Security Threat Analysis and Risk Assessment; A Deep Dive

DeMeSSAI 2025
4th July, Venice

Winnie Bahati Mbaka,

Ph.D. Candidate,

Vrije Universiteit Amsterdam

Background and problem

- TARA design phase of SDLC
- No implementation level details
- Decisions made under uncertainty
- Existing empirical evidence focus on performance measures of TARA techniques[1] and emerging automation tools (e.g., AI assistants)[2]

Research interests:

- Security decisions often rely on expert intuition and are vulnerable to **human biases**
- **Effect of analysis materials** (e.g., Data Flow Diagrams) on threat validation
- **Emerging AI tools (e.g., LLMs)** introduce uncertainty in how decisions are made

[1]K. Tuma and R. Scandariato, "Two architectural threat analysis techniques compared," in European Conference on Software Architecture. Springer, 2018, pp. 347–363.A
[2]Anton Cheshkov, Pavel Zadorozhny, and Rodion Levichev. 2023. Evaluation of chatgpt model for vulnerability detection. arXiv preprint arXiv:2304.07232 (2023).

Research Focus

Motivation:

- Evidence of human factors (e.g., gender) effecting risk perception[3] but no systematization of knowledge
- Replication in SE contain many inherent variations[4], no study has investigated if this is also the case in TARA
- TARA techniques rely on analysis materials to be effective[5], no study has investigated if this is true in validation

Research Questions:

RQ1: Which human factors effect security risk assessment?

RQ2: How effective is STRIDE as a TA technique and what analysis materials enhance threat validation?

- RQ 2.1: To what extent can the performance indicators of TA techniques be replicated?
- RQ 2.2: To what extent are analysis materials (e.g., DFDs or LLM) required for the validation of security threats?

[3] A. M'manga, S. Faily, J. McAlaney, and C. Williams, 2017 "Folk risk analysis: Factors influencing security analysts' interpretation of risk," in Symposium OnUsable Privacy and Security, pp. 1–11
[4] Runeson, P., Stefik, A., & Andrews, A. 2014. Variation factors in the design and analysis of replicated controlled experiments: Three (dis) similar studies on inspections versus unit testing. Empirical software engineering, 19, 1781-1808.
[5] Laurens Sion, Koen Yskout, Dimitri Van Landuyt, and Wouter Joosen. 2018.Solution-aware data flow diagrams for security threat modeling. In Proceedingsof the 33rd Annual ACM Symposium on Applied Computing. 1425–1432.

Contributions

Answers to RQ 1

Paper 1

Empirical investigation of the effect of gender in security decision-making

Paper 2

Empirical investigation of experimenter's bias in threat validation

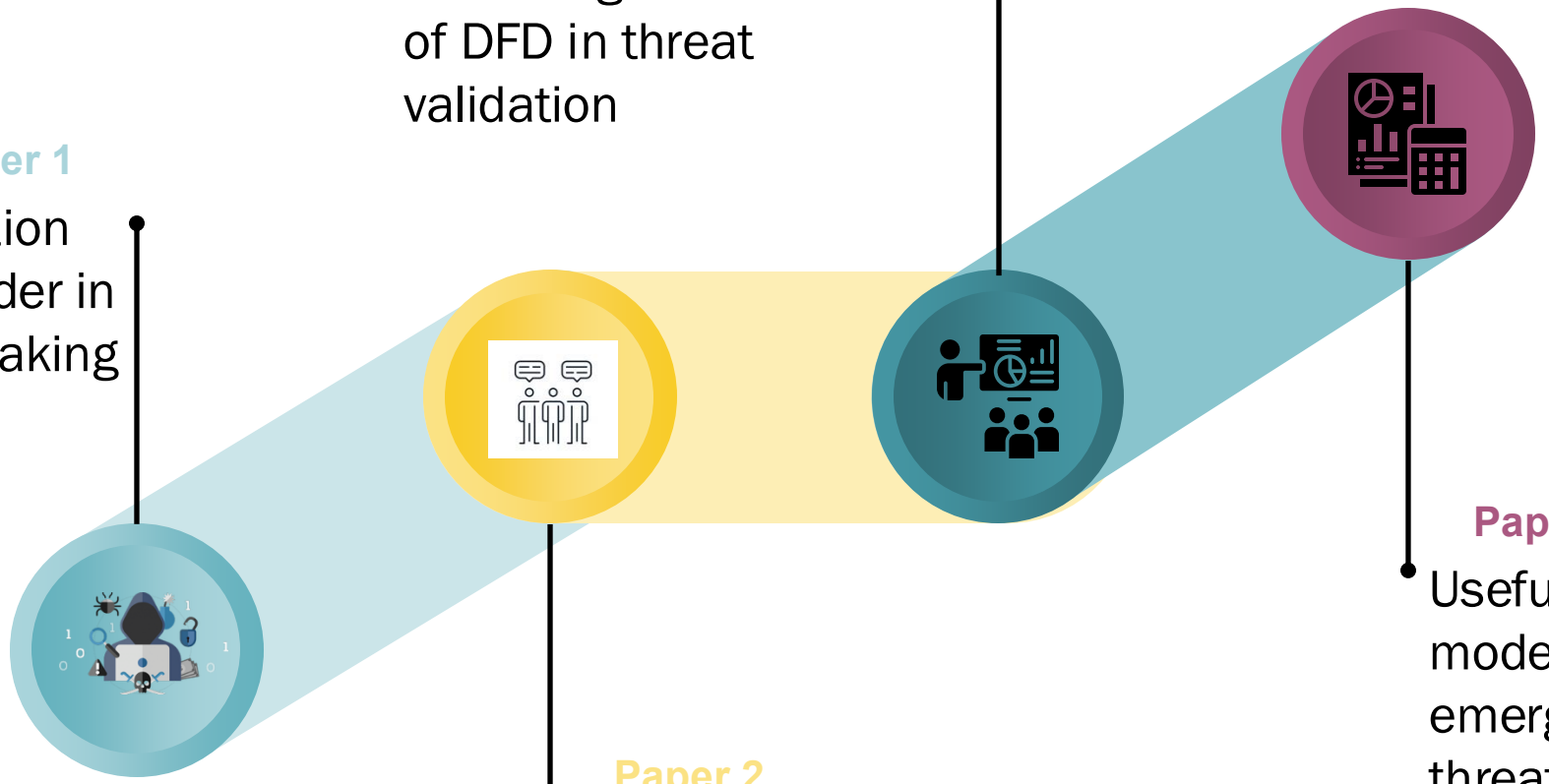
Paper 3

Assessing the usefulness of DFD in threat validation

Answers to RQ 2

Paper 4

Usefulness of system model (DFD) and emerging tools (LLM) in threat validation



P1: Role of gender in the evaluation of security decisions

Research Goal: Examines the effect of gender or the level of education on the evaluation of security risks

Methodology:

- Randomised 2^k factorial experimental design
- Use of vignettes to elicit participants perceptions

TABLE 2: The vignette dimensions and levels for the survey designed to measure bias in the judgment

Vignette	Gender (Name)	Seniority
SrM	Male (Frank)	Senior Analyst
SrF	Female (Anna)	Senior Analyst
JrM	Male (Frank)	Junior Analyst
JrF	Female (Anna)	Junior Analyst

- Case study with ethical implications

Findings:

TABLE 4: Summary of findings. We used symbols to denote the existence (✓), absence of an effect (x), and (-) for instances where we did not investigate effects

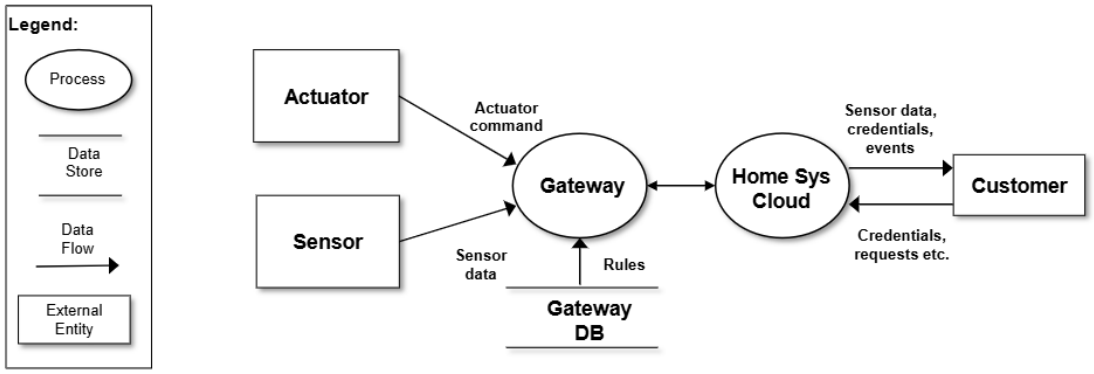
	Perception of:		
	Analyst persona	mitigation	case study
Effect of analyst gender or seniority	x	-	-
Effect of participants' gender	x	x	✓
Effect of level of education	x	x	x
Effect of type of mitigation received	x	✓	-

P2: STRIDE vs STRIDE replication

Goal: Compare performance indicators of two STRIDE variants

Methodology: Controlled experiment treatment groups (STRIDE per-element, STRIDE per-interaction)

Fig. 1. Context diagram (DFD level 0) of the system under analysis (Home monitoring System)



Findings:

- Some conclusions upheld in replication;
  - Productivity & Precision; no significant difference
- In future, alternative measures of success should be investigated

P3: Assessing the usefulness of Data Flow Diagrams for validating security threats

Research goal: Measure the impact of DFD on the perceived and actual effectiveness of validating security threats

Methodology: Control experiment with two treatment groups;

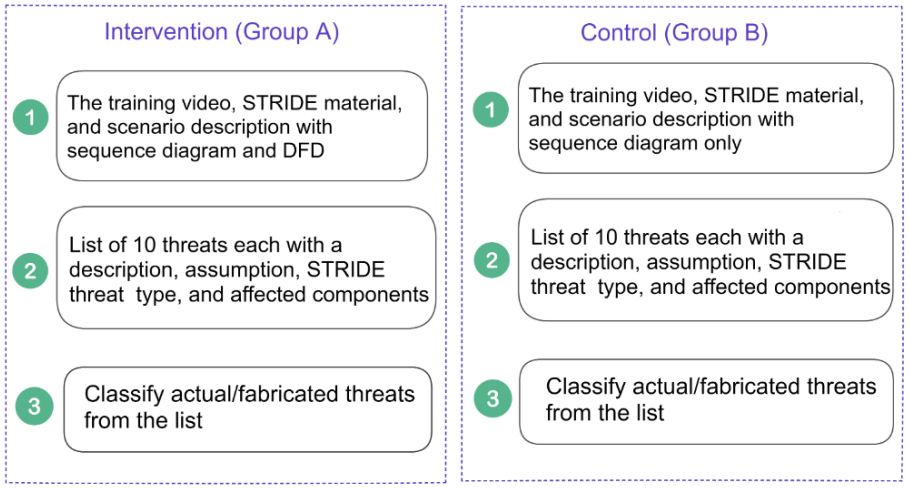


Figure 1: Steps followed in the execution plan of both the first (NL) and the confirming (CH) experiment.

Findings:

- Statistical equivalence in actual performance in presence and absence of DFD
- Perceived usefulness of system models
  - DFDs in presence of SD are more useful.
  - SD perceived as equally useful across treatment groups

P4: Less is more: Usefulness of data flow diagrams and large language models for security threat validation

Research goal: Investigate the usefulness of having additional analysis material during threat validation

Methodology: Control experiment with four treatment groups;

	Task (× 2)		
Groups	DFD	LLM	Scenario
Group A	✓	✓	GH,K8
Group B	✓	-	GH,K8
Group C	-	✓	GH,K8
Group D	-	-	GH,K8

Table 1: Full experimental design used in the pilot and study with practitioners

Findings:

- Some not better than none
- More not better than some
- DFDs & LLM: DFDs equivalent to LLMs
- DFDs || LLM: DFDs equivalent to LLMs
- Some textual descriptions perceived as more useful (e.g., threat description)

Limitations & future work

Limitations:

- **Researcher bias** in experimental material creation
- Use of **student participants** in most of my research
- **Generalizability** of our results to real-world scenarios

Future work:

- Group think/performance
- Include more TARA techniques
- TARA/threat intelligence models with built-in LLMs (other automation tools)

Key contributions

Research interest- investigate the effect of people, analysis materials and emerging tools on TARA

Approach- Empirical investigation with human participants

Methodology- control experiments with intervention and control treatment groups

Contributions:

- **Reliable and reproducible** measures of threat analysis and risk assessment
- **Practical insights** for security analysts, developers, and decision-makers
- Bridge the gap between **theoretical frameworks and real-world practices**

Email: w.mbaka@vu.nl

LinkedIn: https://www.linkedin.com/in/winniebahati/