BWI GmbH stärkt ihr SOC mit einem Governed Cyber Threat Model

Case Study 1 von 2 · Automatisierte Veredelung von Cyber Threat Intelligence zu nachvollziehbaren, evidenzbasierten Attack Graphen

Ergebnisse auf einen Blick

Die Cyber Threat Modelling (CTM)-Plattform des Herstellers IdoubleS veredelt Cyber Threat Intelligence (CTI), die in natürlicher Sprache vorliegt, automatisiert zu einem **Governed Cyber Threat Model**: einem Attack Graphen, in dem jede Aussage auf ihre Textstelle im Bericht zurückführbar ist, eine Konfidenz trägt und von Analysten validiert wird, bevor sie in SIEM, EDR, Threat Hunting oder Incident Response weiterverwendet wird. „Veredeln" ist wörtlich gemeint: Aus dem Rohstoff Sprache entsteht ein strukturiertes, geprüftes, wiederverwendbares Produkt.

Diese Case Study ist die erste von zwei: Sie beschreibt die automatisierte Modellierung bei der BWI GmbH (Testszenario 2 des gemeinsamen Validierungsplans), umgesetzt von SVA als Hauptauftragnehmer – Value-Added Reseller und Integrationspartner von IdoubleS. Die zweite beschreibt die Operationalisierung des Modells – von validierten STIX-Patterns bis zu nativen SIEM-Regeln, nachgewiesen durch Adversary Emulation (Testszenarien 3 und 4).

- **Präzision gegen eine manuelle Baseline:** Die *Exakte Baseline-Übereinstimmung von Taktiken, Techniken und Prozeduren (TTPs)* erreichte **99 %** im Szenario *Emotet Delivery* und **84,17 %** im Szenario *Ryuk Ransomware* – Zielwert des Validierungsplans: mindestens 80 %.
- **Qualität der Relationen:** Die Bedeutung der *Tripletts* (Objekt–Relationship–Objekt) wurde von Experten auf der zweiten Abstraktionsebene mit **96,14 %** (Emotet) und **93,57 %** (Ryuk) bewertet.
- **Mehr als die Baseline:** Die Plattform identifizierte bis zu **51,85 %** zusätzliche, von Analysten als korrekt bestätigte TTPs, die bei der sorgfältigen manuellen Erarbeitung entgangen waren.
- **Gemessen, nicht behauptet:** Sechs automatisierte Durchläufe je Bericht, jede Aussage mit Quelle – und ein Ergebnis, das nach einer Konzeptüberarbeitung von rund 61 % auf 99 % stieg. Die manuelle Baseline benötigte je Bericht rund vier Wochen; heute prüfen und validieren die Analysten, statt Berichte manuell zu erfassen.

Kundenprofil & Ausgangssituation

Die Bundesrepublik Deutschland hat die BWI GmbH mit der Modernisierung und dem zuverlässigen und wirtschaftlichen Betrieb eines Großteils der Informations- und Kommunikationstechnik der Bundeswehr beauftragt. Gesellschafter der BWI ist zu 100 % der Bund. Als IT-Systemhaus der Bundeswehr ist die BWI ihr primärer Digitalisierungspartner in Frieden, Krise und Krieg.

Angesichts der globalen Gefahrenlage im Cyberraum baut die BWI ihre Detektions- und Reaktionsfähigkeiten im operativen Informationssicherheitsumfeld kontinuierlich aus. Cyber Threat Intelligence ist dafür der Rohstoff – und zugleich der Engpass: Sie kommt in natürlicher Sprache, in großer Menge und hoher Detailtiefe. Die BWI entschied, die Cyber-Threat-Modelling-Plattform des Herstellers IdoubleS in einem zukunftsweisenden

Projekt aufzubauen, unter realen Bedingungen zu validieren und zu betreiben – als Early Adopter. Mit der Umsetzung beauftragte sie SVA als Value-Added Reseller und Integrationspartner von IdoubleS – mit Unterstützung des Herstellers.

Herausforderung

Anspruchsvolle strategische, operationale und taktische CTI wird von Threat-Intelligence-Providern und der Cyber-Defence-Community in Berichten in natürlicher Sprache bereitgestellt. Um sich gegen relevante Bedrohungsakteure und -szenarien zu verteidigen, müssen Analysten Dutzende dieser Berichte manuell prüfen, interpretieren und die unstrukturierten Inhalte zu maßgeschneiderten Cyber-Threat-Modellen für die eigene Umgebung veredeln – ein unerlässlicher, aber wegen der intensiven manuellen Arbeit nicht skalierbarer Prozess.

Generative Sprachmodelle lesen solche Berichte heute in Sekunden. Für ein SOC entsteht damit jedoch eine neue Frage: Die Ausgabe eines Sprachmodells hat keinen dauerhaften Zustand, keine Herkunft je Aussage, keine Konfidenz, kein Schema, keinen Lebenszyklus und keine Validierung – und ein größeres Modell schließt keine dieser Lücken. Derselbe Bericht ergibt beim nächsten Durchlauf ein anderes Ergebnis; Detection Engineering, Threat Hunting und Incident Response lesen ihn jeweils erneut, in eigenen Tools und Vokabularen. Ungeprüfte Aussagen verbreiten sich mit Maschinengeschwindigkeit – und geprüft und plausibel sehen gleich aus. Die Frage ist also nicht mehr Geschwindigkeit oder Volumen, sondern Vertrauen: evidenzbasiertes, nachvollziehbares, validiertes und gepflegtes Wissen, das kein Modell von sich aus liefert.

„Früher war die Flut an CTI-Berichten manuell nicht zu bewältigen. Heute veredelt die IdoubleS-Plattform sie zu einem Modell, dem wir vertrauen: nachvollziehbar bis zur Quelle, mit Konfidenz bewertet, von Analysten validiert – in einer Qualität und Geschwindigkeit, die vorher undenkbar war.“

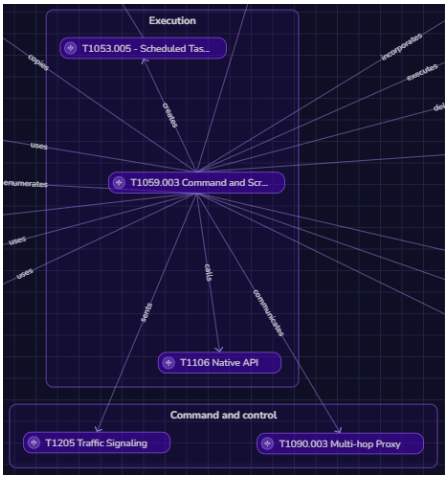
[Head of Cyber Defence Center, BWI]

Zielsetzung

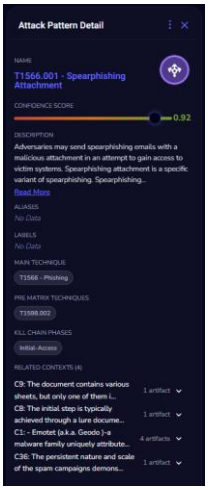
Die BWI formulierte ihre Erwartung nicht als Absichtserklärung, sondern als Validierungsplan mit messbaren Kriterien. Die Plattform sollte das Threat-Intelligence-Team dabei unterstützen, CTI automatisiert zu verarbeiten, auf den für die BWI relevanten Anteil einzuschränken, zu veredeln und zu priorisieren – und dabei vom Beta-Stadium in einer relevanten Testumgebung (Technology Readiness Level 6) bis zum Betrieb in der operativen Umgebung (TRL 9) reifen.

Für die automatisierte Modellierung wurden drei Fähigkeiten festgelegt: die automatische Erstellung bedrohungszentrischer Attack Graphen aus Berichten in natürlicher Sprache, deren manuelle Veredelung und Erweiterung sowie die Extraktion cyber-domänenspezifischer Entitäten und MITRE-ATT&CK-TTPs. Der Zielwert: Attack Graphen mit einer Accuracy von durchschnittlich mindestens 80 %, gemessen mit vorab gemeinsam definierten Metriken gegen eine von BWI- und IdoubleS-Analysten erarbeitete Baseline. Die Zusammenarbeit folgt dem Early-Adopter-Prozess des Herstellers: zweiwöchige Bewertung der Ergebnisse, priorisierte Erweiterungswünsche, nach Schweregrad eingeordnete Verbesserungen, Fehlerbehebung in Sprints und Re-Validierung von beiden Seiten – der Kunde prägt so die Roadmap des Produkts.

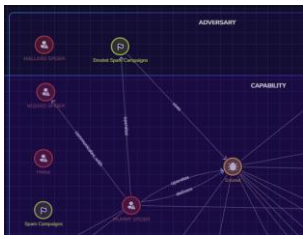
Lösung – von Sprache zu Struktur



Die Plattform des Herstellers IdoubleS ist ein offenes System, das anspruchsvolle kommerzielle oder Open-Source-CTI in natürlicher Sprache maßgeschneidert für eine Kundenumgebung auf zwei Abstraktionsebenen veredelt. Die KI-gestützte Veredelung erfolgt automatisiert per natürlicher Sprachverarbeitung (NLP) in einer domänenspezifischen Pipeline: Der Bericht wird in Mikro-Kontexte zerlegt – Absätze, Tabellenzeilen, Codeblöcke und Bilder –, die als Artefakte erhalten bleiben. Ein eigens trainiertes Transformer-Modell klassifiziert die TTPs und ordnet sie der jeweiligen Taktik zu; eine erweiterte Named-Entity-Recognition extrahiert cyber-domänenspezifische Objekte wie Malware, Infrastruktur oder Tools; die Relationship Extraction verbindet sie. Ergebnis ist ein bedrohungszentrischer Attack Graph: auf der ersten Ebene die Techniken und ihre Beziehungen entlang der MITRE-ATT&CK-Kill-Chain, auf der zweiten Ebene die Prozedur der Technik, die durch Objekte, Entitäten und Relationen das Verhalten konkret machen.



Governed heißt: Jede Aussage im Graphen trägt ihre Evidenz – die Textstelle, Tabellenzeile oder das Bild, aus dem sie stammt –, einen Konfidenz-Score, Zeitstempel und einen Validierungsstatus. Alles Extrahierte entsteht als Entwurf; ein Analyst validiert, ersetzt oder verwirft – nie der Urheber der Aussage selbst. Erst validiertes Wissen wird nach außen gegeben: als STIX-2.1-Objekte mit ATT&CK-, CAPEC-, CWE-, CVE- und CPE-Identifikatoren, über offene Schnittstellen (STIX/TAXII, APIs, MCP) – lesbar für jedes SIEM, jedes EDR und jeden KI-Agenten. Das so entstandene Modell ist ein Asset des Kunden, nicht des Herstellers: Es liegt vollständig in offenen Standards vor, lässt sich jederzeit exportieren und bleibt sein Eigentum – auch wenn CTI-Quelle, SIEM, EDR oder KI-Plattform wechseln. Kein Wissen bleibt in einem Herstellerprodukt gebunden. Und die Plattform läuft dort, wo die Daten bleiben müssen: On-Premises oder abgeschottet (air-gapped), Sprachmodelle eingeschlossen, auf Open-Source-Frameworks.



Bedrohungszentrische Attack Graphen werden Asset- und Systeminformationen in einem Attack Tree gegenübergestellt, um Risikoprofile gegenüber den Kritischen Funktionen des Kunden zu bilden – ein ganzheitliches Verständnis dafür, wie relevante Bedrohungsakteure diese Funktionen angreifen würden. Der Attack Graph bildet zugleich den roten Faden für Detektion, Threat Hunting und Incident Response: Er spiegelt den gesamten Angriffslebenszyklus eines Szenarios im Kontext der MITRE ATT&CK Matrix for Enterprise, ICS und Mobile wider und ist die Grundlage, aus der die Plattform KI-gestützt STIX-Patterns und native SIEM-/EDR-Regeln ableitet – der Gegenstand der zweiten Case Study.

Validierungsergebnisse

Ein Modell, dem ein SOC vertrauen soll, muss gemessen werden – vom Kunden, gegen eine eigene Baseline, mit vorab festgelegten Metriken. Grundlage waren zwei zusammenhängende Szenarien aus CrowdStrike-Intelligence-Berichten, deren Verwendung CrowdStrike genehmigt hat: *Emotet Delivery* (CSIT-22052) und *Ryuk Ransomware* (CSIT-21260). Analysten der BWI und des Herstellers erarbeiteten je

Bericht manuell eine Baseline – rund vier Wochen Arbeit, verstanden als eine richtige Option von vielen. Die Plattform erzeugte je Bericht sechs automatisierte Durchläufe, die auf beiden Abstraktionsebenen mit der Baseline verglichen wurden.

Quantitativ mit zwei Metriken aus dem Baseline-Vergleich – *Exakte Baseline-Übereinstimmung von TTPs* (wie viele der manuell erfassten TTPs die KI identifiziert hat) und *TTP-Ergänzungsrate* (korrekte TTPs über die Baseline hinaus, von Analysten geprüft) – und **qualitativ** mit der *Triplett-Bedeutung*: Quell-Objekt, Ziel-Objekt und Relationship werden je Triplett von Analysten auf Übersichtlichkeit, Verständlichkeit und Nachvollziehbarkeit geprüft.

Die ersten Durchläufe aus Q4/2025 verfehlten das Ziel: Die *Exakte Baseline-Übereinstimmung* lag beim Emotet-Bericht bei rund 61 %, beim Ryuk-Bericht unter 50 %; auch die zweite Abstraktionsebene blieb hinter den Erwartungen zurück. Weil das Ergebnis gemessen und nicht geschätzt war, war die Konsequenz klar: fundamentale Konzeptüberarbeitungen auf beiden Ebenen – Context-Building mit Mikro-Kontexten, Bildanalyse, ein Transformer-basierter TTP-Klassifizierungsstream mit Taktik-Differenzierung, eine erweiterte NER-Pipeline und Threat-Actor-Intelligence. Das neue Release wurde Ende April 2026 beim Kunden deployed und in Q2/2026 re-validiert:

Metrik	Emotet Delivery	Ryuk Ransomware
Exakte Baseline-Übereinstimmung von TTPs	99 %	84,17 %
TTP-Ergänzungsrate	51,85 %	15,71 %
Triplett-Bedeutung	95 %	68,31 %

Tabelle 1 – Erste Abstraktionsebene, Durchschnitt aus sechs Durchläufen

Metrik	Emotet Delivery	Ryuk Ransomware
Triplett-Bedeutung	96,14 %	93,57 %

Tabelle 2 – Zweite Abstraktionsebene, Durchschnitt aus sechs Durchläufen

Ein Prüfkriterium wurde während der Validierung neu gefasst: Statt des Aufwands für die *Nachbearbeitung* automatisiert erzeugter Graphen wird der Aufwand für die *Quality Assurance* bewertet – die KI hatte viele zusätzliche valide TTPs, Objekte und Relationen gefunden, die den Analysten entgangen waren. Konfidenz-Scores geben dafür Anhaltspunkte für eine stichprobenartige Prüfung mit moderatem Aufwand. Die weiteren Validierungsschritte – STIX-Patterns und native SIEM-Regeln – sind Gegenstand der zweiten Case Study.

Use Cases

Ein Governed Cyber Threat Model wird einmal erstellt und mehrfach genutzt – von operativen Teams in SOC und CERT ebenso wie von CISO- und Compliance-Teams:

- **Detection Engineering** – Pseudocode (STIX-Patterns) und native Detektionsregeln für SIEM und EDR aus dem validierten Graphen, mit Abdeckung je Technik.
- **Threat Hunting und Incident Response** – der Graph als strukturierter Leitfaden: Technik, Logquelle, Ereignis in der Reihenfolge des Angreifers.
- **Regulatorik** – bedrohungsgeleitete Szenarien je Kritischer Funktion als Nachweis für DORA und als Grundlage für den Threat Intelligence Report nach TIBER-EU.
- **Agentischer SOC** – KI-Agenten jeder Plattform lesen validierte Aussagen und schreiben Entwürfe zurück; das Wissen bleibt beim Kunden.