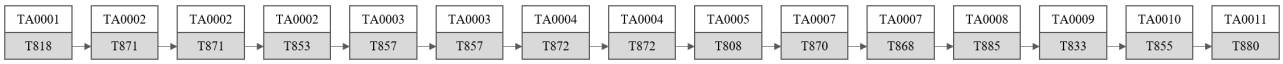

Figure 8: Success case of attack sequence generated by the HMM for Triton incident

D FORWARDING ALGORITHM IN THE HMM

$$\alpha_t(j) = \sum_{i=1}^{n} \alpha_{t-1}(i) a_{ij} b_j(o_t) \tag{1}$$

- $\alpha_{t-1}(i)$: previous forward path probability from the previous time step
- a_ij : transition probability from previous state q_i to current state q_j
- $b_j(o_t)$: state observation likelihood of the observation symbol o_t given the current state j