TEK5520 - Sikkerhet i industrielle systemer

Analyzing and visualizing the TRITON attack

Julia Vister

November 2, 2025

Contents

0.1 Q1: Attack flow diagram

The attack flow diagram (see figure 1) models the TRITON campaign C0030 from weaponization to realized impact. Each node contains a description, MITRE technique mapping, confidence rating, associated tools/artifacts (such as trilog.exe and Mimikatz) and preconditions. [1]

Figure 1: Attack flow

0.2 Q2: Extended and AI-enhanced attack flow

The initial attack flow was extended by integrating AI-enabled and automated techniques from the MITRE ATLAS framework. [3] This was aimed to explore how AI can make the attack easier or faster to execute, more stealthy or more impactful. The updated flow added specific AI-related actions, labeled AML.TXXXX, which where added as new nodes and connections. See figure 2.

Figure 2: Attack flow with AI techniques

0.3 Q3: Detection, Prevention, and Response Table

The TRITON attack in 2017 compromised SIS to potentially cause harm, and it is attributed to a Russian government-linked institution (see figure 1 for specifics). This attack involved prolonged network access, malware deployment, and reprogramming for remote takeover. [4] Based on this, the top 3 mitigations for Q1 attack flow is [2]:

1. **Multi-Factor Authentication with access controls:** This addresses initial access and credential theft, which are key entry points in TRITON's reconnaissance and persistence. For criminal threat actors, MFA deters simple phishing or credential stuffing which is common in industrial attacks. When it comes to nation-state actors, like Russia, it forces escalation to more detectable methods. The IEC 62443 controls SR* 1.1 (user id and auth.) and SR 2.1 (auth. enforcement) ensures authentication integrity and limits access to SIS controllers - which here was inadequate. For defense-in-depth strategy, implementing Security Level 2-4 (basic MFA to advanced RBAC), protects the perimeter and the core systems. This would also be low-cost and scalable.

2. **Network Segmentation with intrusion detection:** This targets reconnaissance and lateral movement. Criminals are stopped by basic zones, while nation-state actors face delayed persistence, exposing their activity. This aligns with IEC 62443 controls SR 5.1 (network segmentation) and SR 4 (data confidentiality), which restrict flows to protect SIS data. For defense-in-depth strategy, SL* 3 to SL 4 to isolate IT/OT boundaries. This should use existing firewalls, and is sufficient for SL 3-4 hig-risk zones.

3. **Regular Backups and configuration management:** This mitigates impact from reprogramming as TRITON's SIS takeover attempted. Criminals, through e.g., ransomware, are countered by restores. Nation-state sabotage is reversed, which prevents shutdown. This meets the IEC 62433 controls FR 7 (Resource Availability) SR 7.1 (DoS protection) specifically. SL2 up to SL4 adds recovery layers as a defense-in-depth strategy. This would be automated, and aligns with SL 4 maintenance schedules.

These mitigation can reduce the risk both for criminal actors and nation-states, which is consistent with Security Level 2-4's objectives. MFA presents high reduction at a low cost, segmentation leverages existing infrastructure and backups automate recovery, which can keep costs under budget while ensuring continuity.

Moving on to the analysis of the second flow from Q2 with AI extended techniques, these are the top 3 mitigations that also builds upon the ones from the first flow [2]:

1. **MFA with AI anomaly detection:** This counters AI-assisted initial access and credential actions, for example AI-generated phishing to infiltrate. Criminal actors get deterred by anomaly detection, while nation state actors face delays in reconnaissance as AI tools generate detectable behavioral patterns, potentially slowing TRITON infiltration. This enhances IEC 62443 control SR 1.1/2.1 by integrating AI-monitoring for integrity checks. Security Layer 2 to 4 combined human verification with automated alerts to protect the SIS perimeter, while being low cost and scalable with SL 4.

2. **Network segmentation with AI firewalls:** This addresses AI-enhanced reconnaissance and lateral movement, where AI fastens the mapping of the Triconex controllers or simulates evasion tactics. Criminals will be restricted by static segments and limited on low-effort scans, while nation-state's probing are neutralized by AI firewalls and pattern blocking. IEC 62443 controls SR 5.1 and FR 4 (Data Confidentiality) are strengthened with AI rules. This applies SL 3 to SL 4, and creates channels that isolates OT from IT. Cost-wise is focused on upgrading existing firewalls, but creates high return of investment.

3. **Backups with model retraining:** This mitigation counters AI persistence and impact, such as backdoor machine learning models in SIS for reprogramming. Criminals' AI-ransomware variants will be reversed via restores, and nation-states AI backdoor will be deleted through retraining, which will prevent TRITON-like shutdowns escalating. This supports IEC 62443 controls SR 7.1 and FR 7 by verifying backups against AI artifacts, and extends SL 2 up to SL 4 that layers recovery with integrity checks. However, retraining and tools is costly, but possible with off-peak scheduling to avoid operational costs.

Despite these defenses, there are still gaps where both the original and AI-extended attack flows could succeed. Missing IEC 62443 controls include SR 3.3 (security functionality verification) which is inadequate for a AI-generated malware or custom one like TRITON, and FR 7 lacks outage prevention and FR 4 is lacking of AI interference exfiltration. There are also additional security levels needed, such as SL 4 for advanced detection and AI robustness in petrochemical plants. Balancing security and operational requirements for TRITON-targeted systems involve using passive IDS to avoid shutdowns and phasing mitigations during downtime to protect safety-critical processes. This would ensure continuity with minimal disruption.

*SR = System Requirement *SL = Security Level

Note: I have combined related actions, so that each step represents a phase (e.g., all reconnaissance related actions into Step 1), and included the second attack flow with the AI-extended techniques into one table. *AI-enhanced actions (Q2 flow)* are shown in *italics*. All other columns reflect the original TRITON (Q1) mitigations. Confidence values (0–100) represent effectiveness against criminal and nation-state actors, and is unchanged from Q1. See table 1 and 2.

Step	Action (Grouped)	Detect	Prevent/Limit	Respond/ Recover	IEC 62443 Control	Confidence
1	Reconnaissance + Tool Development *(AI: Gen. AI malware, auto-recon)*	Monitor code repositories, tool downloads, encryption patterns, and network scans via SIEM.	Implement whitelisting, network segmentation, and key management.	Isolate environments, block tools, disrupt channels, and enhance monitoring.	FR 4, SR 2.1	70
2	Initial Access *(AI: Gen. phishing, prompt injection)*	Monitor VPN and authentication logs for unauthorized attempts.	Enforce MFA, patch systems, and restrict remote access.	Disable access, reset credentials, and audit logs.	SR 1.1, SR 1.2	70
3	Command and Script Execution *(AI: Gen. obfuscated scripts)*	Detect unauthorized command and script execution via logs.	Use whitelisting and limit script and privilege use.	Quarantine systems, terminate scripts, and restore backups.	SR 2.1, SR 3.1	70
4	Remote Services + ICS Scripting *(AI: Auto-protocol fuzzing)*	Detect unauthorized remote connections and ICS command patterns.	Restrict remote access with MFA VPN and limit ICS script execution.	Disconnect services, block scripts, and revert configurations.	SR 5.1, SR 7.1	72
5	Evasion + Privilege Escalation *(AI: Adaptive evasion paths)*	Monitor evasion techniques and privilege changes via logs.	Deploy IDS/IPS, enforce role-based access, and encrypt traffic.	Analyze evasion, update defenses, and revoke privileges.	SR 6.1, SR 2.1	62

Table 1: Steps 1–5: TRITON (Q1) + AI-Extended (Q2) Attack Flow

Step	Action (Grouped + AI)	Detect	Prevent/Limit	Respond/ Recover	IEC 62443 Control	Confidence
6	Credential Access + Persistence *(AI: Auto-credential harvest)*	Monitor credential access, memory patterns, and task creation via SIEM.	Harden credential storage, restrict dump tools, and limit scheduling.	Reset credentials, isolate systems, and remove tasks.	SR 1.1, SR 3.1	65
7	Data Transfer *(AI: Stealth exfil via ML)*	Detect unencrypted data transfers.	Enforce encryption for all uploads.	Contain transfer and encrypt data.	SR 4.1	75
8	Command Execution + Safety Control *(AI: Optimized SIS payload)*	Monitor command authenticity, safety manipulation, and malicious execution.	Validate commands, harden controllers, and restrict execution.	Reject invalid commands, restore settings, and halt execution.	SR 3.1, SR 6.2	70
9	Impact + Downtime Issues *(AI: Amplified fail-safe)*	Monitor fail-safe triggers and production downtime alerts.	Implement redundancy and validate test procedures.	Reset fail-safe and activate recovery plan.	FR 7, SR 6.2	67
10	Network + Controller Manipulation *(AI: Dynamic mode switch)*	Monitor unusual data flows and programming changes.	Implement traffic filtering and lock controller mode.	Block malicious flows and revert configurations.	FR 5, SR 7.1	72

Table 2: Steps 6–10: TRITON (Q1) + AI-Extended (Q2) Attack Flow

0.4 Q4: Using ChatGPT

I prompted ChatGPT with the same Question 3 (Section 3) query as i received. Below I compare ChatGPT's response to the analysis, which is based on my two attack flows from Section 1 and 2. I will focus on key elements such as the table, top 3 mitigations, gaps and balancing security and operations. Overall I would say I agree somewhat to ChatGPT's recommendations, as they align with the IEC 62443 controls and TRITON's TTPs. However, i do disagree on some details and specifics where ChatGPTS's output feels less tailored to my flow that has 10 steps.

ChatGPT's table include 12 steps, and therefore expands on mine by splitting phases, also adding AI-specific rows in the "Action" column. I find this valid for illustrating AI extension, however i do believe that it's not as accurate for direct TRITON mapping as the malware did not use AI and i prefer to structure the table so that AI and actual actions are clearly separated (something i did by having AI-techniques in italics). The other columns entries in thee table are similar, but it uses ranges for confidence level whereas i use single values. I do find that to be less precise as I score based on threat actors - high for criminals, and low for nation-states, however it is not a too great difference. The IEC control mappings overlaps as well (e.g., SR 1.1), but ChatGPT do omit some of mine making it incomplete.

I agree with ChatGPT's prioritization of network segmentation as I have included it as 2nd priority, and also MFA controls which I included as my 1st. I find ChatGPT's 2nd prioritization about change management and code signing valid, and it complements my 3rd priority about backup and configuration management. I might have under-emphasized FR 7 (Resource Availability) in my analysis while ChatGPT emphasized it, however I disagree with omitting backups entirely as TRITO's fail-safe trigger shows how much recovery is an important role. ChatGPT's cost ratings are a bit more generic than my own, but explicitly stated impact.

ChatGPT's gaps are valid and also expand on mine, and correctly identifies weaknesses in TRITON. We both also recommend SL 4 upgrades. It mentioned SR 6.1 as a gap, however i have mentioned monitoring (SIEM/IDS) in my table (Step 1). It also overlooks AI-specific gaps like model poisoning which i included in my extended flow analysis. We agree on passive monitoring to avoid shutdowns in continuous petrochemical operations, which is a big lesson from the TRITON incident. ChatGPT's layered defense mirrors my SL 2-4 strategy, however i disagree slightly on its recommendation of frequent patching as the SIS cannot do this without downtime, making my phased maintenance more practical.
All in all, ChatGPT's analysis is a good benchmark, it is broad, actionable and IEC aligned. However, it is less customized to my flow and TRITON specifics.

Note: I have included ChatGPT's output to the question in section 3 in the appendix. See Appendix A for full ChatGPT output and prompt.

Bibliography

[1] MIRE ATTCK. Triton safety instrumented system attack, 2024.

[2] CISCO. A cybersecurity framework to protect industrial automation and control systems, 2024.

[3] MITRE Corporation. Atlas matrix, 2025.

[4] Daniel Kapellmann Zafra Dan Caban Steve Miller, Nathan Brubaker. Triton actor ttp profile, custom attack tools, detections, and attck mapping, 2019.

Appendix A

ChatGPT Interaction Log and Output

This appendix contains the complete conversation with ChatGPT used for Q4 analysis.

USER: Solve this question regarding the TRITON malware incident in 2017. I will give you two attack flows, one origial and one with enhanced AI-techniques. Create a table listing all actions (nodes) in your attack flow as rows, and the following columns Step, Action, (Detection, Prevention/limitation, Respond/Recovery), IEC 62443 Control, Confidence. For each action, briefly describe the relevant detection, prevention/limitation, and response/recovery approaches, the corresponding IEC 62443 control(s), and assign a confidence value (0–100) reflecting how effective these measures are against the specific threat actors (consider two threat actors: Nation-state and Criminal). Then address the following:
-Identify your top 3 mitigations that you believe have the highest priority for defending against this attack.
-Explain why you selected each mitigation, considering both types of threat actors. Justify your selection by addressing: 1-How each mitigation addresses specific IEC 62443 security objectives. 2-The defense-in-depth strategy across different security levels. and 3- Cost-effectiveness considering the security level requirements.
-Identify gaps where the attack could still succeed despite these defenses: 1-Indicate which IEC 62443 controls are missing or inadequate? 2- what additional Security Level (SL) might be needed? 3-Reflect on how would you balance security and operational requirements, especially for safety-critical or continuously operating systems.
- Briefly discuss how your prioritized mitigations achieve an effective balance between risk reduction and implementation effort/cost, consistent with the chosen security level objectives.
Perform this analysis for both the flows you created in Q1 and Q2.

ChatGPT's response:

TRITON Malware Attack (2017) - Q3 Detection, Prevention, and Response Analysis

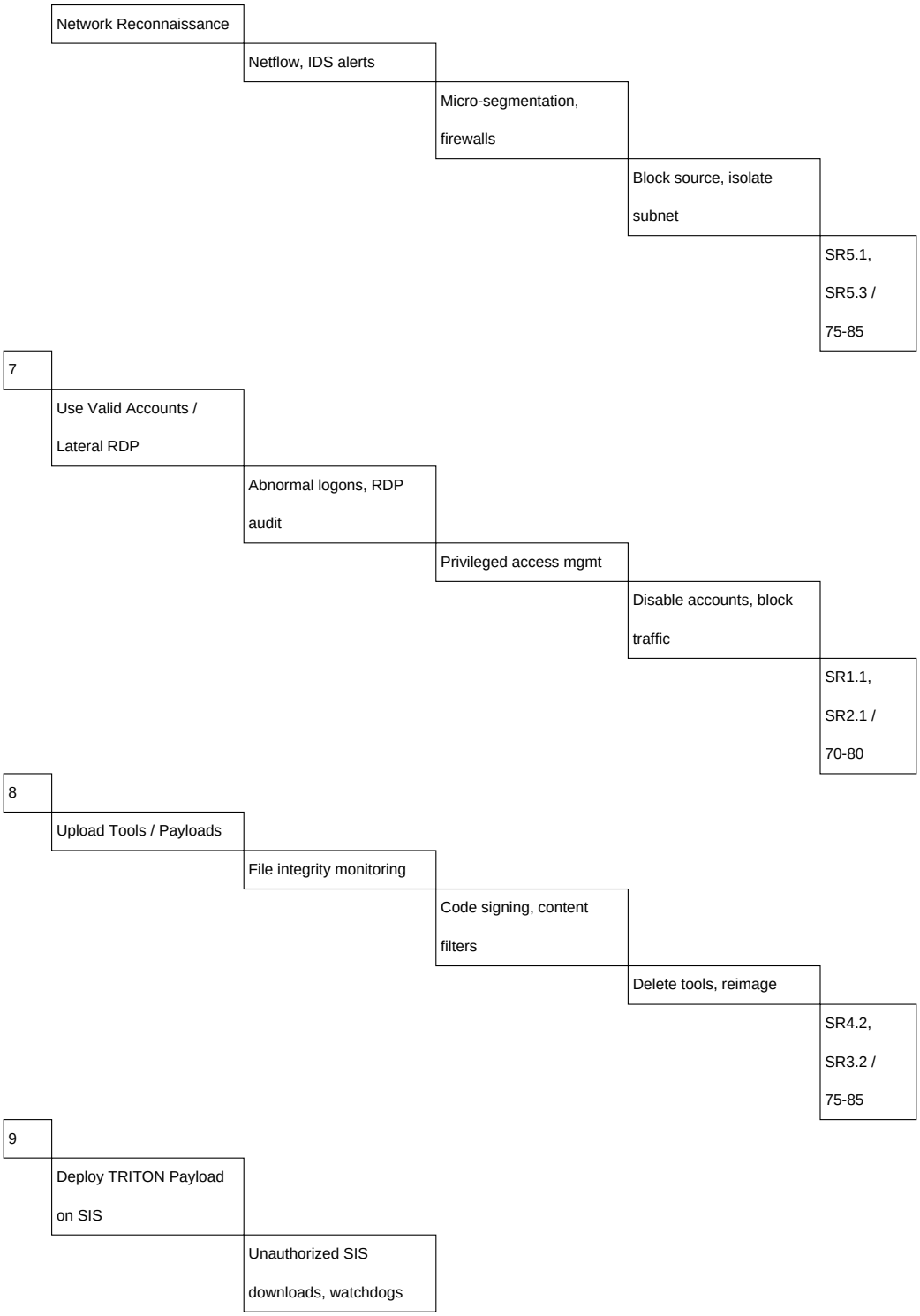
This report analyzes the 2017 TRITON (TRISIS) malware incident targeting a petrochemical plant's Safety Instrumented System (SIS). It includes a consolidated detection, prevention, and response table for both the original and AI-enhanced attack flows, mapped to IEC 62443 controls, and provides prioritized mitigations, gap analysis, and implementation guidance.

---

1. Consolidated Detection, Prevention, and Response Table

#	Action (Node)	Detect	Prevent / Limit	Respond / Recover	IEC 62443 / Confidence
1					
	Phishing / Credential Harvesting				
		Email filters, URL sandboxing, anomaly detection			
			Phishing-resistant MFA, user awareness		
				Revoke creds, reset passwords	
					SR1.1, SR2.1 / 65-80
2					
	AI Spear-phishing (Enhanced)				
		Language anomaly detection, behavioral email AI			
			Deepfake filtering, call-back verification		
				Forensic review, revoke	

			access	SR1.1, SR6.1 / 40-60
3	Credential Dumping (LSASS, Mimikatz)	EDR memory alerts, abnormal LSASS reads	AppLocker, LSA protection	Reimage host, rotate keys	SR3.2, SR2.1 / 60-75
4	Exploit Vendor VPN / Remote Access	VPN logs, auth anomalies	MFA, JIT vendor access	Disable account, audit logs	SR1.1, SR5.2 / 70-80
5	Living-off-the-Land Tools	PS logging, AMSI telemetry	Whitelisting, restricted scripting	Kill process, restore host	SR3.2, SR2.6 / 65-80
6					


Network Reconnaissance
Netflow, IDS alerts
Micro-segmentation, firewalls
Block source, isolate subnet
SR5.1, SR5.3 / 75-85
7
Use Valid Accounts / Lateral RDP
Abnormal logons, RDP audit
Privileged access mgmt
Disable accounts, block traffic
SR1.1, SR2.1 / 70-80
8
Upload Tools / Payloads
File integrity monitoring
Code signing, content filters
Delete tools, reimage
SR4.2, SR3.2 / 75-85
9
Deploy TRITON Payload on SIS
Unauthorized SIS downloads, watchdogs

			Hardened eng. station, code signing		
				Safe state, reload clean logic	
					SR7.6, SR4.1 / 85-90
10					
	Modify Safety Logic / Program Download				
		SIS change alerts			
			Dual-authorization, signed firmware		
				Restore certified image	
					SR7.6, SR1.2 / 90-95
11					
	AI Recon & Exploit Automation				
		Anomaly in scanning patterns			
			Patch mgmt, virtual patching		
				Block scans, sandbox analysis	
					SR3.1, SR5.1 / 45-60
12					
	AI Obfuscation / Polymorphism				
		Behavioral EDR detection			
			Runtime attestation,		

heuristics		
	Rollback, isolate VMs	
		SR3.2, SR4.2 / 35-55

---

2. Top 3 Priority Mitigations

---

1) Network Segmentation and Unidirectional Gateways (SR5.x)

Prevents lateral movement from IT to OT/SIS; forms the foundation of Defense-in-Depth.

Cost: Moderate | Impact: High | Effective vs Nation-state & Criminal actors.

2) Strong Change Management, Code Signing & Dual Authorization (SR4.x, SR7.x)

Protects against TRITON's core tactic of altering SIS logic. Ensures verified integrity and safe states.

Cost: Medium-high | Impact: Critical | Required for SL3-SL4 safety systems.

3) MFA and Vendor Access Control (SR1.x, SR6.x)

Prevents initial access through stolen vendor credentials, low-cost and rapid to deploy.

Cost: Low | Impact: High | Complements network and application-layer defenses.

---

3. Gaps and Security Level Recommendations

- Missing SR6.x: Supply chain and vendor assurance still weak.
- Missing SR4.2/SR7.6: Controller runtime attestation not common.
- SR6.1: OT visibility/logging insufficient in many plants.

Recommended SL: Upgrade SIS & engineering workstations to SL3-SL4.

---

4. Balancing Security and Operations

- Continuous systems cannot patch frequently: use virtual patching & passive monitoring.

- Layered defense: MFA (access layer), segmentation (network), signed logic (application).
- Prioritize high-ROI controls: MFA, vendor controls, immutable logs.

---

5. AI-Enhanced Threat Observations

- AI automation increases stealth and scanning efficiency.
- Reduces confidence of traditional signature-based detection.
- Counter with behavioral analytics, immutable logs, and stronger supply chain assurance.

---

6. Implementation Roadmap

Immediate (0-30 days): MFA, vendor access control, immutable logging.

Short-term (1-3 months): Network segmentation, engineering workstation hardening, signed SIS logic.

Medium-term (3-12 months): OT-SIEM deployment, vendor assurance, hardware attestation.

Conclusion

Layered implementation of MFA, segmentation, and SIS logic integrity controls ensures compliance with IEC 62443 SL3 objectives and creates a robust, cost-effective defense posture against both traditional and AI-enhanced TRITON-style attacks, balancing risk reduction with operational continuity.

Appendix B

Attack Flow Diagram (Q1 + Q2 Combined)

The full attack flow diagram (TRITON + AI-Enhanced) is included below in high resolution. Due to size, it is split across multiple pages for readability. The steps are counted as the blue action boxes, meaning step 1 is the first action box in the flow.

Source Files (submitted separately):

- TRITONAIflow.afb – Attack Flow Builder file

Page 1 of 3: Steps 1–7

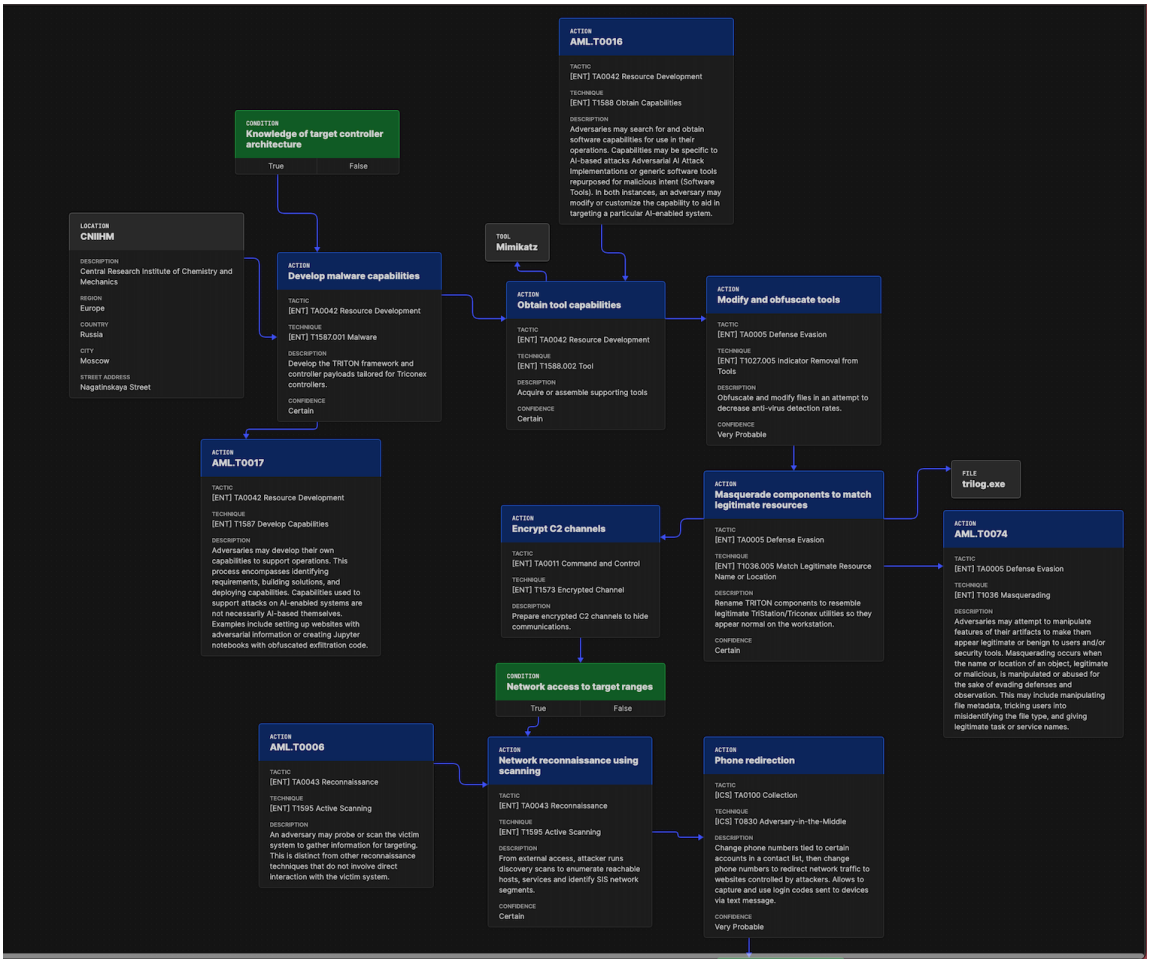


Figure B.1: Attack Flow: Steps 1–7 (TRITON + AI)

Page 2 of 3: Steps 7.5–15

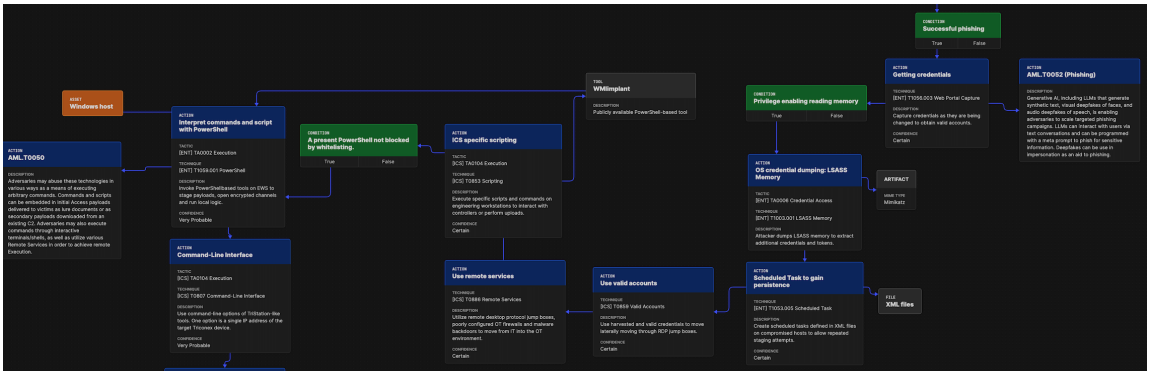


Figure B.2: Attack Flow: Steps 7.5–15 (TRITON + AI)

Page 3 of 3: Steps 15–20

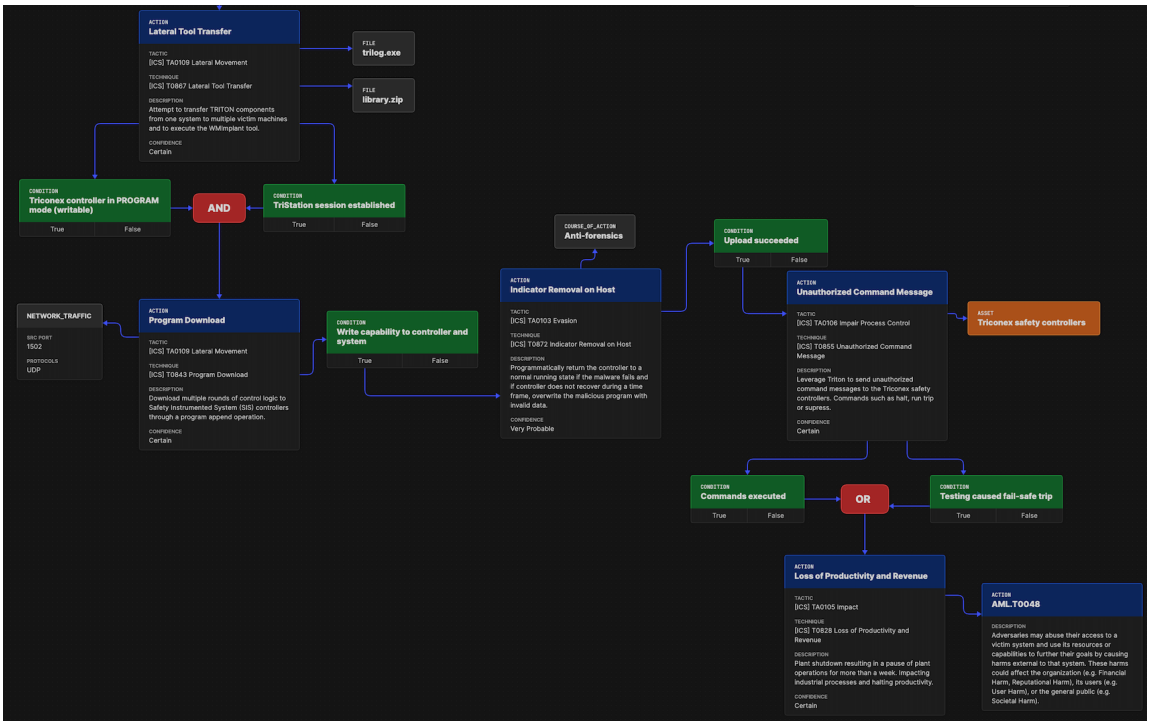


Figure B.3: Attack Flow: Steps 15–20 (TRITON + AI)