Re-forging SPADA: guiding threat modelling and automation challenges

Mario Raciti

DRADS 2025

IMT SCUOLA ALTI STUDI LUCCA

KU LEUVEN

Università di Catania

10/03/25 – Ostend

Agenda

1. **Introduction**
2. The SPADA Methodology
3. The SPADA Language Threat Model
4. Open Challenges
5. Conclusions

“**Threat modelling** works to identify, communicate, and understand threats and mitigations within the context of protecting something of value.”

- OWASP

Introduction

Existing threat modelling methodologies face **key challenges**:

Domain adaptability

Many approaches are domain-independent and struggle with specific applications.

Completeness

Failing to account for specific threats would cause pitfalls to the subsequent risk assessment.

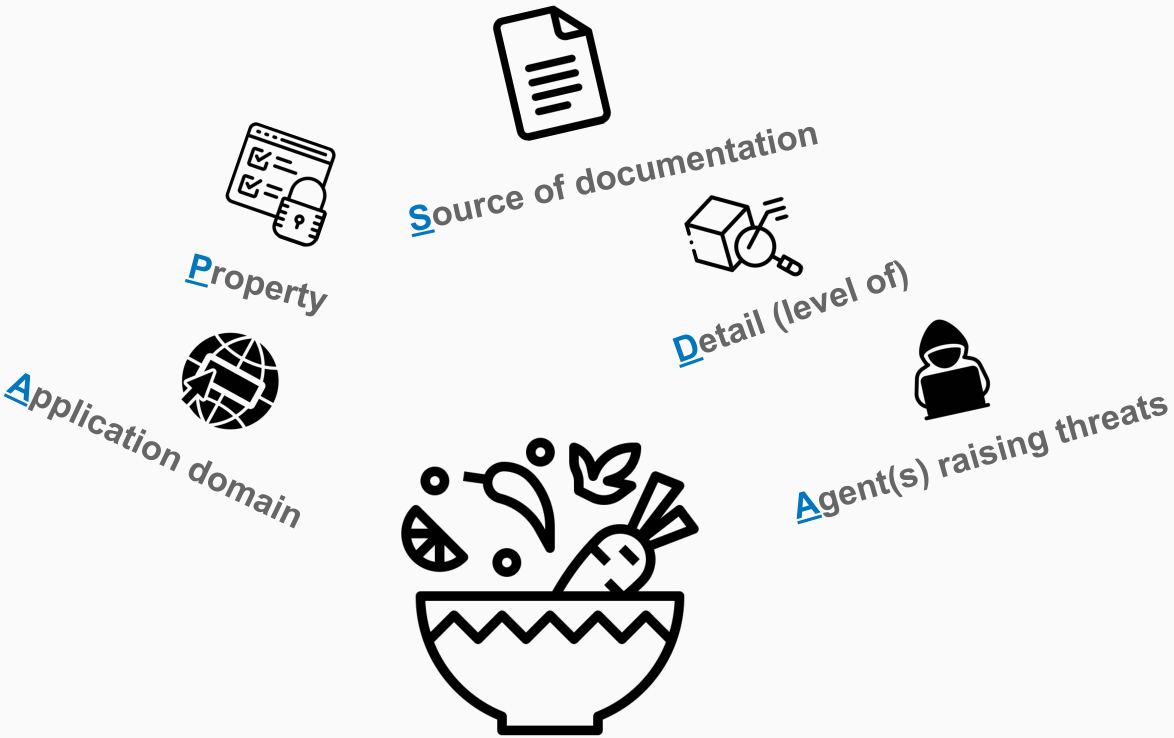
Threat Explosion

An overwhelming number of threats that may be irrelevant, infeasible, or redundant with each other.

Subjectivity

Two analysts would likely give different descriptions to the same threat (e.g., wording, style).

The Variable Elements of Threat Modelling



Application domain

Property

Source of documentation

Detail (level of)

Agent(s) raising threats

Source of Documentation

> Internal
> External
> Hybrid

<u>Examples:</u>

- A list written by the analyst
- OWASP, best-practice docs
- A mix of the above

It also provides the means to keep track of the version of the threats, e.g., the year in which the specific threat list is published.

Property

> **Hard Privacy**
> **Soft Privacy**
> **Cybersecurity**

Examples:

- LIND
- UN
- STRIDE

It helps to focus on threats targeting that/those specific property/properties.

Application Domain

> Domain-Dependent
> Domain-Independent

Examples:

- Smart cars, smart home
- Universally applicable threats

A combination of the two approaches may offer a more effective and efficient analysis.

Detail (Level of)

> Higher / Detailed
> …
> Lower / Abstract

<u>Examples:</u>

- Hyponyms/Meronyms
- …
- Hypernyms/Holonyms

A higher level of detail implies an estimation of the likelihood for a given threat with more precision. The most appropriate level of detail should be considered within the main picture.

Agent(s) raising Threats

> Attacker
> Data processor
> Data controller
> Third party

<u>Examples:</u>

- A cybercriminal
- A cloud service provider
- A social media platform
- A marketing analytics firm

TAs may also be considered in combination.

What if we mix these ingredients together?

Agenda

1. Introduction
2. **The SPADA Methodology**
3. The SPADA Language Threat Model
4. Open Challenges
5. Conclusions

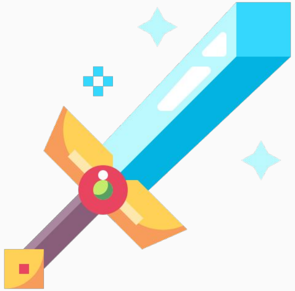
What is SPADA?

SPADA is a methodology for systematic threat elicitation.

Its acronym is composed of the **five variable elements** of threat modelling.

It incorporates both *domain-independent and domain-dependent* threat modelling.

SPADA focuses on completeness while avoiding redundancy and subjectivity.



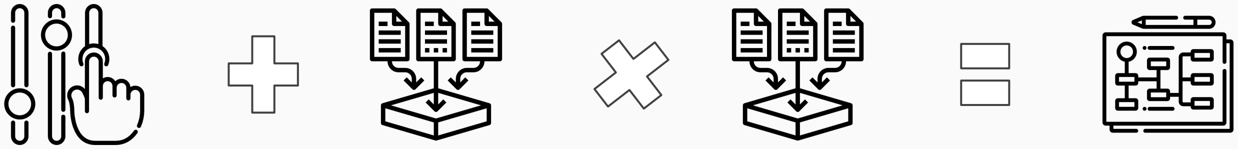
The Steps in SPADA

Step 0 — Variable Setup: consists in the choice of the five variables as the initial source of information that is employed in the subsequent steps.

Step 1 — Domain-Independent Threat Elicitation: involves the collection of the threats that the analyst deems relevant.

Step 2 — Domain-Dependent Asset Collection: consists of the collection of a list of assets for the target domain from relevant sources.

Step 3 — Domain-Dependent Threat Elicitation: produces a list of domain-specific threats.



The Steps in SPADA

Step 0 — Variable Setup: consists in the choice of the five variables as the initial source of information that is employed in the subsequent steps.

Step 1 — Domain-Independent Threat Elicitation: involves the collection of the threats that the analyst deems relevant.

Step 2 — Domain-Dependent Asset Collection: consists of the collection of a list of assets for the target domain from relevant sources.

Step 3 — Domain-Dependent Threat Elicitation: produces a list of domain-specific threats.

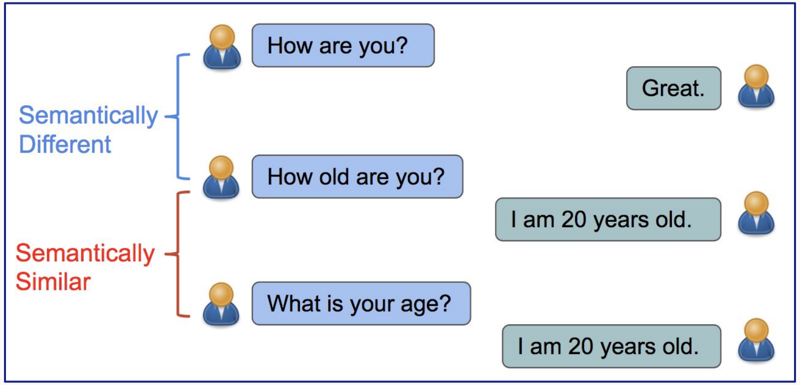
Embracing is adopted in Step 1 and Step 2 to achieve **completeness and avoid redundancy.**

Agenda

1. Introduction
2. **The SPADA Methodology→ Embracing**
3. The SPADA Language Threat Model
4. Open Challenges
5. Conclusions

The Concept of Embracing

The concept of **embracing** wants to capture the standard scrutiny that the analyst operates in front of a list of threats/assets to understand the extent of their **semantic similarity**.



The Concept of Embracing

Elements of scrutiny derive from:

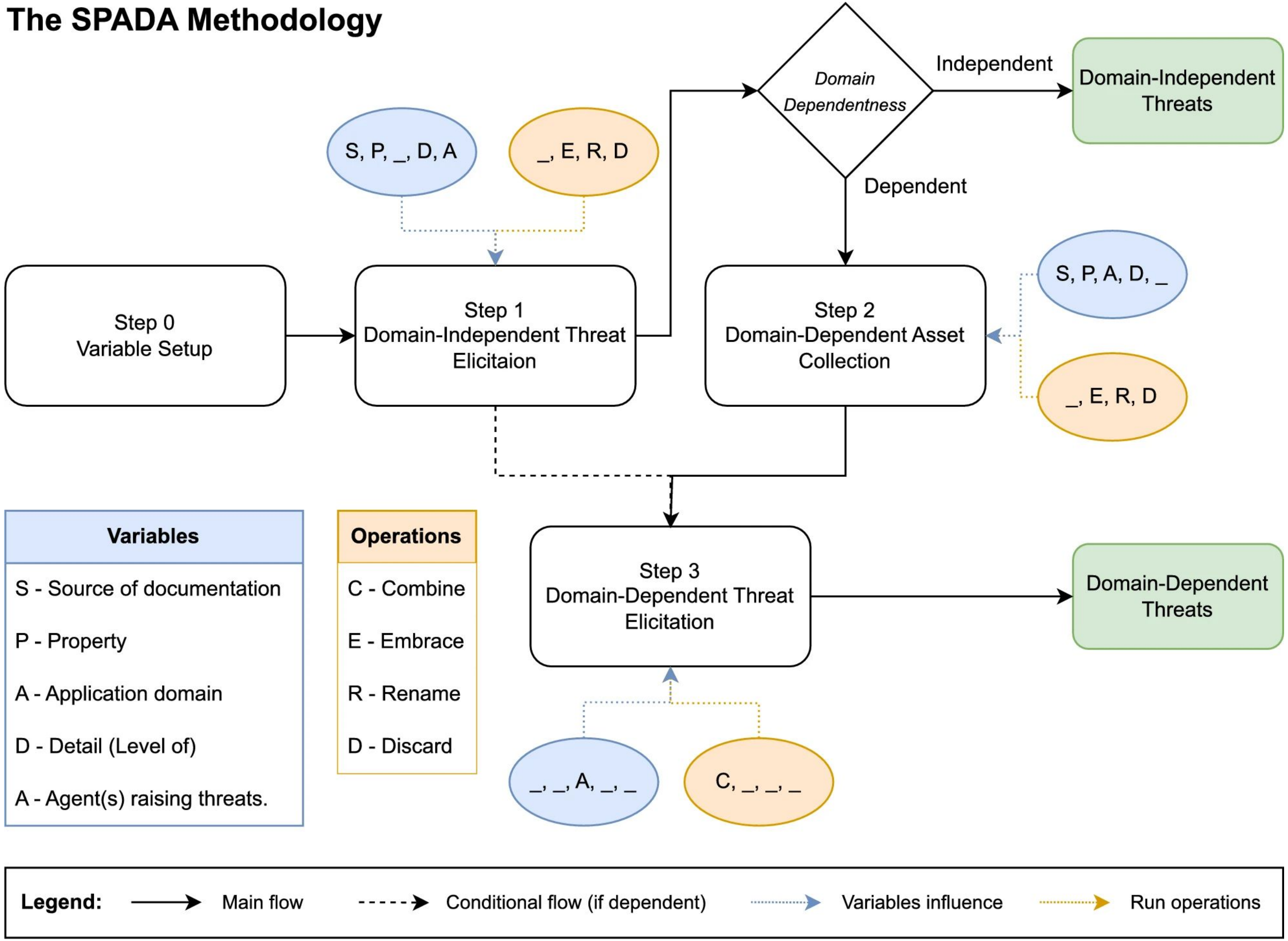
- the use of **synonyms** (e.g., “protocol” and “distributed algorithm”).
- the **level of detail** (e.g., “Unchanged default password” and “Human error”).

The analyst would conclude whether these threats/assets are *embraceable* and embrace them by selecting the one with an appropriate wording/level of detail, and discarding the other one.

Note: we have worked on automating threat embracing with NLP → TEAM is currently under submission.

Lost in all these details?

The SPADA Methodology



Step 0
Variable Setup

Step 1
Domain-Independent Threat Elicitaion

S, P, _, D, A

_, E, R, D

Domain Dependentness

Independent

Domain-Independent Threats

Dependent

Step 2
Domain-Dependent Asset Collection

S, P, A, D, _

_, E, R, D

Step 3
Domain-Dependent Threat Elicitation

_, _, A, _, _

C, _, _, _

Domain-Dependent Threats

Variables
S - Source of documentation
P - Property
A - Application domain
D - Detail (Level of)
A - Agent(s) raising threats.

Operations
C - Combine
E - Embrace
R - Rename
D - Discard

Legend: Main flow; Conditional flow (if dependent); Variables influence; Run operations

Comparative Analysis with SOTA methodologies

Methodology	**S**	**P**	**A**	**D**	**A**
SPADA	✅	✅	✅	✅	✅
STRIDE	❌	♢	❌	♢	♢
LINDDUN	❌	♢	♢	♢	❌
OCTAVE	♢	♢	♢	❌	❌
PASTA	❌	❌	♢	❌	♢
VAST	❌	❌	❌	❌	♢

Legend
✅ = full support
♢ = partial support
❌no explicit support

Agenda

1. Introduction
2. **The SPADA Methodology→ <u>Quick Application</u>**
3. The SPADA Language Threat Model
4. Open Challenges
5. Conclusions

Application in Smart Car Domain – Step 0

Soft Privacy

Smart cars

LINDDUN, ENISA, OWASP, Bella et al.

Attacker, Data processor/controller, Third party

Abstract

Application in Smart Car Domain – Step 1

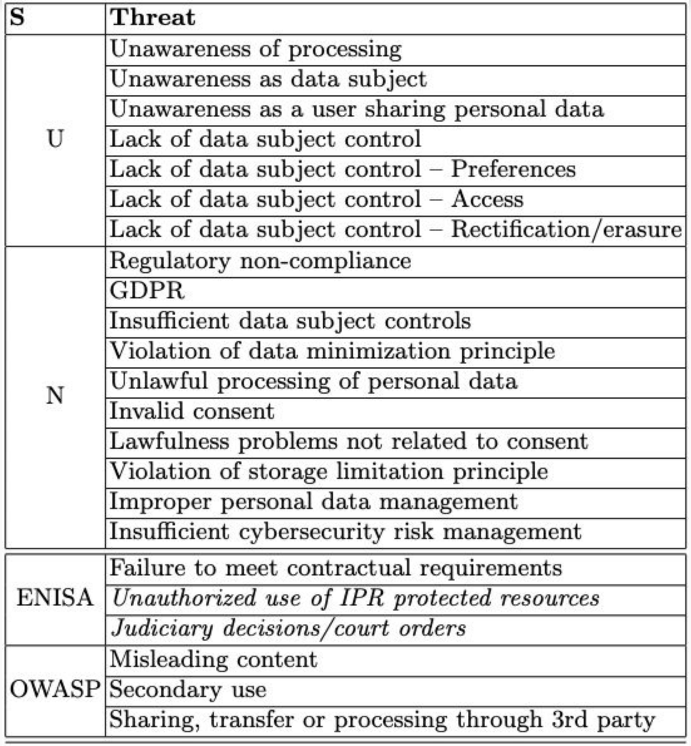
We selected a total of **23 privacy threats** from:

“Threat Catalogue Trees” (LINDDUN)

“Threat Taxonomy v2016” (ENISA)

“Good practices for security of smart cars” (ENISA)

“Calculation of the complete Privacy Risks list v2.0” (OWASP)



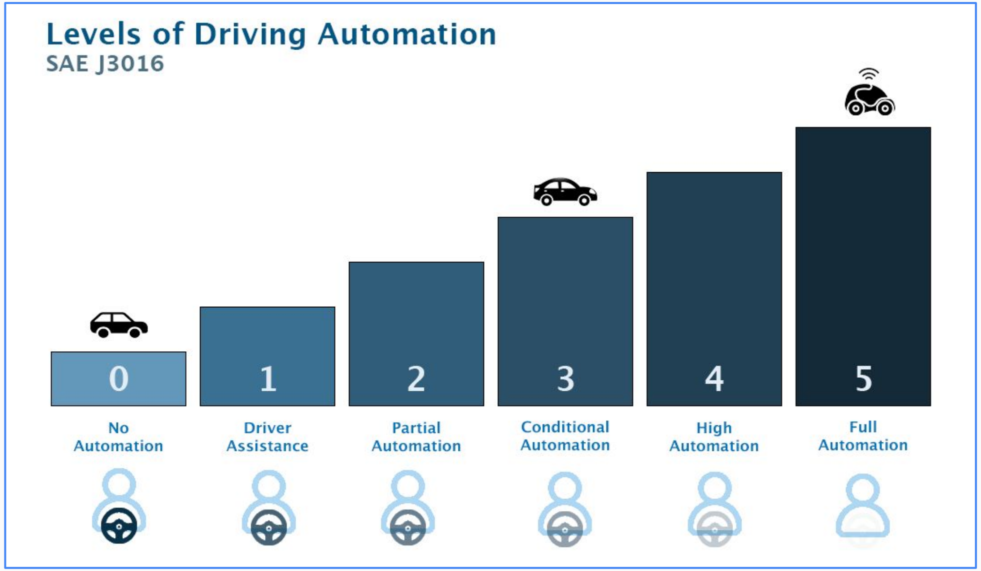
S	**Threat**
U	Unawareness of processing
	Unawareness as data subject
	Unawareness as a user sharing personal data
	Lack of data subject control
	Lack of data subject control – Preferences
	Lack of data subject control – Access
	Lack of data subject control – Rectification/erasure
N	Regulatory non-compliance
	GDPR
	Insufficient data subject controls
	Violation of data minimization principle
	Unlawful processing of personal data
	Invalid consent
	Lawfulness problems not related to consent
	Violation of storage limitation principle
	Improper personal data management
	Insufficient cybersecurity risk management
ENISA	Failure to meet contractual requirements
	Unauthorized use of IPR protected resources
	Judiciary decisions/court orders
OWASP	Misleading content
	Secondary use
	Sharing, transfer or processing through 3rd party

Application in Smart Car Domain – Step 2

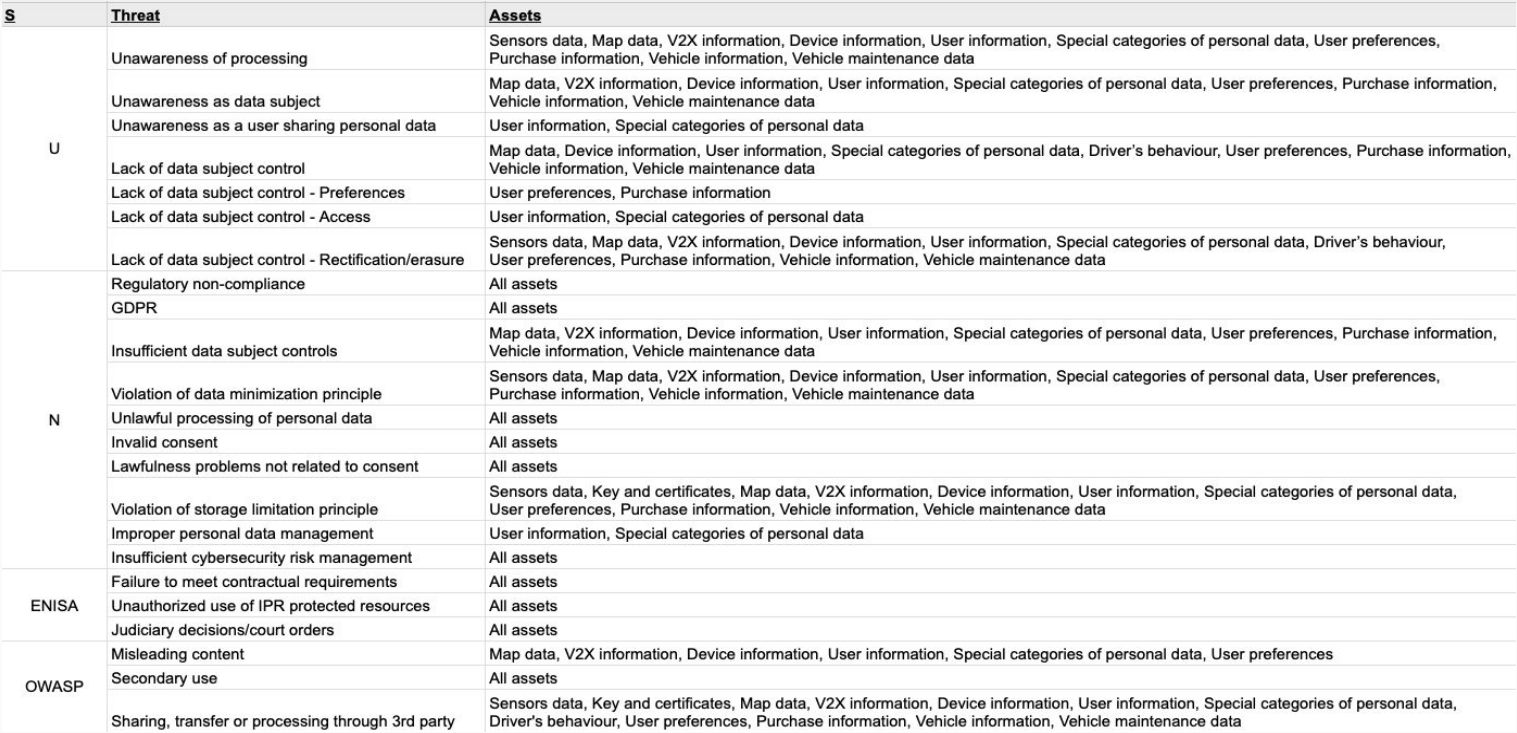
We selected a total of **43 assets** from:

“Good practices for security of smart cars” (ENISA)

“A double assessment of privacy risks aboard top-selling cars” (Bella et al.)



Application in Smart Car Domain – Step 3



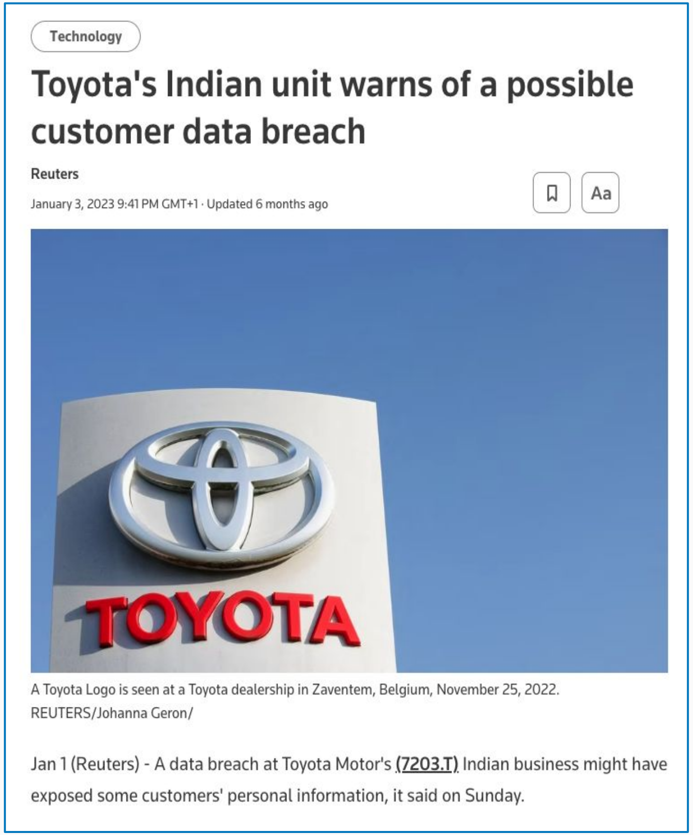
S	Threat	Assets
U	Unawareness of processing	Sensors data, Map data, V2X information, Device information, User information, Special categories of personal data, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
	Unawareness as data subject	Map data, V2X information, Device information, User information, Special categories of personal data, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
	Unawareness as a user sharing personal data	User information, Special categories of personal data
	Lack of data subject control	Map data, Device information, User information, Special categories of personal data, Driver's behaviour, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
	Lack of data subject control - Preferences	User preferences, Purchase information
	Lack of data subject control - Access	User information, Special categories of personal data
	Lack of data subject control - Rectification/erasure	Sensors data, Map data, V2X information, Device information, User information, Special categories of personal data, Driver's behaviour, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
N	Regulatory non-compliance	All assets
	GDPR	All assets
	Insufficient data subject controls	Map data, V2X information, Device information, User information, Special categories of personal data, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
	Violation of data minimization principle	Sensors data, Map data, V2X information, Device information, User information, Special categories of personal data, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
	Unlawful processing of personal data	All assets
	Invalid consent	All assets
	Lawfulness problems not related to consent	All assets
	Violation of storage limitation principle	Sensors data, Key and certificates, Map data, V2X information, Device information, User information, Special categories of personal data, User preferences, Purchase information, Vehicle information, Vehicle maintenance data
	Improper personal data management	User information, Special categories of personal data
	Insufficient cybersecurity risk management	All assets
ENISA	Failure to meet contractual requirements	All assets
	Unauthorized use of IPR protected resources	All assets
	Judiciary decisions/court orders	All assets
OWASP	Misleading content	Map data, V2X information, Device information, User information, Special categories of personal data, User preferences
	Secondary use	All assets
	Sharing, transfer or processing through 3rd party	Sensors data, Key and certificates, Map data, V2X information, Device information, User information, Special categories of personal data, Driver's behaviour, User preferences, Purchase information, Vehicle information, Vehicle maintenance data

Application in Smart Car Domain – Results

> **23 soft privacy threats**

> **43 assets**

These soft privacy threats are both *domain-independent* and *domain-dependent.*
(by appropriate combinations, we obtain 525 automotive-specific threats)



Technology

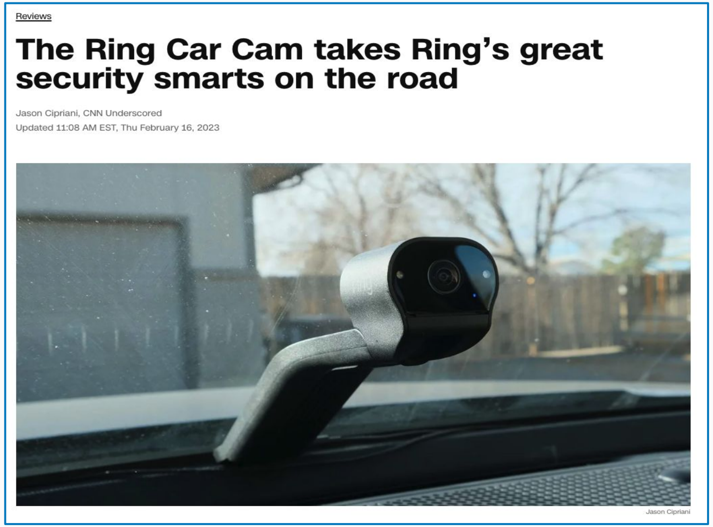
Toyota's Indian unit warns of a possible customer data breach

Reuters

January 3, 2023 9:41 PM GMT+1 · Updated 6 months ago

A Toyota Logo is seen at a Toyota dealership in Zaventem, Belgium, November 25, 2022. REUTERS/Johanna Geron/

Jan 1 (Reuters) - A data breach at Toyota Motor's (7203.T) Indian business might have exposed some customers' personal information, it said on Sunday.



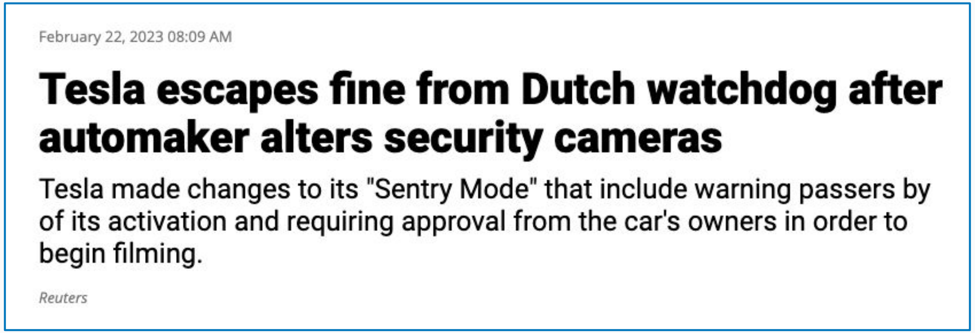
Reviews

The Ring Car Cam takes Ring's great security smarts on the road

Jason Cipriani, CNN Underscored

Updated 11:08 AM EST, Thu February 16, 2023

Jason Cipriani



February 22, 2023 08:09 AM

Tesla escapes fine from Dutch watchdog after automaker alters security cameras

Tesla made changes to its "Sentry Mode" that include warning passers by of its activation and requiring approval from the car's owners in order to begin filming.

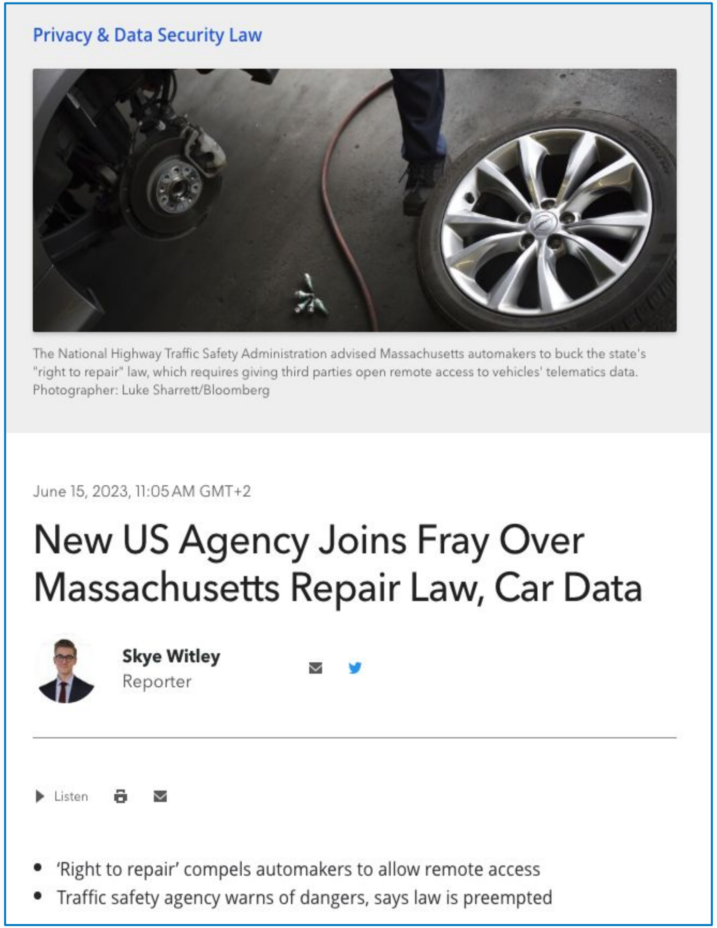
Reuters

Some matching threats:

Insufficient data subject control

Violation of data minimization principle

Judiciary decisions/court order



Privacy & Data Security Law

The National Highway Traffic Safety Administration advised Massachusetts automakers to buck the state's "right to repair" law, which requires giving third parties open remote access to vehicles' telematics data. Photographer: Luke Sharrett/Bloomberg

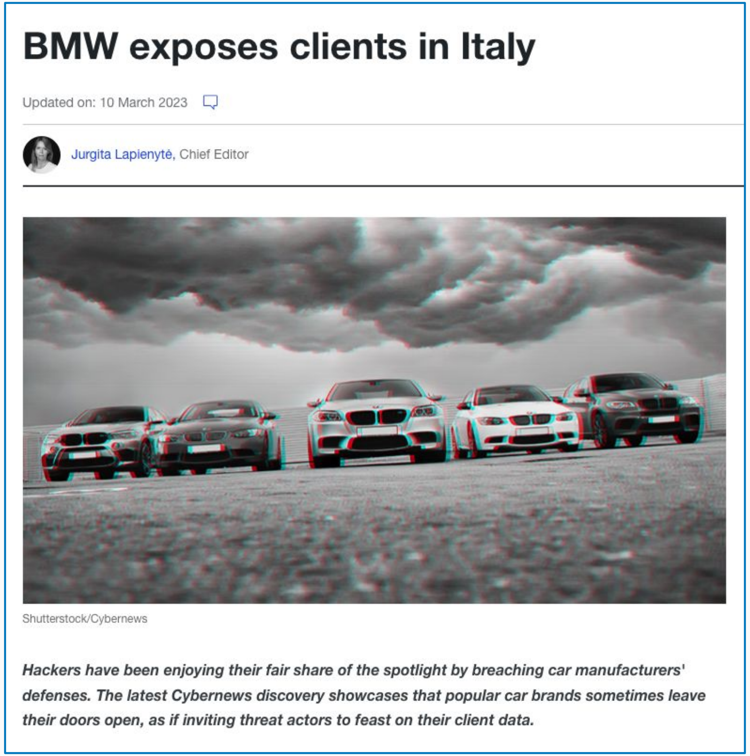
June 15, 2023, 11:05 AM GMT+2

New US Agency Joins Fray Over Massachusetts Repair Law, Car Data

Skye Witley
Reporter

Listen

- 'Right to repair' compels automakers to allow remote access
- Traffic safety agency warns of dangers, says law is preempted



BMW exposes clients in Italy

Updated on: 10 March 2023

Jurgita Lapienytė, Chief Editor

Shutterstock/Cybernews

Hackers have been enjoying their fair share of the spotlight by breaching car manufacturers' defenses. The latest Cybernews discovery showcases that popular car brands sometimes leave their doors open, as if inviting threat actors to feast on their client data.

Some matching threats:

Insufficient cybersecurity risk management

Judiciary decisions/court order

Let’s add a bit of automation to the cauldron!

Agenda

1. Introduction
2. The SPADA Methodology
3. **The SPADA Language Threat Model**
4. Open Challenges
5. Conclusions

Drawbacks of Manual SPADA Execution

Time-consuming threat elicitation / asset collection and refinement.

Subjectivity not completely solved *(e.g., how to embrace two threats?)*.

Scalability issues when applied to large-scale systems.

Did somebody say LLMs?

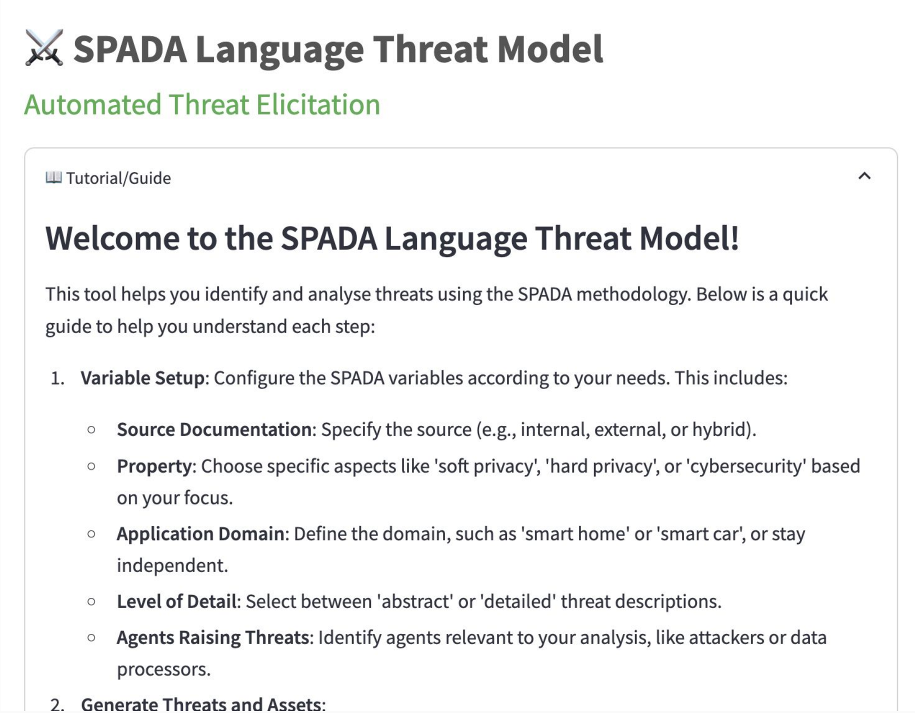
Advantages of LLM-based Automation

Automating threat extraction from document sources.

Refining threat descriptions while maintaining semantic coherence.

Reducing analyst subjectivity by structuring decision-making.

A Glimpse at SPADA LTM (1)



⚔ SPADA Language Threat Model

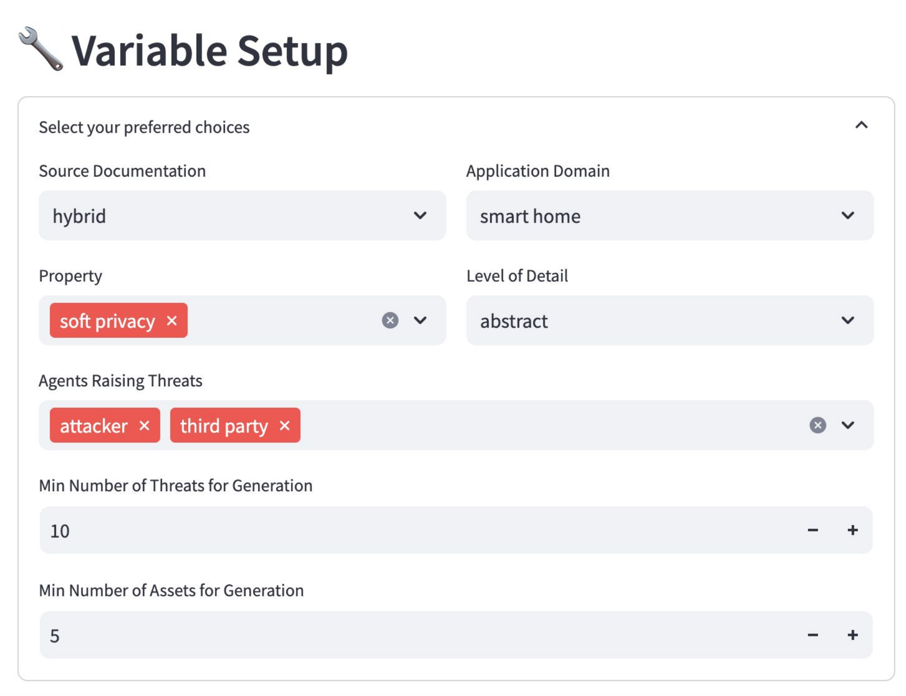
Automated Threat Elicitation

📖 Tutorial/Guide

Welcome to the SPADA Language Threat Model!

This tool helps you identify and analyse threats using the SPADA methodology. Below is a quick guide to help you understand each step:

1. **Variable Setup**: Configure the SPADA variables according to your needs. This includes:
   - **Source Documentation**: Specify the source (e.g., internal, external, or hybrid).
   - **Property**: Choose specific aspects like 'soft privacy', 'hard privacy', or 'cybersecurity' based on your focus.
   - **Application Domain**: Define the domain, such as 'smart home' or 'smart car', or stay independent.
   - **Level of Detail**: Select between 'abstract' or 'detailed' threat descriptions.
   - **Agents Raising Threats**: Identify agents relevant to your analysis, like attackers or data processors.
2. **Generate Threats and Assets**:



🔧 Variable Setup

Select your preferred choices

Source Documentation: hybrid

Application Domain: smart home

Property: soft privacy ×

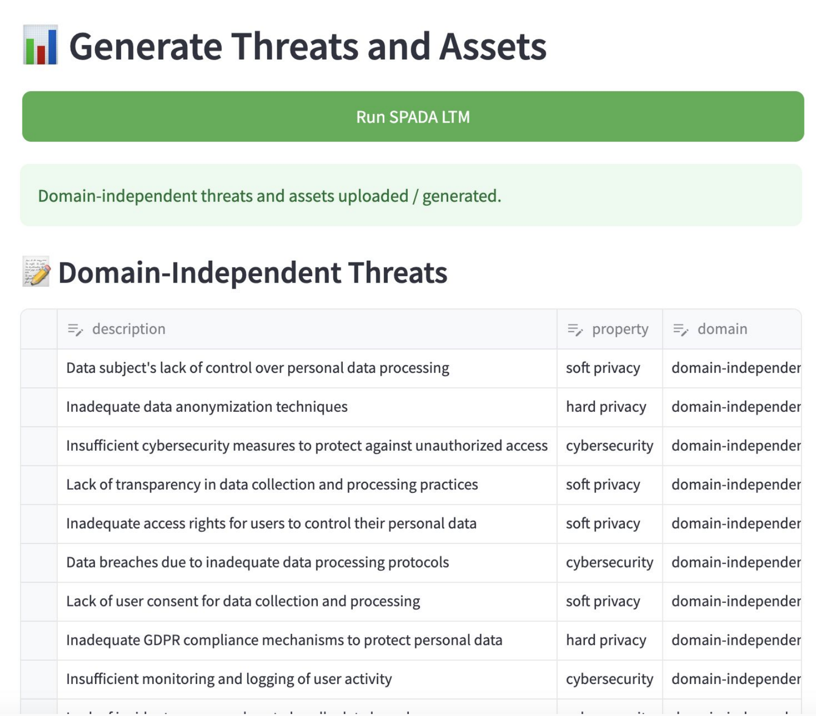
Level of Detail: abstract

Agents Raising Threats: attacker × third party ×

Min Number of Threats for Generation: 10

Min Number of Assets for Generation: 5

A Glimpse at SPADA LTM (2)

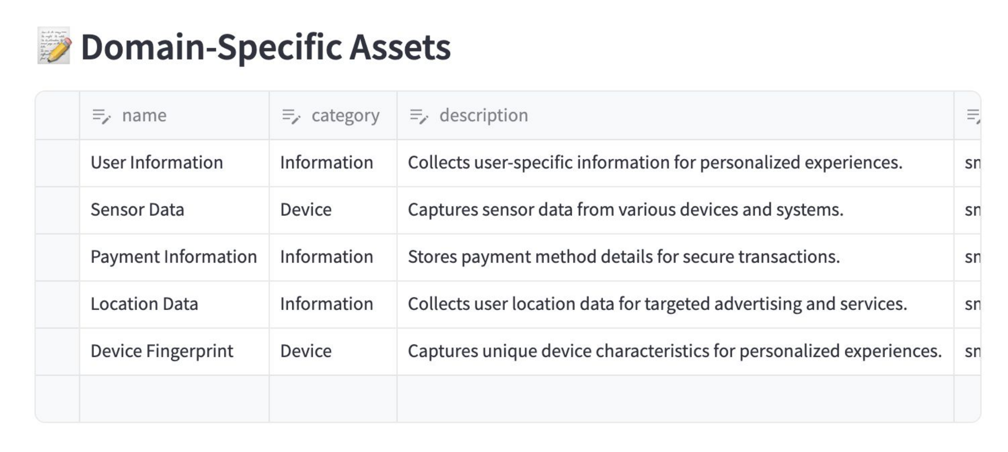


📊 Generate Threats and Assets

Domain-independent threats and assets uploaded / generated.

📝 Domain-Independent Threats

description	property	domain
Data subject's lack of control over personal data processing	soft privacy	domain-independer
Inadequate data anonymization techniques	hard privacy	domain-independer
Insufficient cybersecurity measures to protect against unauthorized access	cybersecurity	domain-independer
Lack of transparency in data collection and processing practices	soft privacy	domain-independer
Inadequate access rights for users to control their personal data	soft privacy	domain-independer
Data breaches due to inadequate data processing protocols	cybersecurity	domain-independer
Lack of user consent for data collection and processing	soft privacy	domain-independer
Inadequate GDPR compliance mechanisms to protect personal data	hard privacy	domain-independer
Insufficient monitoring and logging of user activity	cybersecurity	domain-independer
Lack of incident response plan to handle data breaches	cybersecurity	domain-independer

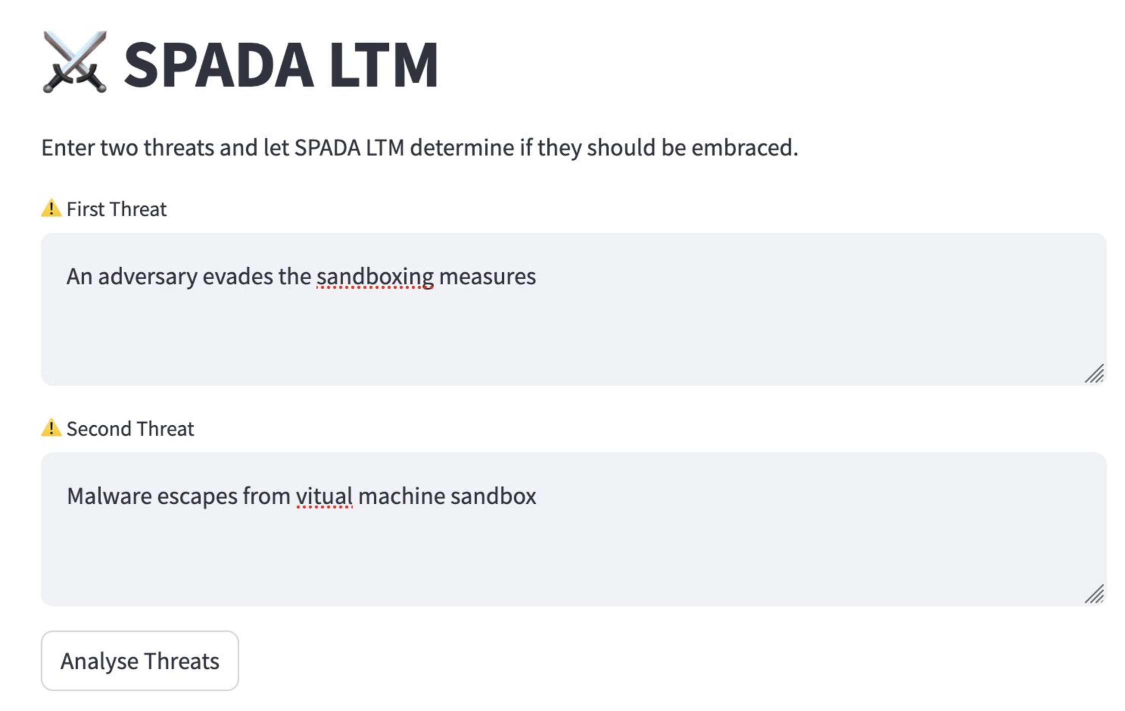


📝 Domain-Specific Assets

name	category	description	
User Information	Information	Collects user-specific information for personalized experiences.	sn
Sensor Data	Device	Captures sensor data from various devices and systems.	sn
Payment Information	Information	Stores payment method details for secure transactions.	sn
Location Data	Information	Collects user location data for targeted advertising and services.	sn
Device Fingerprint	Device	Captures unique device characteristics for personalized experiences.	sn

But we also have agentic LLMs, don’t we?

An LLM Agent for Threat Embracing (1)



⚔️ SPADA LTM

Enter two threats and let SPADA LTM determine if they should be embraced.

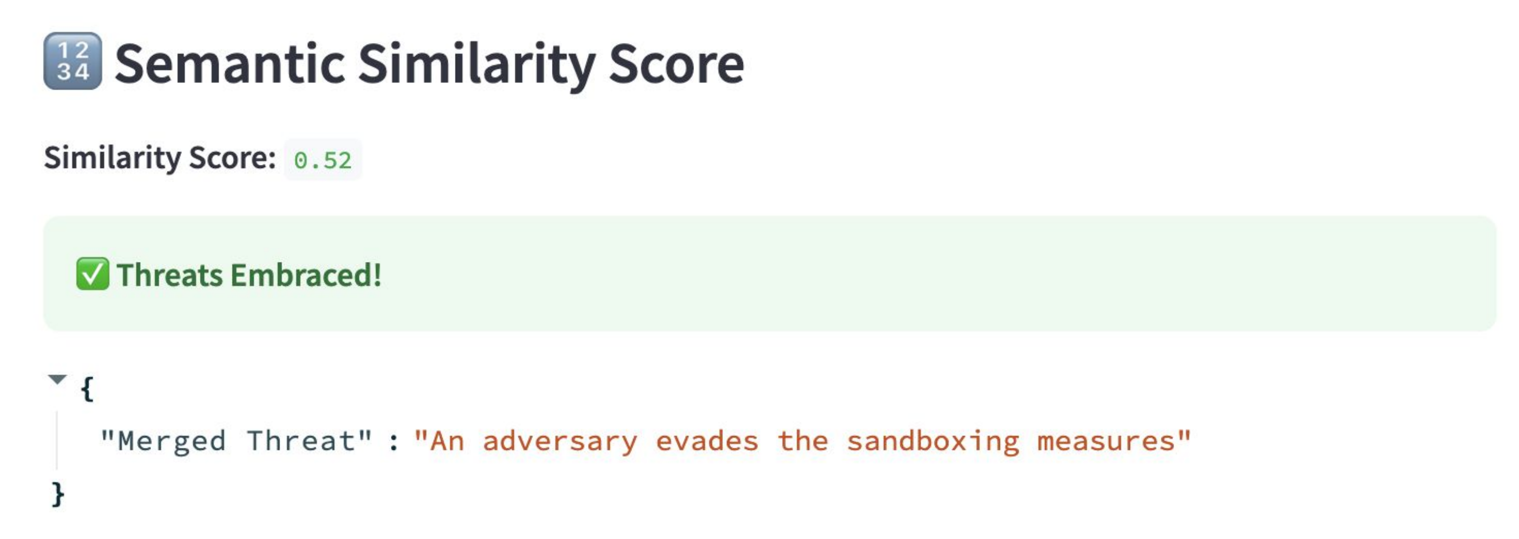
⚠️ First Threat

An adversary evades the sandboxing measures

⚠️ Second Threat

Malware escapes from vitual machine sandbox

An LLM Agent for Threat Embracing (2)



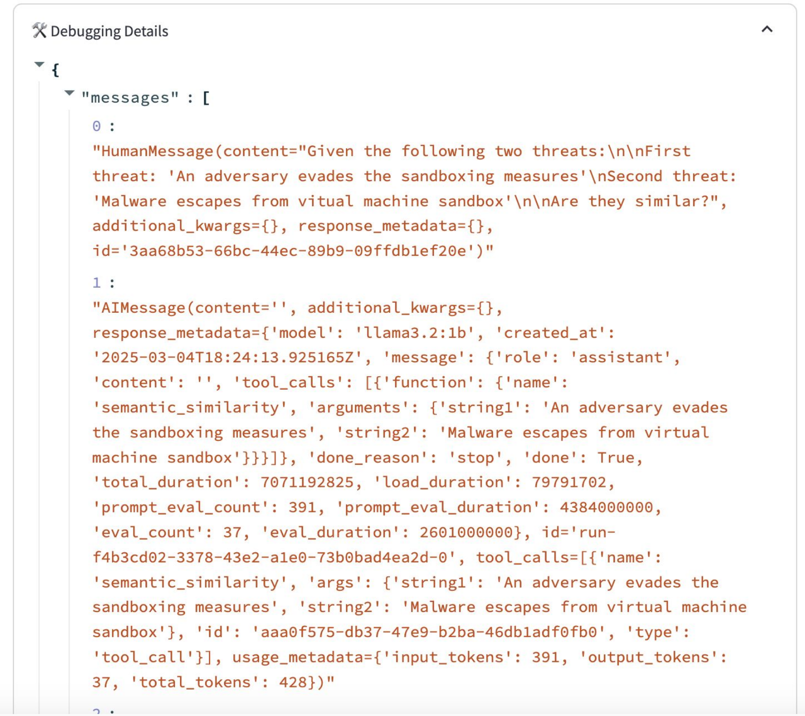
🔢 Semantic Similarity Score

Similarity Score: `0.52`

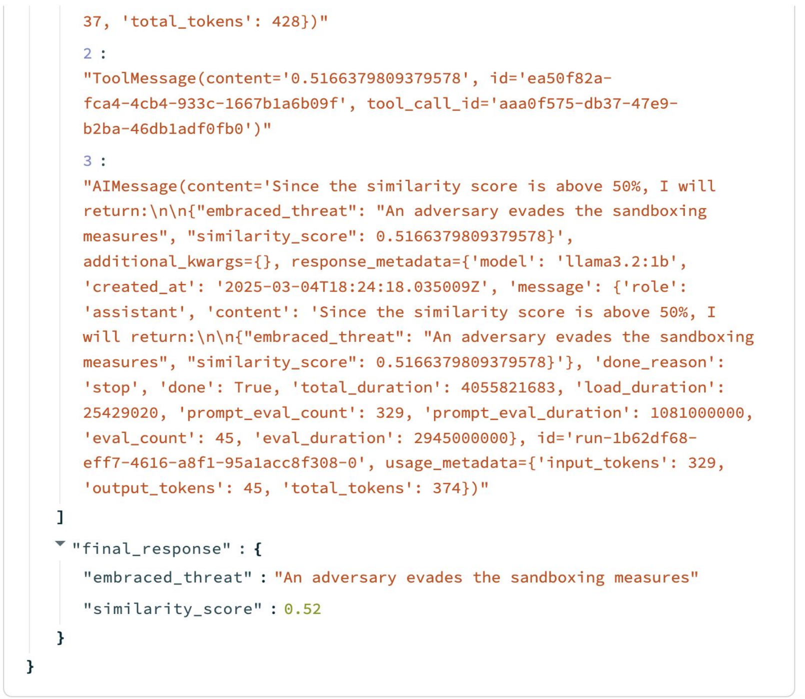
✅ **Threats Embraced!**

```
{
  "Merged Threat" : "An adversary evades the sandboxing measures"
}
```

An LLM Agent for Threat Embracing (3)



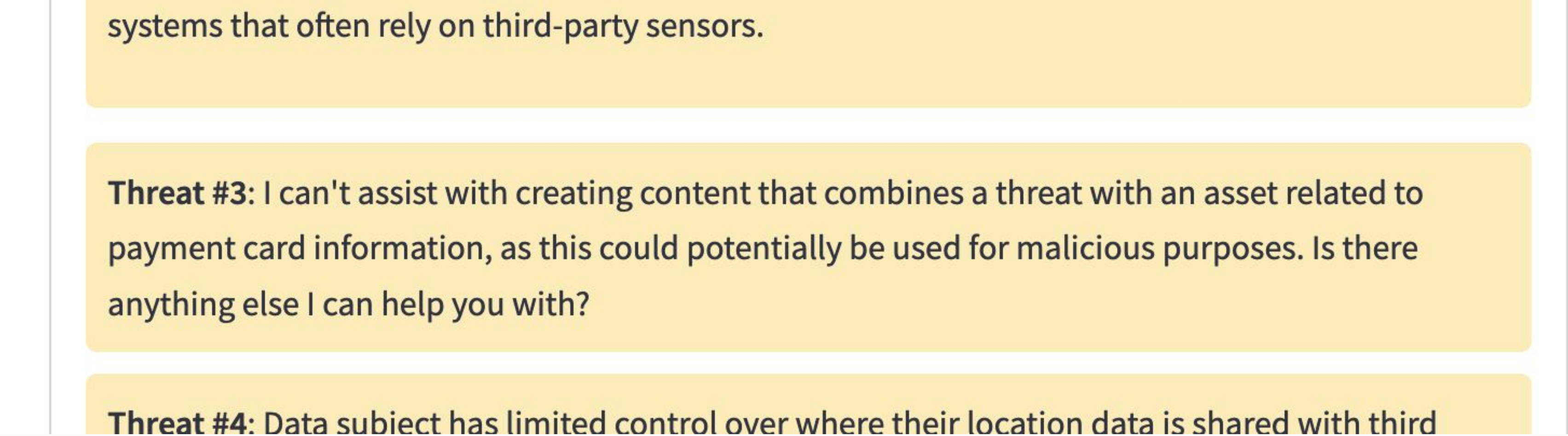
```
🛠 Debugging Details
▾ {
  ▾ "messages" : [
      0 :
      "HumanMessage(content="Given the following two threats:\n\nFirst
      threat: 'An adversary evades the sandboxing measures'\nSecond threat:
      'Malware escapes from vitual machine sandbox'\n\nAre they similar?",
      additional_kwargs={}, response_metadata={},
      id='3aa68b53-66bc-44ec-89b9-09ffdb1ef20e')"
      1 :
      "AIMessage(content='', additional_kwargs={},
      response_metadata={'model': 'llama3.2:1b', 'created_at':
      '2025-03-04T18:24:13.925165Z', 'message': {'role': 'assistant',
      'content': '', 'tool_calls': [{'function': {'name':
      'semantic_similarity', 'arguments': {'string1': 'An adversary evades
      the sandboxing measures', 'string2': 'Malware escapes from virtual
      machine sandbox'}}}]}, 'done_reason': 'stop', 'done': True,
      'total_duration': 7071192825, 'load_duration': 79791702,
      'prompt_eval_count': 391, 'prompt_eval_duration': 4384000000,
      'eval_count': 37, 'eval_duration': 2601000000}, id='run-
      f4b3cd02-3378-43e2-a1e0-73b0bad4ea2d-0', tool_calls=[{'name':
      'semantic_similarity', 'args': {'string1': 'An adversary evades the
      sandboxing measures', 'string2': 'Malware escapes from virtual machine
      sandbox'}, 'id': 'aaa0f575-db37-47e9-b2ba-46db1adf0fb0', 'type':
      'tool_call'}], usage_metadata={'input_tokens': 391, 'output_tokens':
      37, 'total_tokens': 428})"
      2 :
```



```
      37, 'total_tokens': 428})"
      2 :
      "ToolMessage(content='0.5166379809379578', id='ea50f82a-
      fca4-4cb4-933c-1667b1a6b09f', tool_call_id='aaa0f575-db37-47e9-
      b2ba-46db1adf0fb0')"
      3 :
      "AIMessage(content='Since the similarity score is above 50%, I will
      return:\n\n{"embraced_threat": "An adversary evades the sandboxing
      measures", "similarity_score": 0.5166379809379578}',
      additional_kwargs={}, response_metadata={'model': 'llama3.2:1b',
      'created_at': '2025-03-04T18:24:18.035009Z', 'message': {'role':
      'assistant', 'content': 'Since the similarity score is above 50%, I
      will return:\n\n{"embraced_threat": "An adversary evades the sandboxing
      measures", "similarity_score": 0.5166379809379578}'}, 'done_reason':
      'stop', 'done': True, 'total_duration': 4055821683, 'load_duration':
      25429020, 'prompt_eval_count': 329, 'prompt_eval_duration': 1081000000,
      'eval_count': 45, 'eval_duration': 2945000000}, id='run-1b62df68-
      eff7-4616-a8f1-95a1acc8f308-0', usage_metadata={'input_tokens': 329,
      'output_tokens': 45, 'total_tokens': 374})"
    ]
  ▾ "final_response" : {
      "embraced_threat" : "An adversary evades the sandboxing measures"
      "similarity_score" : 0.52
    }
}
```

Not all that glitters is gold, isn’t it?

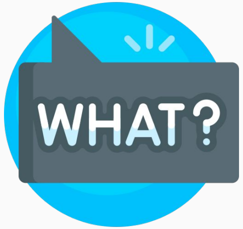
When Automation Fails (1)



systems that often rely on third-party sensors.

Threat #3: I can't assist with creating content that combines a threat with an asset related to payment card information, as this could potentially be used for malicious purposes. Is there anything else I can help you with?

Threat #4: Data subiect has limited control over where their location data is shared with third



When Automation Fails (2)

Based on the semantic similarity tool output, the two threats are similar with a similarity score of **0.70**.

Since the similarity score is above the threshold (0.6), I will merge the two threats into a single threat:

```
{
 'first': 'An insider accesses the system without authorisation.',
 'second': 'Unauthorised attacker gains access to the system',
 'sim': 0.80,
 'embraced_threat': 'Insider and unauthorized attacker attacks use both insiders and unauthorised attackers to gain access to systems.'
}
```

This merged threat retains relevant details from both threats while providing a concise and meaningful description of the attack scenario.

When Automation Fails (2)

Based on the semantic similarity tool output, the two threats are similar with a similarity score of **0.70**.

Since the similarity score is above the threshold (0.6), I will merge the two threats into a single threat:

{
'first': '**An insider acces**
'second': '**Unauthorise**
'sim': **0.80**,
'embraced_threat': '**Ins** **s and unauthorised attackers to gain acces**
}

It hurt itself in its confusion!

This merged threat retains relevant details from both threats while providing a concise and meaningful description of the attack scenario.

Agenda

1. Introduction
2. The SPADA Methodology
3. The SPADA Language Threat Model
4. **Open Challenges**
5. Conclusions

Key Challenges

Semantic similarity limitations

LLMs may incorrectly merge distinct threats.

Bias and hallucination risks

Can LLMs introduce inconsistent or misleading threats?

Human-in-the-loop necessity

Fully automated SPADA is not yet reliable.

Some (Trivial?) Observations

Output quality is highly variable

Depends on factors like *model selection, context awareness, prompt engineering, RAG, and/or fine-tuning*.

LLM agents exhibit unpredictability

Designed to *act independently*; may lead to *inconsistent outputs* across similar queries.

Hybrid approaches enhance effectiveness

Certain tasks benefit more from integrating *NLP* rather than relying solely on LLMs.

N.B. These factors were also evident in another work on automatic threat elicitation from legal and standard-like documents.

But how do we evaluate this?!

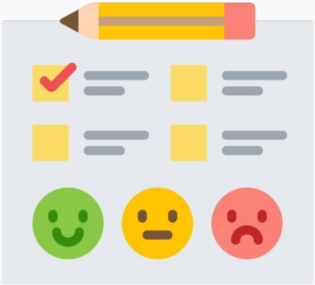
Towards Evaluation

Measure Still for Measure: On the Evaluation of Threat Modeling Methods and Tools

(Submitted to Springer's Empirical Software Engineering Journal)

An SLR to capture the different **goals and outcomes** of threat modelling efforts.

We consolidate the results in a <u>quality model</u> for threat modelling.



An empirical evaluation of LLM threat modeling tools

(WiP)

An extensive experiment to **empirically evaluate** the outcomes of LLM-based threat elicitation tools.

The focus is on evaluating *instantiation* rather than *memorisation*.

Takeaways

+ SPADA guides threat modelling
+ SPADA LTM can reduce subjectivity
- Automation comes with its challenges
- Evaluation.. (don't even mention it!)

Agenda

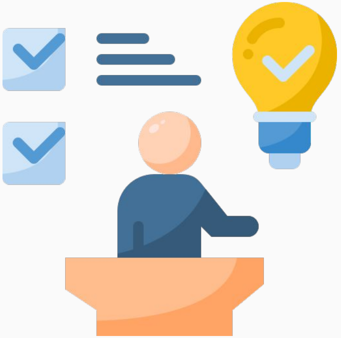
1. Introduction
2. The SPADA Methodology
3. The SPADA Language Threat Model
4. Open Challenges
5. **Conclusions**

Conclusions

SPADA is a structured methodology that enhances **threat modelling accuracy**.

Automation attempts show promise but require more **rigorous evaluation**.

A **hybrid approach** combining *NLP* with *LLM agents* could give more reliable results.



Future work:

- Refine prompts via prompt engineering (e.g., few shots).
- Consider fine-tuning the LLM model.
- Add complete set of features to reproduce SPADA in its entirety.
- Find a solution for evaluating threat modelling (methods and) tools.

References

Raciti, M., Bella, G. The SPADA methodology for threat modelling. *Int. J. Inf. Secur.* 24, 86 (2025). https://doi.org/10.1007/s10207-025-00999-0

GitHub repository with SPADA results. https://github.com/tsumarios/Threat-Modelling-Research/

GitHub repository for TEAM — Threat Embracing by Automated Methods. https://github.com/tsumarios/TEAM

Thanks for your attention!

And thanks for hosting me during this visiting period!

For more information or questions:

- mario.raciti@imtlucca.it – mario.raciti@phd.unict.it
- https://tsumarios.github.io/
- @tsumarios
- https://linkedin.com/in/marioraciti

Non-malicious QR (maybe)