IdoubleS CTM: An automated bespoke Cyber Threat Modelling system

Rukhsar Khan and Yusuf Khan
IdoubleS Research & Development
https://www.idoubles.net/
rukhsar.khan@idoubles.net, yusuf.khan@idoubles.net

***Abstract* - Currently it requires numerous threat intelligence analysts to manually consume dozens of strategic, tactical and operational threat intelligence reports written in natural language prose to get an exact understanding of the strategy, ecosystem, attack methodology and tradecraft of relevant threat actors. The lack of an AI-driven system that would process and automate these large amounts of threat intelligence reports for generating Cyber Threat Models prevents this situation from scaling. Consequently, producing Cyber Threat Models is extremely time-consuming today and leaves the organization's assets at risk for a prolonged time. Also, the high number of analysts required makes it considerably expensive.**

To address the limitations of non-scalable manual processes, this paper introduces IdoubleS CTM (Cyber Threat Modelling), an open and automated system designed to process vast quantities of OSINT and/or commercial threat intelligence data provided in natural language. Leveraging the capabilities of generative AI, IdoubleS CTM creates bespoke, threat-, system-, and asset-centric Cyber Threat Models by integrating sophisticated cyber threat intelligence with asset data. This AI-driven approach enables the scalable and efficient generation of detailed threat scenarios, significantly reducing operational costs and resource requirements while ensuring timely and precise threat modelling.

1 Introduction

Organizations operate within specific sectors and maintain complex relationships with associated industries and regions through their supply chains, partner networks, and customers. These connections, combined with their unique value chains and brands define the products and services they offer. Underpinning these operations are people, technology, and processes, which together form the organization's assets. These assets, critical for business continuity, are continually exposed to evolving cyber threats. Identifying the mission-critical assets[1] to protect, alongside the corresponding threats that pose the highest risk to them, is a challenging task requiring bespoke Cyber Threat Models (CTMs) that are tailored to the organization's specific context.
Currently, constructing CTMs relies heavily on Threat Intelligence Analysts manually consuming and extensively analysing threat intelligence reports, written in natural language, to model the threat-centric component of the overall threat. Analysts must also identify the organization's crown jewels and the supporting systems and services to build the system- and asset-centric parts of the threat. These distinct components are then juxtaposed to create complete threat scenarios that provide actionable insights into how relevant threat actors might target the organization's assets.

Despite the growing need for scalability and efficiency in cyber threat modelling, no market-ready product exists to automate the processing of unstructured data from threat intelligence reports. As a result, analysts must manually refine this data, structure it, and align it with the system- and asset-centric elements of the overall threat. Additionally, they must manually define attack paths to represent potential relationships between attackers and the organization's assets.

This reliance on manual processes is not only time-consuming and expensive but also limits the ability to adapt to the growing complexity and volume of cyber threats. The increasing frequency of new attack techniques and the need for seamless mapping of these threats to organizational assets presents a significant challenge, generating an urgent demand for a scalable and automated solution to streamline the threat modelling process. [1] [2].

Figure 1 represents the core functionality of IdoubleS CTM, a software-based Automated Cyber Threat Modelling Platform specifically designed to assist Threat Intelligence Analysts in generating bespoke CTMs. This platform automates the generation of sophisticated threat-centric Knowledge Graphs (KG) from unstructured Cyber Threat Reports, leveraging advanced Natural Language Processing (NLP) techniques. Simultaneously, critical organizational functions are modelled in the form of Attack Trees (AT), which are derived from system- and asset-centric data. These ATs provide detailed insights into the organization's key systems, services, and potential exposures. By juxtaposing the threat-centric KGs with

[1] Throughout this document we are using the terms crown jewels, key or mission-critical assets and critical functions interchangeably. Attack surface is the sum of all the assets including their underlying systems and services.

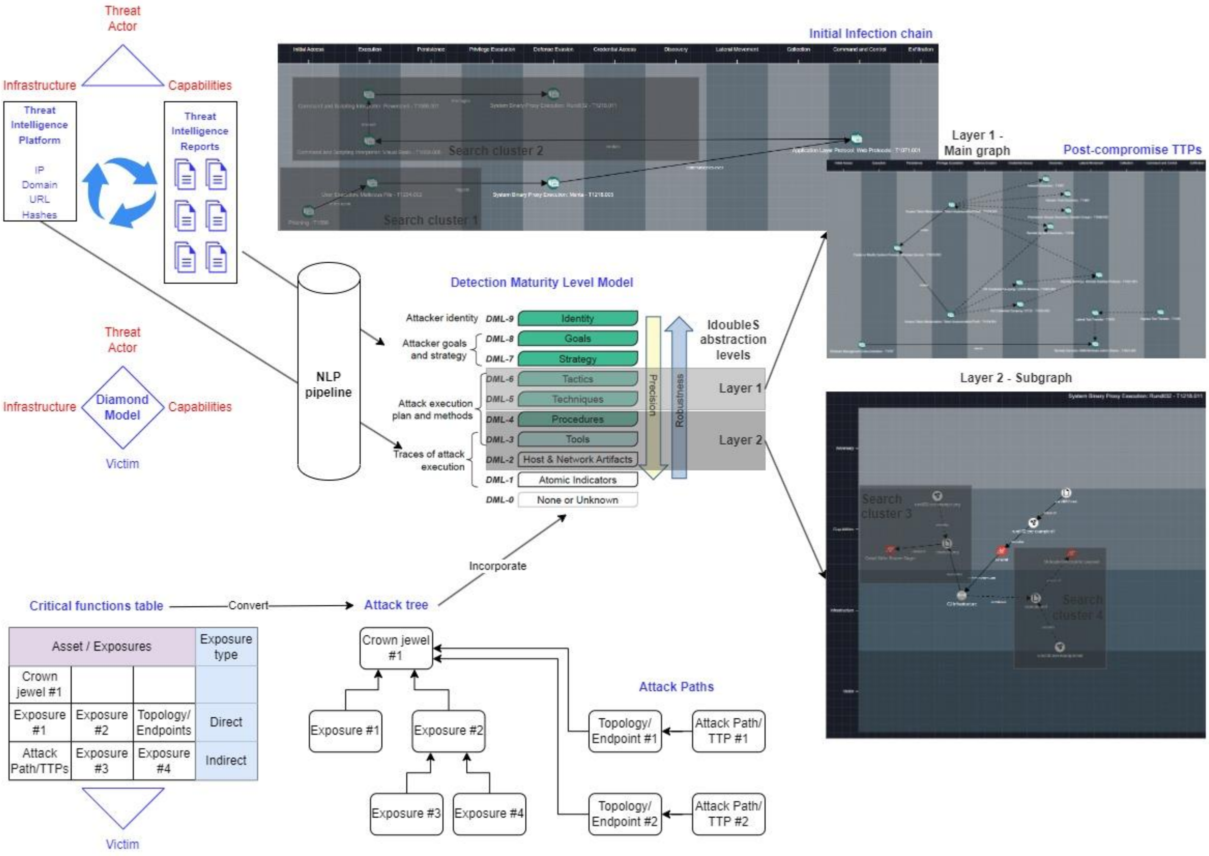

Figure 1 - IdoubleS CTM core functionality

the system- and asset-centric ATs, the platform calculates realistic Attack Paths (AP). These APs represent the pathways an adversary would expose to compromise critical assets. This integration of threat intelligence with system vulnerabilities and exposures enables the automated generation of actionable recommendations for Cyber Security teams, aiding in pre-emptively reducing exposures from the attack surface as well as prioritizing and optimizing security operations.

2 Related Work

In the paper "Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence" [3], Bromander and Mavroeidis emphasize the importance of defining ontologies for Cyber Threat Intelligence (CTI), particularly for strategic, operational, and tactical CTI, which are predominantly based on natural language prose. This aligns with earlier observations by Bromander et al. [4] in 2016, where they highlighted the abstraction levels of the Detection Maturity Level (DML) Model [5], stating: *"The further up the stack you get, the more seldom you find machine-readable results from the analysis and work that is done."*

To address this gap and enable an end-to-end cyber defense methodology, IdoubleS formalizes its ontology by integrating and harmonizing vocabularies from the following key concepts, semantics, taxonomies, and ontologies:

- The Diamond Model of Intrusion Analysis [6], referred to as the Diamond Model.
- MITRE ATT&CK Matrix for Enterprise [7], for standardized TTP representation.
- Detection Maturity Level (DML) Model [5], to map abstraction levels in CTI.
- Structured Threat Information Expression (STIX), to enable machine-readable representations.
- TIBER-EU Framework [8], for guiding the testing and evaluation of cyber resilience.

These foundational elements provide a robust basis for automated analytical reasoning at defined abstraction levels of IdoubleS, as depicted in Figure 1.

2.1 Automated CTM Generation

IdoubleS follows many of the principles outlined in the TIBER-EU framework [1] for generating CTMs. Unlike the manual natural language input and output processing introduced by TIBER-EU, which lacks programmatic utility, IdoubleS leverages automation through NLP. This automation enables the platform to understand, extract, and relate entities and objects from natural language CTI reports, generating KGs directly in machine-readable formats such as STIX, ready for immediate operationalization. This approach is comparable to SecIE [9] and RelExt [10], however, IdoubleS differentiates itself by structuring KGs into multiple layers of abstraction, primarily aligned with the DML model's DML2 through DML6 abstraction levels. This structured approach allows IdoubleS to encompass not only the threat-centric aspects of an attack but also the system- and asset-centric dimensions, creating a comprehensive, multi-dimensional representation of cyber threats.

IdoubleS advances the state of CTI by addressing critical limitations in manual methodologies. By operationalizing CTI through automated methods, it bridges the gap between natural language prose and machine-readable data, providing enhanced precision, scalability, and usability for cyber defence strategies. This layered ontology-driven approach ensures a holistic understanding of threats, systems, and assets, empowering stakeholders to derive actionable insights and implement effective defence measures.

3 IdoubleS CTM

3.1 Threat-centric Knowledge Graph structure

IdoubleS CTM leverages STIX 2.1 to represent Threat Intelligence objects as part of a structured, layered KG. The KG is organized into two distinct layers to provide a comprehensive and granular view of cyber threats. By integrating TTPs (Tactics, Techniques, and Procedures) into the first layer of the KG, IdoubleS CTM provides a systematic representation of attack patterns. The first layer nodes represent MITRE ATT&CK TTPs, while the edges define the relationships and sequences between those TTPs, offering insights into potential attack pathways. The second layer enriches this representation by detailing the cyber domain-specific entities associated within the Procedure of each TTP, creating a highly granular view of how adversaries operate in a specific campaign. Figure 1 illustrates the structure of the first and second layer of the KG. This layered approach ensures a comprehensive and actionable representation of both, high-level attack strategies and their specific implementation details.

The first layer represents TTPs employed during an attack, aligned with the MITRE ATT&CK Matrix for Enterprise [7]. This layer abstracts attack activities at a higher level, illustrating how individual techniques are linked through relationships to form an attack graph. By chaining relevant techniques into a graph, this layer provides an overview of adversary behaviour and their operational methodologies.

The second layer delves deeper into each MITRE ATT&CK technique, representing detailed Procedures primarily on a tactical rather than on a specific technical level (Indicators of compromise) to describe specific entities including their relationships involved in the attack. These entities include:

- *Threat Actors*, *Malware* and *Tools* used by adversaries.
- Artifacts such as *Hashes*, *IPs*, *Registry Keys*, *URLs*, *Domains* and *Code Snippets*.

Each entity is further aligned with the Diamond Model of Intrusion Analysis, classifying them into three of the four available Diamond Events:

- *Capabilities*: What the adversary can do (e.g., tools, malware).
- *Infrastructure*: Resources and platforms used to execute or support the attack.
- *Adversary*: The threat actor or group responsible for the attack.

3.2 Automated Knowledge Graph creation

Cyber Threat Reports serve as the primary input for the automated generation of the KG. As illustrated in Figure 2, the content of these reports undergoes a structured parsing and semantic analysis process. This process segments the reports into clusters, bundling contextual content to ensure coherent and meaningful organization.

The pipeline for creating Layer 1 of the KG includes:

- The segmented clusters are classified into TTPs through Text Classification (TC), powered by the Large Language Model (LLM) Meta Llama3:8B.
- Not all clusters can be directly classified into TTPs, as some represent contextual information that does not specifically map to a TTP. These contextual clusters are retained to enrich the broader understanding of the threat landscape.
- Relationships between the extracted TTPs are identified using Relation Extraction (RE), leveraging the capabilities of Llama. This process identifies three key entities:
  - A source TTP (origin of the relationship).
  - A target TTP (destination of the relationship).
  - The relationship type (e.g., “leads to,” “enables,” or “facilitates”).
- These relationships enable the creation of a chained attack flow, representing the sequence and dependencies of adversarial actions.
- Once TTPs are classified and their relationships are extracted, the first layer of the KG is completed. This layer is then formatted into STIX objects to align with widely recognized threat intelligence standards:
  - TTPs are represented as AttackPattern objects in STIX terminology.
  - Relationships between TTPs are encoded as STIX Relationship Objects (SROs).
- To enhance the TTP representation, the MITRE ATT&CK TAXII server is integrated, enabling the retrieval of detailed information for each AttackPattern object. This ensures that all relevant attributes and references for the TTPs are included in the KG.

This automated pipeline delivers a structured representation of the threat report content in the form of Layer 1 of the KG, which captures TTPs and their interconnections. By leveraging LLM-driven classification, Relation Extraction, and STIX formatting, the platform provides a scalable and standardized approach to modelling complex cyber threats, laying the foundation for actionable insights and further analysis in subsequent KG layers. As outlined in Figure 2 the second layer delves into the Procedures of specific TTPs, capturing the finer details of how these techniques were executed during an attack. This layer focuses on extracting and organizing cyber domain entities associated with each TTP, providing a granular representation of adversarial activities.

The pipeline for creating Layer 2 of the KG includes:

- Using Named Entity Recognition (NER) powered by the LLM and Regular Expression (REGEX), relevant cyber domain entities are identified and extracted from the raw data.

- The extracted cyber domain entities are connected by identifying relationships between them using RE. This step uncovers:
    - Interactions: How one entity influences or depends on another (e.g., a File executing a Process or a URL delivering a Malware sample).
    - Causal Links: Sequence or dependencies between entities (e.g., a Registry Key modification triggered by a Process).
- These relationships are essential for creating a detailed graph that contextualizes the role and behaviour of each entity within the specific TTP.
- All cyber domain entities associated with a specific TTP are translated into their corresponding STIX Domain Objects (SDOs) and STIX Cyber Observable Objects (SCOs). Entities that cannot be directly mapped to predefined SDOs or SCOs are converted into Custom STIX Objects, ensuring that all relevant information is represented in the KG. For example, code snippets are mapped to a custom object type, X-Code, designed specifically for this purpose.
- The relationships between SDOs and SCOs are captured using SROs, which define the nature of the connections between entities (e.g., "related-to," "indicates," or "derived-from"). These relationships create a structured and actionable representation of how entities interact within the context of the TTP.

3.3 STIX Mapping

The Layer 1 and Layer 2 KGs are represented in separate STIX bundles, ensuring modularity and clarity. Each bundle encapsulates the relevant STIX objects and their relationships.

The Layer 1 bundle focuses on representing AttackPattern SDOs along with their corresponding SROs, encapsulated within the *objects* parameter. During the TC process, when a cluster is assigned a TTP number (e.g., T1210), a database lookup retrieves comprehensive TTP details from previously loaded data obtained via the MITRE ATT&CK TAXII server. This step is essential because the TC process outputs only the TTP number, without additional descriptive information. Each cluster is represented by its own AttackPattern object, which resides within a dedicated STIX bundle forming a unique sub-graph. This design ensures that every classified cluster preserves its distinct relationships and contextual data within its associated STIX bundle.

The Layer 2 bundle extracts cyber domain entities using NER and REGEX processing. REGEX proves particularly effective for extracting predefined entity formats such as IP addresses, hashes, Registry Keys, and Domain names, owing to their structured and predictable nature. These precise extraction techniques enrich the representation of Layer 2, providing a detailed, granular view of the entities associated with each AttackPattern object. The relationships between these objects, defined in SROs, will also chain the cyber domain entities.

The following list demonstrates the mapping between cyber domain entities and their corresponding STIX objects (including SDOs, SCOs, and custom STIX objects):

Entity	Mapped STIX Object
Threat Actor	STIX SDO: Threat Actor
Malware	STIX SDO: Malware
Tool	STIX SDO: Tool
MD5	STIX Custom Object: Hash
SHA256	STIX Custom Object: Hash
IP Version 4	STIX SCO: IPv4 Address
IP Version 6	STIX SCO: IPv6 Address
Registry Key	STIX SCO: Windows Registry Key Object
URL	STIX SCO: URL Object
Domain	STIX SCO: Domain Name Object
Code Snippet	STIX Custom Object: Code

Table 1 - STIX mapping for cyber domain entities

To ensure data integrity and avoid duplication, all entities extracted through NER and REGEX are consolidated. This involves merging similar or identical values into a single STIX object. The current implementation does not yet include all possible STIX objects. As part of planned improvements, additional entities and their corresponding STIX objects will be supported and integrated into Layer 2 in the future.

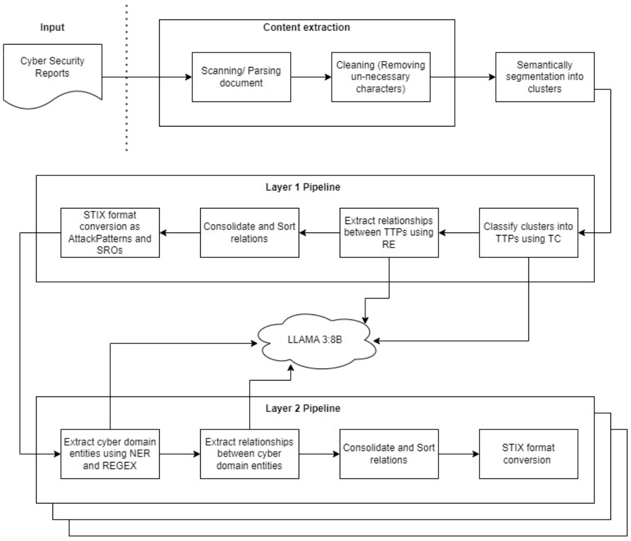
Input
Cyber Security Reports
Content extraction
Scanning/ Parsing document
Cleaning (Removing un-necessary characters)
Semantically segmentation into clusters
Layer 1 Pipeline
STIX format conversion as AttackPatterns and SROs
Consolidate and Sort relations
Extract relationships between TTPs using RE
Classify clusters into TTPs using TC
LLAMA 3:8B
Layer 2 Pipeline
Extract cyber domain entities using NER and REGEX
Extract relationships between cyber domain entities
Consolidate and Sort relations
STIX format conversion

Figure 2 - Layer 1 and Layer 2 processing pipeline

4 Results and Evaluation

To evaluate the effectiveness of the automated KG generation process, we visualized the output using the Open-Source *CTI STIX Diamond Activity Attack Graph* visualizer. For this test, we used the public Threat Report *#StopRansomware: LockBit 3.0*, published by the Cybersecurity & Infrastructure Security Agency (CISA) [11]. This report had previously been manually modelled into a KG by Cyber Threat Intelligence Analysts, providing a

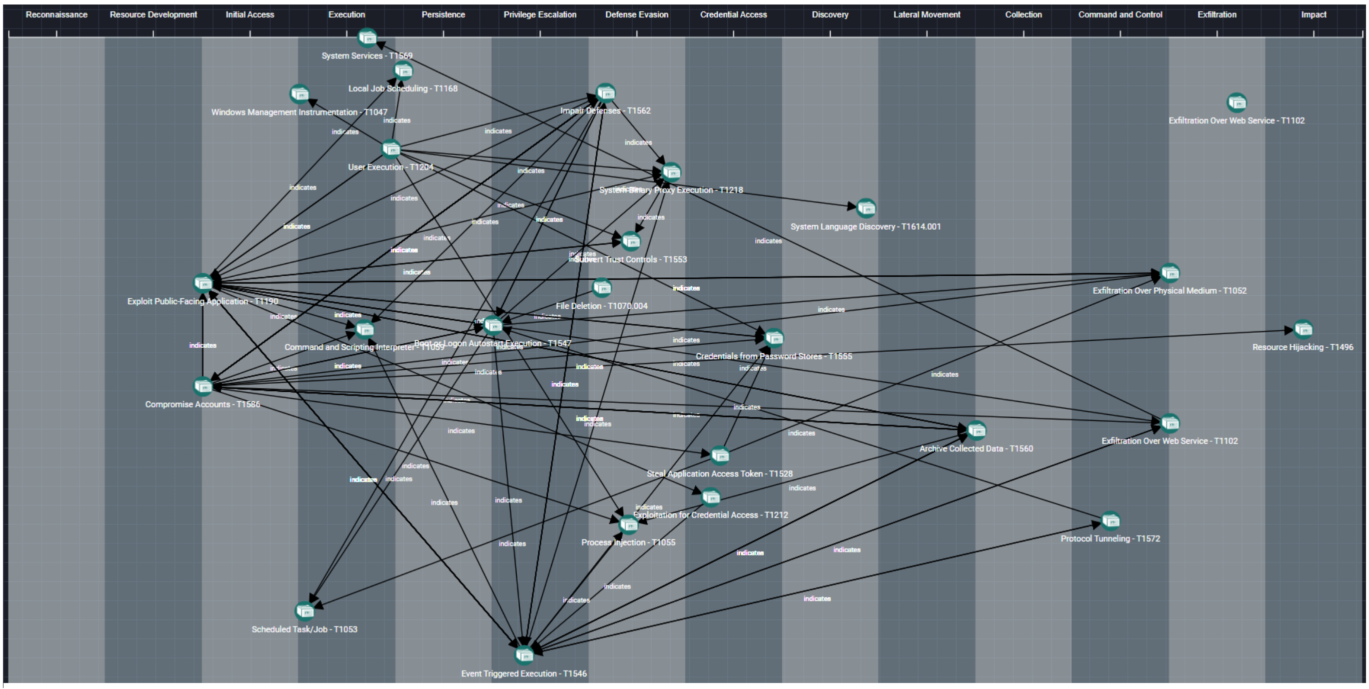
Reconnaissance
Resource Development
Initial Access
Execution
Persistence
Privilege Escalation
Defense Evasion
Credential Access
Discovery
Lateral Movement
Collection
Command and Control
Exfiltration
Impact
System Services - T1569
Local Job Scheduling - T1168
Windows Management Instrumentation - T1047
Impair Defenses - T1562
Exfiltration Over Web Service - T1102
User Execution - T1204
System Binary Proxy Execution - T1218
System Language Discovery - T1614.001
Subvert Trust Controls - T1553
Exploit Public-Facing Application - T1190
Exfiltration Over Physical Medium - T1052
File Deletion - T1070.004
Command and Scripting Interpreter - T1059
Boot or Logon Autostart Execution - T1547
Credentials from Password Stores - T1555
Resource Hijacking - T1496
Compromise Accounts - T1586
Archive Collected Data - T1560
Exfiltration Over Web Service - T1102
Steal Application Access Token - T1528
Exploitation for Credential Access - T1212
Process Injection - T1055
Protocol Tunneling - T1572
Scheduled Task/Job - T1053
Event Triggered Execution - T1546
indicates

Figure 3 - AI processed KG

reliable baseline for comparison.

The evaluation compared the manually constructed KG to the AI-generated KG. Figure 3 illustrates the output of the automated processed KG. Below are observations and feedback provided by Cyber Threat Intelligence Analysts regarding the results of the AI-generated KG:

- Layer 1: Some AttackPattern objects were duplicated in the graph. These duplicates could have been merged into a single AttackPattern object, improving clarity and reducing redundancy.
- Layer 1: Certain relationships between AttackPatterns were incorrectly generated as bidirectional. These should be corrected to reflect their appropriate unidirectional nature, ensuring consistency with the logic of the threat model.
- Layer 1: The accuracy of the TTP classification needs improvement. Some AttackPatterns generated by the AI were not relevant to the threat scenario, highlighting the need for further refinement in the text classification process.
- Layer 2: During the NER process, some entities were mapped to the wrong STIX object types. For instance, specific cyber domain entities such as IP addresses or file hashes were misclassified, which impacts the precision of the detailed representation in Layer 2.

The evaluation demonstrates the potential of the automated KG generation process but also highlights areas requiring improvement:

- Redundancy Reduction: The handling of duplicate objects and bidirectional relationships should be optimized to ensure a clean and logical graph structure.
- Classification Refinement: The enhancement of the LLM for TTP classification could increase the relevance and accuracy of the generated AttackPatterns.
- Entity Mapping: Improvements in the NER and STIX object mapping process are essential to ensure the correctness of Layer 2 representations.

Despite these limitations, the automated system successfully produced a KG with considerable alignment to the manually modelled baseline. With targeted refinements, the system can significantly enhance efficiency and scalability in CTM generation, reducing reliance on time-intensive manual processes. This evaluation establishes a roadmap for further development and optimization to achieve higher accuracy and operational reliability in future iterations.

5 Future Work

The evaluation results have highlighted areas where the accuracy of TC, RE, and NER need to be enhanced to improve the overall quality of the generated KG. To address these challenges, we have initiated a dedicated project focusing on data collection, preparation, and formatting. This project aims to curate high-quality datasets specifically tailored for fine-tuning the LLM, enabling better context understanding and output precision. By refining the underlying model, we aim to significantly improve the accuracy and relevance of the generated KGs.

Additionally, we are actively working on integrating and optimizing ATs and the automated calculation of APs to identify potential pathways adversaries might exploit to compromise critical infrastructure. This involves analysing and defining common denominators to align and juxtapose the threat-centric KG (Layer 1 and Layer 2) with the system- and asset-centric AT. This integration will enable the seamless identification of attack paths, providing actionable insights into how attackers may navigate through an organization's systems to target critical assets.

Future iterations will focus on:

- LLM Fine-Tuning: Enhancing model capabilities for TC, RE, and NER by leveraging domain-specific datasets.
- AT and AP Development: Refining methodologies to calculate dynamic Attack Paths and establish stronger correlations between KGs and Attack Trees.
- Expanded Entity Support: Incorporating additional STIX objects to represent more diverse cyber domain entities and relationships.

6 Conclusion

In this paper, we introduced IdoubleS CTM, an automated CTM platform designed to address the challenges of scalability, accuracy, and efficiency in cyber threat modelling. By leveraging advanced techniques in NLP and KG generation, IdoubleS CTM enables the transformation of unstructured threat intelligence data into actionable insights. The platform automates the generation of multi-layered KGs, integrating threat-centric, system-centric, and asset-centric data to provide a comprehensive view of potential attack scenarios.

The evaluation of IdoubleS CTM demonstrated its ability to generate KGs, while also highlighting areas for improvement in TC, RE, and NER. Future developments will focus on refining the platform's capabilities, particularly in improving model accuracy, expanding entity support, and juxtaposing asset- and system-centric ATs to KGs.

References

[1] "TIBER-EU Guidance for Target Threat Intelligence Report," European Central Bank, https://www.ecb.europa.eu/paym/cyber-resilience/tiber-eu/shared/pdf/Final_TIBER-EU_Guidance_for_Target_Threat_Intelligence_July_2020.pdf, 2020.

[2] "TIBER-EU Scope Specification Template," European Central Bank, https://www.ecb.europa.eu/paym/cyber-resilience/tiber-eu/shared/pdf/Final_TIBER-EU_Scoping_specification_template_July_2020.pdf, 2020.

[3] S. Bromander and V. Mavroeidis, "Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence," https://arxiv.org/pdf/2103.03530.pdf, 2017 (revised in 2023).

[4] A. Jøsang, S. Bromander and M. Eian, "Semantic Cyberthreat Modelling," http://folk.uio.no/josang/papers/BJE2016-STIDS.pdf, 2016.

[5] R. Stillions, "The DML model," http://ryanstillions.blogspot.com/2014/04/the-dml-model_21.html, 2014.

[6] S. Caltagirone, A. Pendergast and C. Betz, "The Diamond Model of Intrusion Analysis," https://www.activeresponse.org/wp-content/uploads/2013/07/diamond.pdf, 2013.

[7] "MITRE ATT&CK Matrix for Enterprise," The MITRE Corporation, https://attack.mitre.org/matrices/enterprise/, 2015-2023.

[8] "TIBER-EU framework," European Central Bank, https://www.ecb.europa.eu/paym/cyber-resilience/tiber-eu/html/index.en.html, 2018-2023.

[9] Y. Park and T. Lee, "Full-Stack Information Extraction System for Cybersecurity Intelligence," IBM T. J. Watson Research, https://aclanthology.org/2022.emnlp-industry.pdf, 2020.

[10] A. P. S. M. A. J. J. H. a. R. Z. Aditya Pingle, "RelExt: Relation Extraction using Deep Learning approaches for Cybersecurity Knowledge Graph Improvement," arXiv: 1905.02497v2, 2019.

[11] M.-I. F. CISA, "#StopRansomware: LockBit 3.0," 16 March 2023. [Online]. Available: https://www.cisa.gov/sites/default/files/2023-03/aa23-075a-stop-ransomware-lockbit.pdf.