"I'm Getting Information that I Can Act on Now": Exploring the Level of Actionable Information in Tool-generated Threat Reports

ALVI JAWAD, HALA ASSAL, and JASON JASKOLKA, Carleton University, Canada

Existing threat modeling tools have been investigated primarily for their functionality and features but not for the contents that they automatically generate, i.e., threat reports. This paper presents the first study focusing on threat reports; we explore what users consider as "actionable information" in such reports, and assess how well threat reports support users in taking action to address identified threats. Our qualitative study involved semi-structured interviews with 15 participants (primarily software engineers and security experts) from North America. The study involved tasks related to identifying and assessing pieces of actionable information in threat reports generated for a healthcare system case study by two prominent threat modeling tools, namely Microsoft Threat Modeling Tool and OWASP Threat Dragon. Our findings cover five overarching themes determined based on existing literature related to better threat modeling tool outputs (i.e., threat reports); quality of suggestions, clarity of presentation, threat prioritization capability, security decision support, and adequate coverage. Based on our analysis, we found that users consider information detailing threats and mitigation suggestions to be directly actionable, and they consider a threat prioritization scheme and statistical overview of insights as supplementary actionable information. We also assess the level of actionable information present in existing threat reports and outline why the current reports lack adequate coverage of actionable information necessary to make decisions with high confidence. To address the identified shortcomings and satisfy user needs, we provide recommendations for improving the state of threat reports in existing and emerging threat modeling tools.

CCS Concepts: • **Security and privacy → Human and societal aspects of security and privacy**;

ACM Reference Format:
Alvi Jawad, Hala Assal, and Jason Jaskolka. 2024. "I'm Getting Information that I Can Act on Now": Exploring the Level of Actionable Information in Tool-generated Threat Reports. In *The 2024 European Symposium on Usable Security (EuroUSEC 2024), September 30-October 1, 2024, Karlstad, Sweden.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3688459.3688467

1 INTRODUCTION

Threat modeling is a structured process of identifying and understanding common threats against a software system to assess and prioritize mitigations for such threats [45]. Uses for threat modeling include (1) developing a better understanding of system vulnerabilities to elicit a consistent set of security requirements, (2) finding design issues early in the development process to make changes when they are inexpensive, and (3) prioritizing targeted features, allowing the engineering of more secure products that better meet business, customer, and compliance requirements [36]. Organizations can make use of threat modeling to improve their system security. Real-world uses of threat modeling include use in Microsoft's Security Development Lifecycle (SDL) [22], Google's cloud security [16], OWASP's community-driven open-source projects [14], and initiatives from government organizations like the US National Institute of

Authors' Contact Information: Alvi Jawad, alvi.jawad@carleton.ca; Hala Assal, assal@sce.carleton.ca; Jason Jaskolka, jason.jaskolka@carleton.ca, Carleton University, Ottawa, ON, Canada.

Standards and Technology (NIST) in their special publications [20] on various topics. Manually performing the threat modeling process for large and complex systems can be cumbersome and lead to an incomplete threat model [36]. This has led to the development of many automated threat modeling tools including free options like Microsoft Threat Modeling Tool (TMT) [24], OWASP Threat Dragon (TD) [14], Threats Manager Studio [37], Mozilla SeaSponge [26], and PyTM [39], along with commercial options like SD Elements [9], IriusRisk [19], and ThreatModeler [41].

Previous work on threat modeling tools has focused on tool functionality and features [8, 31, 34], the amount of detail captured in the tool's threat model [8, 17, 34], tool usability [8, 31, 34], and the degree of tool adoption in the Software Development Life Cycle (SDLC) [4, 8, 34]. To the best of our knowledge, this study is the first to focus specifically on tool-generated threat reports—reports that enumerate potential threats in the evaluated systems. These reports are typically used to inform security requirements and identify design issues early in the SDLC to engineer more secure products [36]. More explicitly, we address the following research questions:

RQ1 What do users consider as actionable information in a threat report?
RQ2 In the current form of threat reports, how actionable is the information presented?

We use the term *actionable information* to refer to information that users can leverage to act, e.g., to develop a better understanding of security requirements either to elicit new security requirements or revisit existing ones, thus supporting secure system development. Examples of actionable information in threat reports include adequate information on threats that can be used to identify existing system vulnerabilities or critical system areas, or adequate information identifying critical threats on which attention and mitigation efforts should be focused.

Our study focuses on threat reports generated by two prominent threat modeling tools [8]: Microsoft TMT [44] and OWASP TD [14]. Among all existing tools, Microsoft TMT is the oldest (2016) and most mature tool with comprehensive documentation [44]. In contrast, OWASP TD is a relatively new (2020) open-source tool that is considered a strong contender to Microsoft TMT [4], notably with its feature-rich rule engine [8]. Both tools provide an intuitive interface to create system diagrams [34], which are then used to automatically apply rules from the tools' threat library to generate potential threats in a threat report. Additionally, both tools are available for free, are able to categorize threats based on the popular STRIDE classification [35] (see Table 1), and have been used by previous studies as standards for comparing different aspects of threat modeling tools [8, 17, 34] or to derive requirements for new tools [31].

Contributions: In this work, we conduct the first study on *actionable information in tool-generated threat reports* from users' perspective. The study involved semi-structured interviews with 15 participants from North America (primarily software developers and security experts) and included tasks related to identifying and assessing actionable pieces of information in two threat reports generated for an immunization healthcare system by two prominent threat modeling tools, namely Microsoft TMT and OWASP TD [8]. We present our findings within five overarching themes determined based on existing literature related to better threat modeling tool outputs, i.e., threat reports [8, 30, 34]; quality of suggestions, clarity of presentation, threat prioritization capability, security decision support, and adequate coverage (details are presented in Section 4.1). Our main contributions are as follows.

- We explore what users consider as actionable information in threat reports generated by threat modeling tools. Based on our findings, we conclude that above everything, users value detailed threat and mitigation suggestions as directly actionable information. Besides these, users consider a solid threat prioritization scheme and a statistical overview of useful insights as important supplementary actionable information.

- We assess the level of actionable information presented in threat reports. Our results reveal that the current state of reports lacks adequate coverage of actionable information necessary to make concrete decisions, primarily marred by the unsatisfactory quality of threat and mitigation suggestions and ways to prioritize critical threats.
- Based on users' needs identified in our analysis, we provide six recommendations to improve the current state of tool-generated threat reports. Our recommendations primarily involve detailed and specific threat and mitigation suggestions, a useful threat prioritization scheme, and consistent use of clearly defined terminologies to reduce the cognitive load on users and inspire higher confidence in the threat report contents.

This work helps gain a deeper understanding of the current state of automatic threat reporting capabilities of threat modeling tools. We expect that our findings and recommendations will help guide developers of existing and emerging threat modeling tools to generate reports that aid users in making informed security decisions.

2 RELATED WORK

Existing research on threat modeling tools has focused on tool functionality and features, not the generated threat reports. Previous work presented a taxonomy of diagram-based and text-based threat modeling tools with a set of 8 criteria for evaluation [34]; a comparison of Microsoft TMT and OWASP TD to assess the difference in their interface, system modeling capability, and threat generation approaches [8]; and proposed a set of requirements for new threat modeling tools derived by comparing their functionality [31]. Other works have focused on the details that can be captured or presented by threat modeling tools, including a systematic review of automated threat modeling techniques involving three tools to show how the threats generated by different tools contained different levels of detail [17], identifying that most tools provide options to extend the basic threat library and expanding the possibility of the types and number of threats that can be generated [34], and finding how the number of inputs needed to create a detailed DFD in Microsoft TMT can be overwhelming for smaller or less experienced development teams [8]. On the topic of adopting threat modeling in the SDLC, there is a scarcity of studies examining the efficacy of threat modeling tools in aiding security practices [8]. Bernsmed et al. [4] highlighted that aside from Microsoft TMT, threat modeling tools have yet to be adopted in Agile development practices. Shi et al. [34] identified important criteria to consider for encouraging threat modeling tool adoption in the SDLC, including tool interoperability, long-term support, and evolution.

Other works have focused on the usability of threat modeling tools. These include identifying how users can be inundated with the different settings and magnitude of information present when creating a model with Microsoft TMT, whereas OWASP TD has limited support for adjustable or custom diagram properties [31]. Other studies have highlighted how tools with diagram-based inputs are more intuitive despite text-based models being suited to better maintenance and analysis [34]. A study on tool interfaces [8] has found that Microsoft TMT has a steeper learning curve with many customization options and OWASP TD is more user-friendly due to its simpler layout and visuals.

Two recent studies have explored the current state of threat modeling tools. Verreydt et al. [43] focused on practices in the context of large Dutch organizations, whereas Thompson et al. [40] studied how security experts of medical devices conduct threat modeling. Both studies found the overall tool support for threat modeling to be inadequate and tool usage to be limited to creating diagrams. Threat modeling tools often require extensively detailed input to produce useful outputs [43], and the interpretation of this output can be challenging for users depending on their overall experience [40] and specifically their security expertise [43]. Adopting threat modeling tools can be complex, e.g., to integrate into a team's workflow [43] and often conflict with other evaluation tools used [40]. Such workflow integration issues can result in resistance to adopting security tools, and sub-optimal security practices or workarounds [2].

Table 1. STRIDE threat categories, example threats, and related security objectives [33]

Threat Category	Example	Related Objective
Spoofing	Impersonate a user or device within a system, e.g., by using their username or password	Authentication
Tampering	Maliciously modify, corrupt, or destroy data at rest or in transit, e.g., modifying data in a database	Integrity
Repudiation	Deny performing some malicious action due to lack of traceability, e.g., due to lack of auditing	Non-repudiation
Information Disclosure	Get unauthorized access to information at rest or in transit, e.g., to sensitive health information	Confidentiality
Denial of Service	Deny access to the services provided by a system, e.g., by overloading web servers with requests	Availability
Elevation of Privilege	Obtain higher privileged access to resources than intended, e.g., gaining admin or root privileges	Authorization

The quality of outputs generated by threat modeling tools and whether these support users in making security decisions is understudied. Bygdås et al. [8] briefly commented on how handling a PDF file output (OWASP TD) is easier than an HTML file output (Microsoft TMT), especially with OWASP TD's GitHub integration. However, they did not go further on specifics of the output or how it could be improved to support decision-making. Our study provides a more thorough evaluation of such outputs of threat modeling tools (i.e., threat reports) which is a gap in research.

We explore what users consider to be actionable information and assess the level of actionable information in threat reports. Our recommendations aim to aid security decision-making by improving the state of threat reports in existing and emerging threat modeling tools, grounded in data collected from software engineers and security experts.

3 THE CHOSEN THREAT MODELING TOOLS AND CASE STUDY SYSTEM

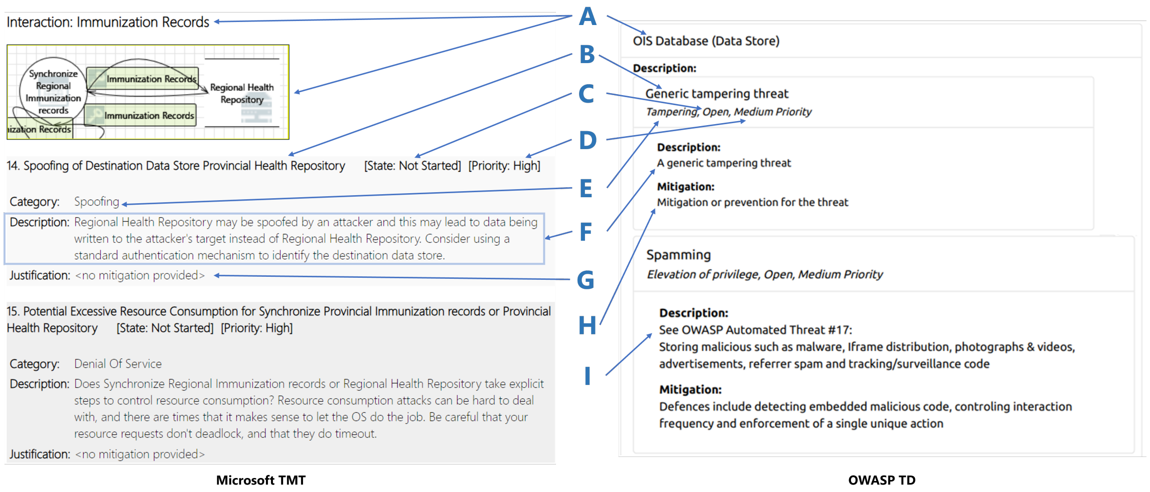
In this section, we present the technical background related to this study, including the two selected threat modeling tools and their threat report elements, data flow diagrams, STRIDE, and the healthcare case study system.

3.1 Microsoft TMT and OWASP TD

In this study, we use threat reports generated by two prominent threat modeling tools, namely Microsoft Threat Modeling Tool (TMT) and OWASP Threat Dragon (TD). Microsoft TMT is the most mature threat modeling tool available [8]. It is free, provides numerous configuration options and templates, and allows for detailed threat generation capabilities, albeit with a steeper learning curve compared to most other threat modeling tools [44]. Microsoft TMT threat reports are generated in the HTML (.htm) format and can only be exported locally. On the other hand, OWASP TD is designed as a multi-platform, lightweight tool that makes use of its threat library and a powerful rule engine [34], making it the second-most used threat modeling tool [8]. OWASP TD threat reports can be exported as a PDF (.pdf) or HTML (.html) file (only in v2.1.3) locally or through the GitHub integration for easier sharing.

Both tools take a system modeled as a Data Flow Diagram (DFD) as input and generate potential threats in a threat report categorized using the STRIDE model. DFDs allow representing the flow of data among different system components (processes and data stores) and external entities, and visualizing how data is used, processed, stored, and manipulated during operation [36]. The primary four components of a DFD are data flows, external entities, processes, and data stores [10] (see Table 4 in Appendix B for details). STRIDE is a mature and optimal approach for identifying threats to system components and their interactions [21]. It consists of six categories of threats [24], namely spoofing, tampering, repudiation, information disclosure, denial-of-service, and elevation of privilege (see Table 1).

Two fragments of the threat reports used in this study, one generated by Microsoft TMT and the other by OWASP TD for our case study system (see Section 3.2) are presented in Figure 1. The labels (A–I) represent elements of the threat reports discussed in this study. Note that labels A–F represent common elements between the two threat reports, whereas labels G–I are unique elements. We provide a brief description of these elements below:



Interaction: Immunization Records

14. Spoofing of Destination Data Store Provincial Health Repository [State: Not Started] [Priority: High]

Category: Spoofing

Description: Regional Health Repository may be spoofed by an attacker and this may lead to data being written to the attacker's target instead of Regional Health Repository. Consider using a standard authentication mechanism to identify the destination data store.

Justification: <no mitigation provided>

15. Potential Excessive Resource Consumption for Synchronize Provincial Immunization records or Provincial Health Repository [State: Not Started] [Priority: High]

Category: Denial Of Service

Description: Does Synchronize Regional Immunization records or Regional Health Repository take explicit steps to control resource consumption? Resource consumption attacks can be hard to deal with, and there are times that it makes sense to let the OS do the job. Be careful that your resource requests don't deadlock, and that they do timeout.

Justification: <no mitigation provided>

OIS Database (Data Store)

Description:

Generic tampering threat

Tampering, Open, Medium Priority

Description:
A generic tampering threat

Mitigation:
Mitigation or prevention for the threat

Spamming

Elevation of privilege, Open, Medium Priority

Description:
See OWASP Automated Threat #17:
Storing malicious such as malware, Iframe distribution, photographs & videos, advertisements, referrer spam and tracking/surveillance code

Mitigation:
Defences include detecting embedded malicious code, controling interaction frequency and enforcement of a single unique action

Fig. 1. Two example fragments of the threat reports included in the study: (left) Microsoft TMT, (right) OWASP TD (Note that these fragments are shown to highlight the different elements of each threat report, and are not presented for a direct comparison).

(A) **Analyzed interaction/component**: The target of analysis for each tool. Microsoft TMT targets the interactions between two DFD components (with an interaction snippet). OWASP TD targets individual DFD components.
(B) **Threat title**: A short title of the identified threat.
(C) **Threat state**: The current state of the threat, e.g., not started, mitigated.
(D) **Threat priority**: Priority assigned to the threat based on how critical it is.
(E) **Threat category**: The category of the threat based on STRIDE.
(F) **Threat description**: A brief description of the threat. Note that threats containing a generic description (e.g., "A generic tampering threat" in Figure 1) in the OWASP TD report are referred to as *generic threats* in this study.
(G) **Justification**: Justification behind why the threat exists or why it is assigned a certain priority level.
(H) **Mitigation description**: A brief description of the potential mitigation of a threat.
(I) **Automated threat numbers**: A numerical value referring to a threat type that is detailed in the OWASP Automated Threat Handbook [13] relating to unwanted automated usage threats to web applications. These threats contain more detail compared to *generic threats* and are referred to as *automated threats* in this study.

3.2 The Healthcare Case Study System

As a case study system, we selected an online immunization system (OIS), a software-based healthcare system that tracks user immunization records (e.g., COVID-19 vaccinations). The OIS is a real-world healthcare system in North America, the name and details of which have been anonymized for confidentiality purposes. Being a healthcare system, the OIS is of particular interest to us as threats to these systems often blur the line between security and safety due to the handling of sensitive patient data, making threat modeling challenging [1, 40]. Users can use the OIS mobile client or web client to record or update vaccination status and access up-to-date immunization records for themselves and their families. Further details on the OIS operation and the OIS DFD can be found in Appendix B.

4 METHODOLOGY

Ethical considerations. We designed and conducted an online semi-structured interview study approved by our institution's Research Ethics Board (REB). We did not collect any personally identifiable information (e.g., name, email address, social security number), and research data was stored in a password-protected computer accessible only to the research team. Interview audio recordings were deleted after transcription, and all data was anonymized.

Participant recruitment. We started recruitment by disseminating an online study invitation through emails, social media, and snowballing from the authors' network. To be eligible for the study, participants had to be familiar with at least one of the two threat modeling tools and be actively engaged in a role suited to performing threat modeling activities (e.g., software engineer/developer, system architect, security expert). Interested and eligible participants were presented with an informed consent form containing further study details, which was signed and returned to the researchers before each interview. We reached data saturation at 11 participants when our analysis did not produce new insights. We recruited 4 more participants to confirm saturation and then stopped recruitment as recommended for qualitative studies [15, 18]. The results discussed herein are from our group of 15 participants. Participation in our study was completely voluntary and participants did not receive compensation.

4.1 Interview

Question formulation. Our interview questions were created based on our previous experience with the tools, discussions with domain experts, and insights from previous work. Discussions with domain experts involved two rounds of semi-informal discussions; round 1 with two experts on threat modeling and system security, and round 2 with an expert on human-centric security. Discussion topics involved the current usage and usefulness of threat modeling tool outputs and their various usability aspects. The discussion sessions led to the potential lack of actionable information, guiding the formulation of our interview questions, that could be broadly grouped into five overarching clarity themes inspired by existing literature related to better threat modeling tool outputs (i.e., threat reports). Three of these themes relate directly to the purpose of threat reports, i.e., *quality of suggestions*, as threat modeling tool outputs are expected to aid and supplement the manual threat identification process [8], *threat prioritization capability*, as threat modeling tool outputs are expected to help evaluate critical threats [34], and *security decision support*, as threat modeling tool outputs are expected to help make decisions on determining security requirements and judiciously allocating security resources [34]. As a technical report that may be presented to non-experts, the *clarity of presentation* of the threat reports became another theme of interest [30]. Lastly, we involved questions related to the theme of *adequate coverage* to understand the shortcomings in existing threat reports and identify potential improvements.

Pilot testing. We pilot-tested our study with our colleagues and recruited two participants (both software engineers) to ensure that our questions were clear, ascertain if the study duration was enough, and identify any other issues with the methodology. The pilot test results led to minimal changes, e.g., merging two questions into one and improving the phrasing of one question, leading to a final set of 8 main questions that can be found in Appendix A. We didn't find any substantial difference between the data collected before and after the changes made based on the pilot study, and thus, data from the two software engineers involved in the pilot study is included in the results discussed herein.

Case study and threat report introduction. At the beginning of each interview session, participants first received a document providing an overview of the OIS healthcare case study system. Participants were asked to familiarize themselves with the OIS objectives, components, and component interactions, and how it processes and stores patient

Table 2. Example of codes that emerged from the inductive data analysis

Theme/Category	Threat Suggestions		Mitigation Suggestions	
	Microsoft TMT	OWASP TD	Microsoft TMT	OWASP TD
Quality of Suggestions	1. Basic technical information 2. Easy-to-follow descriptions 3. Threats in the context of the system	1. Generic threat description 2. Unspecific automated threats 3. Requires users to make assumptions	1. Enough to get started 2. Sometimes Missing 3. Lumped with threat description	1. Generic mitigation description 2. Needs additional details 3. Clearly labeled

immunization information, allowing access only to authorized users. Then participants were provided with two threat reports (see report fragments in Fig. 1), one generated for the OIS by Microsoft TMT (v7.0.8000.0) and the other by OWASP TD (v1.6.1) respectively. Participants received one threat report followed by the other, where the first report provided was alternated for each participant to avoid order bias.

Interview. Once the participants familiarized themselves with the OIS and the two threat reports, we followed up with a semi-structured interview. All interview questions were open-ended and participants' responses were probed further for the report content on which they were basing their answers. Interview sessions lasted 24 minutes on average. Following the interview, we collected demographic information using a post-study online questionnaire, including sex, age group, education level, job title, self-reported security expertise, and level of expertise with the two tools.

4.2 Analysis

We began our analysis by transcribing the audio recordings, removing all references that could identify a participant (e.g., profession/study level) and chronologically assigning an alias to the 15 participants (P1–P15) based on the order in which they were interviewed. Interview data was analyzed using a combination of deductive (theoretical) and inductive thematic analysis [6]. As a first step, we analyzed the data deductively based on the five pre-determined overarching clarity themes. Then, within these themes, the first author conducted inductive analysis with open coding resulting in an initial list of 71 codes. All three researchers then worked together to refine the codes and acquire insights grounded in the data. For example, within the overarching theme of *quality of suggestions*, we identified two categories of *threat suggestions* and *mitigation suggestions*, each containing codes distinct to Microsoft TMT and OWASP TD (see Table 2). Similarly, three categories of *relation of threats and system vulnerabilities*, *syntax*, and *semantics* were identified within the overarching theme of *clarity of presentation.* The rest of the codes were grouped directly within the other three overarching themes. Table 3 provides a summary of the results within each of the overarching themes.

4.3 Study Limitations

Our first 5 participants first read both threat reports and had access to both reports while answering a single round of interview questions. Recognizing that this could lead to order bias, we changed our approach for the last 10 participants. In the second approach, participants answered interview questions about one threat report before having access to the second report, with two separate rounds of interview questions, one for each report. We tried to minimize the order bias in both cases by alternating the first report read for subsequent participants, i.e., 8 participants read the Microsoft TMT report first, whereas 7 others read the OWASP TD report first. We did not find evidence that this modification caused variations in participants' opinions on how actionable the reports were or which report was preferred.

Even though we have reached saturation, our study includes a relatively small number of participants (n=15), thus our results may not be representative of all threat modeling tool users. Despite our efforts to recruit a diverse sample, our

Table 3. Summary of results

Theme	Category	Microsoft Threat Modeling tool threat report	OWASP Threat Dragon threat report
Quality of Suggestions (Section 5.1)	Threat suggestions	The threat suggestions are actionable with clear technical details about the threats, easy-to-follow threat descriptions, and threat descriptions are provided based on the system context.	The threat suggestions are not actionable as the majority of the threat descriptions contain generic information, requiring user assumptions. Automated threat descriptions are more detailed, yet they are not system-specific.
	Mitigation suggestions	The mitigation suggestions (when available) provide guidance as a starting point. However, all threats do not have mitigation suggestions and mitigation descriptions are difficult to find as they are lumped with threat descriptions.	The mitigation suggestions are not actionable as the majority of the mitigation descriptions contain generic information, missing sufficient detail. However, the descriptions are clearly labeled.
Clarity of Presentation (Section 5.2)	Relation of threats and system vuln.	The relation is clearly identifiable due to the threat title and diagram snippets highlighting the affected system component interactions with further details presented in the threat description.	The relation between threats and components is clear. However, the interaction between the components that often cause the vulnerability is not clear.
	Syntax	The syntax is clear and easy to follow, except the report lacks separate fields for threat and mitigation descriptions. Also, the report provides a helpful numerical threat summary as an overview.	The syntax is clear and easy to follow. However, the organization forces prolonged scrolling to find the desired information
	Semantics	Reports contain clear semantics for most report components, albeit with a few ambiguous terms (e.g., "justification" for Microsoft TMT, and "generic" for OWASP TD) and unspecified threat priority levels. Higher security expertise and detailed threat and mitigation descriptions are likely needed for a better interpretation of the contents.	
Threat Prioritization Capability (Section 5.3)		Reports assign a priority level to each threat, however, the assigned priority is the same for all threats ("high" for Microsoft TMT, "medium" for OWASP TD). Some users took alternative prioritization approaches, e.g., finding interactions or components associated with the most threats and some aggregated statistics.	
Capability to support security decisions (Section 5.4)		Concrete security decisions are difficult based on information provided in both threat reports due to the ambiguity of information, the overwhelming number of threats, and the unhelpful prioritization scheme. Users instead consider the threat reports useful as guidance for more detailed analysis as they provide a decent starting point.	
Adequate Coverage (Section 5.5)		Both reports do not provide adequate coverage for what is expected from a threat report. Information considered missing by users include: examples of attacks or countermeasures, links to external resources, a statistical summary of numerical insights, a useful priority scheme, and clear definition of confusing terminologies (See Section 5.5 for other shortcomings).	

sample is relatively young and may have limited experience with the tools. While we did not find any major differences in the responses provided by participants in different age groups, we have noticed that younger participants tended to focus on the technical details in the reports, whereas older participants tended to drive the discussion towards how such information may inform security incident management and response, and organizational practices. Lastly, we only studied two threat modeling tools. However, the two selected tools are typically used as representatives of the current state of threat modeling tools [8], and studying more tools may not necessarily have generated more insights.

5 RESULTS

Participant demographics. Among the 15 recruited participants, most participants' job titles were "software engineers" (n=9) or "security experts" (n=4) along with "system architects" (n=1) and "system designers" (n=1). Nine participants were in the 18-25 age group, five were in the 26-35 age group, and one was in the 36-45 age group. Six of our participants were females, and nine were males. Six participants had a Bachelor's degree, six had a Master's degree, and three had a

PhD. Most participants self-reported high security expertise, and all were familiar with at least one of the two threat modeling tools used in our study. Table 5 in Appendix C provides more details on the participant demographics.

Analysis results. Our data analysis results (summarized in Table 3) are organized by five overarching themes: *quality of suggestions, clarity of presentation, threat prioritization capability, security decision support,* and *adequate coverage.* When applicable, we provide the number of participants contributing to each theme for context only, not as a quantitative measure. Note that we highlight *participants' quotes* in italics and **key takeaways** from the results in boldface.

5.1 Quality of Suggestions

5.1.1 Threat Suggestions. In the Microsoft TMT threat report, **the threat suggestions provided are actionable with clear technical threat details and easy-to-follow threat descriptions. Threat descriptions are provided based on the context of the system.** First, the threat report contained actionable information; participants mentioned getting basic technical information about specific vulnerabilities from the threat reports that they could immediately act on since the threat descriptions often mention very specific technical terms that they could look up. P4 explained:

> *"So if I'm a practitioner, I'm looking at the [threat report] in Microsoft [TMT] and immediately I'm getting to the crux of an issue that I can solve ... The first thing I saw [is a] weakness in SSO authorization. Immediately, I know what that is ... Now you're telling me elevation of privilege ... Now you're giving me more specific technologies that could be causing that - OAuth2 ... I'm getting information that I can act on now."*

Second, the threat report contained easy-to-follow threat descriptions, which allowed participants to gain a rudimentary idea of the severity and exploitation methods of each threat. Third, the threat report identified threats in the context of the system based on the input system model (DFD). Coupled with a helpful snippet of the system model provided just above the threat descriptions, participants were able to identify the link between the described threats and which system components as well as which interactions between components were affected by each threat.

In the OWASP TD threat report, **the threat suggestions provided are not actionable as the majority of the threat descriptions contain generic information, requiring users to make assumptions. The automated threat descriptions are more detailed, yet they are not system-specific.** First, the majority (76%) of the threats in the report generated for the OIS were *generic threats*, containing only a generic threat description, which participants found unhelpful for taking action—the most useful information it provides is the STRIDE category to which the threat belonged (e.g., "a generic [STRIDE_category] threat," see Fig. 1). P10 mentioned their expectation for more details:

> *"So, after telling me that the title is a generic spoofing threat ... the description is pretty much the same [as the title], that's not really helpful. Okay, so now I know it's a spoofing threat. That's good. But I need more details. Like, that's what I would expect from a tool like this."*

Second, in the few cases where there were *automated threats*, participants deemed them as unspecific automated threats, as they found that the threats were not presented in the context of the system or its components. This required participants to manually relate each threat description to the system components to identify any vulnerable areas. Third, participants mentioned that understanding the suggestions for both *generic threats* and *automated threats* required users to make assumptions, which could lead to subjective interpretations of the severity of each threat and its exploitation process. Participants also mentioned that insufficient security expertise could be detrimental as a shallow understanding of the threats without further descriptions or links to external information resources could lead to misinterpretations of potential mitigations and/or improper implementations, which in turn would lead to further issues.

5.1.2 Mitigation Suggestions. In the Microsoft TMT threat report, **the mitigation suggestions (when available) provide guidance as a starting point. However, not all threats have mitigation suggestions and mitigation descriptions are difficult to find as they are lumped with threat descriptions.** First, combined with the detailed threat descriptions, participants found the mitigation suggestions (when available) enough to get started with their own research for more specific details. However, they also mentioned how some mitigation suggestions could be generic, e.g., "consider using a standard authentication mechanism to identify the destination data store" (see Fig. 1).

Second, participants mentioned how the suggested mitigations are sometimes missing for some threats. Participants highlighted how this omission could lead them to misinterpret these threats as ones that do not require mitigation. Third, participants mentioned how mitigation descriptions (when available) are always lumped with the threat description. Participants found this structure unexpected and could lead the mitigation information to go unnoticed. Indeed, some of our participants (n=4) failed in their task of identifying mitigation information in the threat report—they expected the "justification" field (as opposed to the "description" field in Fig. 1) to contain such information. P4 explained:

> *"Now there's one distinction that I feel is important and that is classifying what is a mitigation versus what is a [threat] description, because mixing those two, again, you're losing me. And some mitigations could be missed as well."*

In the OWASP TD threat report, **the mitigation suggestions are not actionable as the majority of the mitigation descriptions contain generic information, missing sufficient detail. However, the descriptions are clearly labeled.** First, participants found the generic mitigation description of *generic threats* ("Mitigation or Prevention for the threat", see Fig. 1) to be unhelpful in their current form, lacking necessary technical details upon which they could act. In contrast, threats classified as *automated threats* contained a more detailed mitigation strategy than the *generic threats*, matching participants' expectations. In terms of readability, participants preferred the OWASP TD threat report as they found the mitigation description to be clearly labeled and separated from the threat description.

Interestingly, our analysis shows that the overabundance of *generic threats* present at the beginning of the OWASP TD threat report and the need to scroll through a long list of threats led some participants (n=6) to completely miss the fact that the report contained *automated threats* with more detail. To illustrate, even for a small case study system like the OIS, the number of generic threats was about 76% (102 out of 133 total threats manually identified). P3 shared their frustration while scrolling through the long OWASP TD threat report containing mostly generic threats:

> *"Honestly, if there is like [a] generic information disclosure threat, I don't know, is there any need to include it here? Maybe ... have a list in the beginning. Like, okay, so these are all the generic threats that we have. This is just a waste of space. Generic, generic, generic."*

Another interesting finding was that external resources listed in the report could go unnoticed as the report does not provide external links for such resources. The OWASP website [29] offers further details on the automated threats; searching the website using the OWASP TD automated threat numbers (e.g., "see OWASP automated threat #17" in Fig. 1) provides more details such as an example diagram of the process described in detail, and CAPEC and CWE references. However, the OWASP TD threat report does not include external links for such additional details, and users would have to take it upon themselves to manually search for them (that is of course assuming that users are aware that such resource exists). This disconnect led the majority (n=13) of our participants to miss out on the potential benefits that could have been provided by the additional steps of searching for and reading the details on the OWASP website. Similarly, Microsoft TMT's website [23] provides more detailed information for each countermeasure outlined in the report. However, assuming participants are aware such information exists, finding it requires manual online searching (e.g., using the mitigation suggestion "Consider using a standard authentication mechanism to identify the destination data store" in Fig. 1). In the current form where mitigation suggestions are lumped with threat descriptions, these additional resources are hidden from the users' vision, and none of our participants thought about isolating the text that relates to mitigations and manually searching for it online to find further information.

5.2 Clarity of Presentation

5.2.1 Relation between Threats and System Vulnerabilities. In the Microsoft TMT threat report, **the relation is clearly identifiable due to the threat title and diagram snippets highlighting the affected system component interactions with further details presented in the threat description.** Participants found the report presentation to be very useful in identifying vulnerable system areas as they could combine the threat title and the associated snippet presenting the component interactions to know exactly which parts of the system were analyzed for each threat. Participants also liked that the threat descriptions contained detailed information on vulnerable system interactions, which allowed them to understand how the vulnerable component interactions were leading to the threat's existence.

Conversely, in the OWASP TD threat report, **the relation between threats and components is clear. However, the interaction between the components that often causes the vulnerability is not.** Participants mentioned that the report's grouping of all threats for a system component under the component allowed them to find the vulnerable component. However, they found the presentation to be not enough for finding vulnerable system areas. They mentioned how each component might be interacting with multiple other components, and threats affecting the interactions between two components would be grouped under both components without any context about the interaction between those components (often the cause of the vulnerability). Furthermore, participants mentioned how this can lead to redundant reporting of threats, exacerbating the problem of an already long list of identified threats.

5.2.2 Syntax. For the Microsoft TMT threat report, participants mentioned how **the syntax is clear and easy to follow, except the report lacks separate fields for threat and mitigation descriptions. Also, participants found the numerical threat summary to be helpful to gain an overview of the system status.** First, participants found the syntax to be easy to follow in a way that someone with no prior experience with threat reports could easily follow along with the content due to its relative simplicity. Second, participants liked that the threat report presented a numerical threat summary at the beginning of the report, which the participants found helpful as a useful overview before diving deeper into the content of the report. However, as mentioned before in Section 5.1.2, participants found the lack of separate fields for mitigation suggestions in the Microsoft TMT report to be confusing; participants had trouble finding the exact mitigation information and sometimes needed to read through a long threat description to identify the parts that were related to mitigations and not threats.

For the OWASP TD threat report, participants mentioned that **the syntax is clear and easy to follow. However, the organization forces prolonged scrolling to find the desired information.** Similar to the Microsoft TMT threat report, participants found the syntax for the OWASP TD threat report to be easy to follow without any glaring downsides. However, they mentioned the issue of prolonged scrolling, where they sometimes needed to scroll through lengthy text to find the right information, and that the issue persists due to an inefficient syntactical organization. Note that this issue has been fixed in the current version (v2.1.3) of OWASP TD.

5.2.3 Semantics. When discussing the report's semantics, **participants mentioned that both threat reports contain clear semantics for most report components, albeit with a few ambiguous terms and unspecified threat priority levels.** Participants generally found both threat reports to have reasonably clear semantics for most of their components, i.e., terminologies, descriptions, and figures, except with two main semantic issues. First, both reports included many ambiguous terms, which made it difficult for participants to interpret the reports' information objectively. For example, participants identified the “state" and “justification" fields in the Microsoft TMT report and the “open", “generic", and “automated threat numbers" terminologies in the OWASP TD report as being vague and open to interpretation. Second, participants highlighted how the “priority" field for both reports contained unclear threat priority levels without a standardized definition or explanation of different threat levels. They noted how just knowing that a threat is of “high” priority was of no practical value since they did not know the whole spectrum of priority levels.

Participants highlighted that **higher security expertise and detailed threat and mitigation descriptions are likely needed for better interpretation of the content.** Participants expected that having higher security expertise could help them better interpret the threat and mitigation information in the reports. They also expressed a need for more detailed information of the threat and mitigation descriptions, as the currently available information lacks depth and is unhelpful in formulating an understanding of a definitive solution.

5.3 Threat Prioritization Capability

Participants took one of two approaches to prioritize critical threats from the report information: (1) assessing available threat priority information (i.e., the “threat priority" field) and (2) looking for alternative metrics in the threat reports.

Participants (n=12) generally did not find the information included in either threat report to be constructive in supporting their efforts to prioritize threats. As a reason, they mentioned how **both threat reports assign a priority level to each threat identified, yet the assigned priority is the same for all threats**—Microsoft TMT assigns all threats “High” priority and OWASP TD assigns all threats “Medium” priority by default. P5 explained:

> *“Looking through [Microsoft TMT threat report] at least quickly, it seems all these [threats] are priority high. Whereas with the OWASP one, I can see [that] they have medium threats ... I don’t know if there [are] any high ones in here"—participant conducts a word search for ‘high’—“So, they’re all medium, which is quite interesting. It doesn’t give me a very good indication of what I should initially put my effort into.”*

In contrast, some participants (n=3) mentioned that they would not rely on the uniform threat priority levels in the reports that do not reflect the actual priority for each threat. **Some users took alternative prioritization approaches, e.g., finding interactions or components associated with the most threats and some aggregated statistics.** First, we found that participants would prioritize the interactions (Microsoft TMT) or components (OWASP TD) that have the most threats associated with them. Participants preferred prioritizing the threats nested under a system component in OWASP TD, as dealing with these threats may allow them to ensure that at least this component was secure. They mentioned how doing so is difficult in the Microsoft TMT report, as a component can be included in many different interactions sprinkled throughout the report. Second, participants mentioned that having some aggregated statistics at the beginning of each report, e.g., about which components had the most threats or which threats affected the most components, would be helpful in their endeavor of prioritizing critical components or threats that need attention.

5.4 Security Decision Support

5.4.1 Capability to Inform Security Decisions. Participants unanimously indicated that **concrete security decisions based on information provided in both threat reports are difficult due to the ambiguity of information, the overwhelming number of threats, and the unhelpful prioritization scheme.** First, they mentioned how both the threat and mitigation descriptions, as well as some terms, were ambiguous and not sufficiently detailed. They mentioned how they would not be confident in their understanding of the underlying system vulnerabilities or potential mitigations, nor would they be able to reach a concrete decision. Second, they mentioned that dealing with the high number of threats was overwhelming and the futile prioritization scheme in the reports exacerbated the issue. With all threats assigned the same prioritization level by default, participants had trouble making concrete decisions when trying to narrow down critical threats for assessment or identify critical components for protection.

Users instead consider the threat reports useful as guidance for more detailed analysis as they provide a decent starting point. They deemed all threats from the Microsoft TMT threat report and the *automated threats* from the OWASP TD threat report to be enough to provide a starting point for further security analysis. According to our participants, a concrete security decision would require steps beyond reviewing the threat reports, highlighting other activities such as further report scrutiny and conducting more rigorous security analyses.

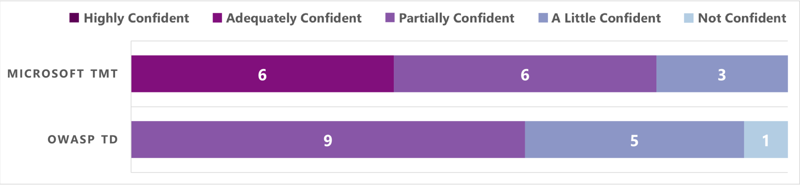


Fig. 2. Level of confidence in making a security decision using the threat reports as the only resource.

To gain further understanding of participants' confidence in making concrete security decisions based on the current state of the threat reports, we asked participants to rate their level of confidence (See Fig. 2) on a 5-point semantic scale (1: Not confident – 5: Highly Confident). Among the 15 participants, the majority (n=12) were either "Adequately Confident" (n=6) or "Partially Confident" (n=6) in making a security decision based on the information provided in the Microsoft TMT report. As reasons, participants primarily highlighted the detailed and system-specific threat suggestions, actionable threat and mitigation information, and helpful guidance from DFD snippets detailed above. Conversely, most participants (n=9) were at most "Partially Confident" in making a security decision by relying on the information in the OWASP TD report. Participants noted the lack of specificity in the suggestions and the existence of too many generic threats as the primary reasons behind that. It is important to note that none of the tool-generated threat reports were able to inspire the highest confidence level in the users, suggesting that there is still room for growth for these tools.

5.5 Adequate Coverage

Considering the overall information presented, all participants found both threat reports to be lacking in a few areas and unanimously concluded that **both reports do not provide adequate coverage for what is expected from a threat report**. However, they found the Microsoft TMT report to be more complete than the OWASP TD report for the reasons identified in Section 5.4. In what follows, we briefly summarize the shortcomings identified by our analysis.

To provide better coverage, participants wanted both reports to include examples of attacks or countermeasures explaining how the identified threats can be realized or mitigated; links to additional resources to provide more insights; a statistical summary; a useful priority scheme; an improved report layout for better readability, syntax, and creative organization (e.g., threats sorted by priority); a legend or table for the different STRIDE categories, terminologies, and levels of priority; clear definition of confusing terms in the report (e.g., "justification" or "open"); more information about how the threats and mitigation information are gathered; threat database transparency, e.g., whether they are static (unchanged or rarely changed) or dynamic (frequently updated), and how frequently they are updated.

Participants also believed that many shortcomings of each threat report (mentioned in Sections 5.1–5.4) could be resolved by taking some pointers from the other. For example, participants believed that the OWASP TD report could be significantly improved by following Microsoft TMT's example in providing more specific threat and mitigation information, better readability (e.g., with the use of DFD snippets), and top-level general statistics. In contrast, participants expected the Microsoft TMT report to include mitigation information for all threats, separate mitigations from threat descriptions, and provide more specific mitigation suggestions; while they liked Microsoft TMT's detailed system-specific threat suggestions, they were greatly disappointed by the unspecific and minimal countermeasure details.

6 DISCUSSION

In this section, we discuss our findings related to the current state of tool-generated threat reports and provide six recommendations based on our analysis of the results for improving the state of existing and emerging tools.

6.1 Current state of threat reports

RQ1: What do users consider as actionable information in a threat report? Above everything, users value the threat details and mitigation suggestions as directly actionable information. Details related to threat suggestions include: (1) a *descriptive threat title*, preferably with helpful diagrams, that allows the users to map the threats to vulnerable system components and/or interactions, (2) a *detailed threat description* that enables the users to get a deeper understanding of the exploitation process of the threat along with its potential severity, and (3) *links to frequently-updated external resources* to allow the users to conduct further research. Similarly, details related to mitigation suggestions include: (1) a *detailed mitigation description* that enables the users to understand how the suggested mitigation helps prevent or mitigate the associated threat, (2) *more specific mitigation descriptions* that allow the user to get an idea about industry-standard defensive techniques, with links to frequently updated external resources for further research.

Aside from detailed threat and mitigation suggestions, users expect threat reports to include meaningful threat priority levels assigned to the identified threats, as well as a statistical overview summarizing critical insights as supplementary actionable information. A useful threat prioritization scheme with well-defined threat priority levels assigned to different threats can help users prioritize high-priority threats from the overwhelmingly large number of threats typically identified in a threat report. Additionally, a statistical overview at the beginning of the threat report containing a numerical summary of information regarding *critical threats*, *critical countermeasures*, and *critical components* (see *Recommendation 6* below for more details) could allow users to get some insights before diving deeper into the threat report details and prove to be useful information in their decision-making process.

While we do not have enough data to make conclusions on how security expertise may relate to the assessment of actionable information in the reports, we observed that participants who reported having higher security expertise were able to more easily handle the complex information and technical jargon in the Microsoft TMT report. Furthermore, participants with the highest reported security expertise found the Microsoft TMT report to be more informative and actionable while the OWASP TD report suggestions were considered unusable. Conversely, participants who reported having less security expertise found the OWASP TD threat report to be more readable due to its inherent simplicity in formatting and description. Lastly, participants who reported being familiar with at least one of the threat modeling tools preferred having the Microsoft TMT report for their analysis due to the greater level of detail provided. These observations align with the findings of Bygdas et al. [8], determining that Microsoft TMT takes a more elaborate approach to generating an intricate and comprehensive threat report, whereas OWASP TD is focused on simplicity, user experience, and tool maintenance.

RQ2: In the current form of threat reports, how actionable is the information presented? Based on our findings, we conclude that the current state of threat reports lacks *adequate coverage* of actionable information necessary to make concrete decisions. While the reasonable *clarity of presentation* contributes to finding the actionable information in the threat reports, the unsatisfactory *quality of suggestions* and unhelpful *threat prioritization scheme* lead users to have low confidence in their *security decisions.* Some of the key issues identified by our analysis are presented below. First, our findings indicate that the current state of threat and mitigation suggestions lack important details that contribute to tool usability such as concrete examples of vulnerabilities and solutions with explanations [3, 27, 38] and external links to other helpful resources [38]. Second, the lack of specificity in the threat suggestions (e.g., not being system-specific) and mitigation suggestions (e.g., generic solutions) may reduce their potential impact [38]. Third, the supplementary output information not meeting user expectations (e.g., unhelpful threat priority levels, missing mitigation information) and specific needs (e.g., priority customization options) may limit their effectiveness [38]. Fourth, the increased complexity of tool outputs stemming from many presentation, phrasing, and organizational issues such as overuse of technical jargon, unclear (e.g., "justification") or undefined (e.g., "generic") terminologies, and long threat reports without a summary of useful insights can make users feel encumbered, discouraging tool adoption [3, 27, 38].

6.2 Recommendations for Improvement

To improve threat modeling tools that generate these threat reports based on the needs identified by our data analysis, we provide six recommendations and their relations to the overarching themes. Note that when adopting any of our recommendations, tool developers should carefully consider the varying levels of security expertise of their target users and their needs, and changes made based on these recommendations should not introduce unnecessary complexity.

Recommendation 1: The tools should provide threat suggestions that are detailed and system-specific, supplemented with external links or examples for details. In their current state, threat reports generated by threat modeling tools often include only a brief threat description without any context about how the threat is related to system components and/or interactions. We argue that aggregating different threats under a component does not provide a holistic picture as many threats exist due to the interaction between two or more components. Thus, we suggest that threat descriptions be system-specific, outlining the components and interactions involved in the existence of the threat. These details could be supplemented by external links to relevant reputable and frequently updated resources (e.g., US National Vulnerability Database (NVD) [28] for a better understanding of known vulnerabilities, advisories, solutions, and tools) or an example of how a threat exploits a vulnerability in the threat report. References to security design patterns (e.g., [11, 12, 32, 42]) can help provide further helpful details such as threat dynamics, implementation hints, and example applications for suggested mitigations. Overall, we expect this recommendation to result in the users obtaining higher confidence in their understanding of the threats and making any related decisions, improving the themes of *quality of suggestions*, *clarity of presentation*, *security decision support*, and *adequate coverage*.

Recommendation 2: The tools should provide more specific mitigation suggestions for all outlined threats, preferably guiding users to find industry-standard solutions. The current state of mitigation suggestions provides users with only a brief, and often vague, description of potential mitigations, leaving the task of determining the next steps to the users' security expertise and interpretation. First, tools should ensure that each threat description is accompanied by a mitigation description and that the mitigation description is context-specific, i.e., it relates directly to the outlined threats and affected components and/or interactions. Next, the tools' mitigation suggestions should provide updated industry-standard mitigation solutions (e.g., the advanced encryption standard (AES) 128-bit for data encryption) directly in the report or guide users to relevant external resources through links. The use of security design patterns mentioned above can play a big role in this regard, allowing the tools to suggest well-established mitigation solutions for known security threats. Alternatively, links to frequently updated external resources (e.g., MITRE's D3FEND™ matrix [25]) could help users find further information to consolidate their understanding of the suggested and alternative mitigation options. We expect this recommendation to improve the themes of *quality of suggestions*, *clarity of presentation*, *security decision support*, and *adequate coverage*.

Recommendation 3: The tools should allow users to choose a critical security objective at the start of the threat modeling process for a useful prioritization scheme. The current threat prioritization schemes in threat reports include assigning the same priority level to all threats, thus users end up with an overwhelming number of threats to deal with. Assigning specific priority levels to each threat is a difficult task to automate due to the numerous different contextual information that must be taken into account for each system. Instead, we suggest that the tools allow users to select a security objective critical to the system's mission, enabling the tool's internal threat classification scheme to use this information for prioritizing threats. An example of this is a user selecting "confidentiality" as the critical security objective for a medical system that contains sensitive patient information. A tool using the STRIDE classification can then relate "confidentiality" to the threat category of "information disclosure" (see Table 1), thereby assigning the highest priority to any threats under that category. The remaining threats could be assigned a label similar to "unassigned" to explicitly denote that these threats have not been assigned any priority and need manual intervention. This approach would allow the users to quickly isolate threats related to their system's critical security objectives, allowing the development of targeted mitigations. We expect this recommendation to improve the themes of *threat prioritization capability*, *security decision support*, and *adequate coverage* of threat modeling tools.

Recommendation 4: The tools should allow specifying threat priority levels tailored to the users' needs. Currently, the tools do not include any definition of the different priority levels assigned to each threat, nor can any tool-defined priority levels apply to all the different types of systems that users will potentially analyze. Therefore, we suggest that the tools allow users to define their own priority levels (e.g., "none", "moderate", "high", or even "critical"), the details of which would determined for the type of system analyzed based on discussions and agreements among all team members in a project. Coupled with *Recommendation 3*, users can then select the different user-defined priority levels to be assigned to each security objective, so that threats related to each objective are assigned priorities accordingly. For example, in the context of a "bus scheduling system", threats related to the most important "availability" objective can be assigned "critical" priority, while threats related to "integrity" and "confidentiality" objectives can be assigned "high" and "none" priorities, respectively. Thus, a combination of these recommendations will allow the tools to provide a prioritization scheme that is tailored to the user's needs. Overall, we expect the combination of Recommendations 3 and 4 to improve the themes of *threat prioritization capability*, *security decision support*, and *adequate coverage*.

Recommendation 5: The tools should make consistent use of clearly defined terminologies throughout the reports. The threat reports in their current form are not self-contained and do not contain details on their terminologies. While the creators of the threat reports may be able to learn the terminologies from the tool's documentation, other users of such threat reports, e.g., managers and investors, may not have the same opportunity. Thus, we suggest the reports choose intuitive terminologies expected by a user and only provide additional details on complex terminologies to avoid clutter. Examples include clear definitions of the different states that a threat can be in (e.g., "unmitigated", "under consideration", and "mitigated"), the different threat priority levels (user-defined descriptions if based on Recommendation 4), the categories of the underlying classification scheme (e.g., "spoofing" and "repudiation" in STRIDE), and any other terms that are not self-explanatory. While such clear definitions and their consistent use throughout the report may not be directly related to actionable information, these will reduce the cognitive load on users (e.g., users from different backgrounds and with varying levels of security expertise) and allow them to focus on the task at hand. We expect this recommendation to improve the themes of *clarity of presentation* and *adequate coverage*.

Recommendation 6: The tools should provide a statistical summary of useful insights at the start of the report. In their current form, the reports lack an overview of useful insights related to the threats. However, users expect a high-level summary at the start of the report similar to an “abstract” or “executive summary”. At a minimum, we suggest that the tools report on the numerical insights such as, (1) *critical threats* that affect a high number of system components and/or interactions, the mitigation of which may protect many components at once, (2) *critical countermeasures* that provide protection against a high number of threats, the deployment of which may protect against many threats at once, and (3) *critical components* that are affected by a high number of threats, the protection of which may require more resources and effort than others. Additionally, tools should summarize the number of threats in each priority level, e.g., 13 “high” priority threats, and threats in different states, e.g., how many threats have been “mitigated” or are “unmitigated (Microsoft TMT does this to an extent). This information could help users gain an overview of the security posture [20] of their system before they dive deeper into the details of the report, and can help them keep track of the improvements in their security posture as new reports are generated after mitigating some threats. We expect this recommendation to improve the themes of *clarity of presentation*, *security decision support*, and *adequate coverage*.

While we argue that our recommendations would apply to most existing and emerging threat modeling tools, we strive for ideal scenarios that are not to be followed blindly. Instead, we urge the developers of threat modeling tools to use our findings and recommendations as guidance to assess and prioritize improving different aspects of their tools.

7 CONCLUSION AND FUTURE WORK

We interviewed 15 participants to explore the information they consider as “actionable" in tool-generated threat reports and to evaluate the current state of actionable information in threat reports. Participants considered the details presented in the threat and mitigation suggestions as directly actionable information. They also deemed a useful threat prioritization scheme and a statistical overview of insights as important supplementary actionable information. However, based on our analysis, we conclude that the current state of threat reports lacks adequate coverage of actionable information necessary for users to make concrete decisions and leads to low confidence in the report content. This is primarily due to the unsatisfactory quality of suggestions and unhelpful threat prioritization capability. Reflecting on our results, we developed six recommendations aimed at improving the quality of threat reports.

Our analysis focuses on the reports generated by the two most prominent threat modeling tools, however, future work could address a wider range of threat modeling tools to identify potential systemic issues. Another possible future research direction could be to explore whether the usability of threat modeling tools affects the quality of the generated threat report, *e.g.,* exploring whether the user would be able to generate a more actionable threat report if the threat modeling tool was more usable. Another line of research could focus on how the security expertise and/or work experience of security experts could affect their evaluation of threat modeling tools and their outputs. In our future work, we plan to create designs for threat modeling tools guided by our recommendations (Sec. 6.2) and apply usability design principles, principles of Universal Design [7] to ensure usability for a diverse range of population, and the Cognitive Dimensions of Notations framework [5] to guide us when making trade-off design decisions. We will employ an iterative design/test/redesign process to improve our designs and ensure their usability.

ACKNOWLEDGMENTS

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Hussain Almohri, Long Cheng, Danfeng Yao, and Homa Alemzadeh. 2017. On Threat Modeling and Mitigation of Medical Cyber-Physical Systems. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. 114–119. https://doi.org/10.1109/CHASE.2017.69

[2] Hala Assal and Sonia Chiasson. 2018. Security in the software development lifecycle. In *Fourteenth symposium on usable privacy and security (SOUPS 2018)*. 281–296.

[3] Hala Assal and Sonia Chiasson. 2019. *'Think Secure from the Beginning'*: A Survey with Software Developers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300519

[4] Karin Bernsmed, Daniela Soares Cruzes, Martin Gilje Jaatun, and Monica Iovan. 2022. Adopting threat modelling in agile software development projects. *Journal of Systems and Software* 183 (2022), 111090.

[5] Alan Blackwell and Thomas Green. 2003. Notational Systems–the Cognitive Dimensions of Notations Framework. *HCI models, theories, and frameworks: toward an interdisciplinary science. Morgan Kaufmann* 234 (2003).

[6] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[7] Sheryl Burgstahler. 2009. Universal Design: Process, Principles, and Applications. *DO-IT* (2009).

[8] Erlend Bygdås, Lars Andreassen Jaatun, Stian Brekken Antonsen, Anders Ringen, and Erlend Eiring. 2021. Evaluating Threat Modeling Tools: Microsoft TMT versus OWASP Threat Dragon. In *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. 1–7. https://doi.org/10.1109/CyberSA52016.2021.9478215

[9] Security Compass. [n. d.]. Efficient Threat Modeling for Secure and Compliant Software at Scale. https://www.securitycompass.com/sdelements/threat-modeling/ (accessed on 2024-05-04).

[10] Tom DeMarco. 2001. *Structure Analysis and System Specification*. Springer Berlin Heidelberg, 255–288. https://doi.org/10.1007/978-3-642-48354-7_9

[11] Eduardo B. Fernandez, Nobukazu Yoshioka, Hironori Washizaki, and Joseph Yoder. 2022. Abstract security patterns and the design of secure systems. *Cybersecurity* 5, 1 (April 2022), 7.

[12] Eduardo Fernandez-Buglioni. 2013. *Security Patterns in Practice: Designing Secure Architectures Using Software Patterns* (first ed.). Wiley Publishing.

[13] OWASP Foundation. [n. d.]. OWASP Automated Threat Handbook. https://github.com/OWASP/www-project-automated-threats-to-web-applications/tree/master/assets/files/EN accessed on 2024-05-04.

[14] OWASP Foundation. [n. d.]. OWASP Threat Dragon. https://owasp.org/www-project-threat-dragon/ (accessed on 2024-04-28).

[15] Barney Glaser and Anselm Strauss. 2017. *Discovery of Grounded Theory: Strategies for Qualitative Research*. Routledge, New York. https://doi.org/10.4324/9780203793206

[16] Google. [n. d.]. Threat Models and Cloud Security. https://cloud.withgoogle.com/cloudsecurity/podcast/threat-models-and-cloud-security/ (accessed on 2024-05-04).

[17] Daniele Granata and Massimiliano Rak. 2023. Systematic analysis of automated threat modelling techniques: Comparison of open-source tools. *Software Quality Journal* (2023), 1–37.

[18] Greg Guest, Kathleen M. MacQueen, and Emily E. Namey. 2012. *Applied Thematic Analysis*. SAGE.

[19] IriusRisk. [n. d.]. Automated Threat Modeling Tool. https://www.iriusrisk.com/ (accessed on 2024-05-04).

[20] Arnold Johnson, Kelley Dempsey, Ron Ross, Sarbari Gupta, and Dennis Bailey. 2019. *Guide for security-focused configuration management of information systems*. Technical Report NIST SP 800-128. National Institute of Standards and Technology, Gaithersburg, MD. NIST SP 800–128 pages. https://doi.org/10.6028/NIST.SP.800-128

[21] Kyoung Ho Kim, Kyounggon Kim, and Huy Kang Kim. 2022. STRIDE-based threat modeling and DREAD evaluation for the distributed control system in the oil refinery. *ETRI Journal* 44, 6 (2022), 991–1003.

[22] Microsoft. [n. d.]. Microsoft Security Development Lifecycle Practices. https://www.microsoft.com/en-us/securityengineering/sdl/practices (accessed on 2024-05-04).

[23] Microsoft. [n. d.]. Microsoft Threat Modeling Tool mitigations. https://learn.microsoft.com/en-us/azure/security/develop/threat-modeling-tool-mitigations Accessed on 2024-05-04.

[24] Microsoft. 2022. Threats - Microsoft Threat Modeling Tool - Azure. https://learn.microsoft.com/en-us/azure/security/develop/threat-modeling-tool-threats Accessed on 2024-05-04.

[25] MITRE. [n. d.]. MITRE D3FEND Knowledge Graph. https://d3fend.mitre.org/ (accessed on 2024-04-30).

[26] Mozilla. [n. d.]. SeaSponge. https://mozilla.github.io/seasponge/#/about (accessed on 2024-05-04).

[27] Ivoline C. Ngong, Brad Stenger, Joseph P. Near, and Yuanyuan Feng. 2024. Evaluating the Usability of Differential Privacy Tools with Data Practitioners. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX Association. (in press).

[28] NIST. [n. d.]. NVD - Search and Statistics. https://nvd.nist.gov/vuln/search (accessed on 2024-04-30).

[29] OWASP. [n. d.]. OWASP Automated Threats to Web Applications. https://owasp.org/www-project-automated-threats-to-web-applications/ Accessed on 2024-05-04.

[30] Dietrich Rathjens. 1985. The seven components of clarity in technical writing. *IEEE transactions on professional communication* 4 (1985), 42–46.

[31] Andreas Schaad and Tobias Reski. 2019. "Open Weakness and Vulnerability Modeler" (OVVL): An Updated Approach to Threat Modeling. In *16th International Joint Conference on e-Business and Telecommunications (ICETE 2019)*. 417–424. https://doi.org/10.5220/0007919004170424

[32] Markus Schumacher, Eduardo Fernandez-Buglioni, Duane Hybertson, Frank Buschmann, and Peter Sommerla. 2006. *Security Patterns: Integrating Security and Systems Engineering.* Wiley.

[33] Microsoft Security. 2007. Microsoft Security Blog | STRIDE Chart. https://www.microsoft.com/en-us/security/blog/2007/09/11/stride-chart/.

[34] Zhenpeng Shi, Kalman Graffi, David Starobinski, and Nikolay Matyunin. 2022. Threat Modeling Tools: A Taxonomy. *IEEE Security & Privacy* 20, 4 (July 2022), 29–39. https://doi.org/10.1109/MSEC.2021.3125229

[35] Adam Shostack. 2008. Experiences Threat Modeling at Microsoft. *MODSEC@ MoDELS* 2008 (2008), 35.

[36] Adam Shostack. 2014. *Threat Modeling: Designing for Security.* John Wiley & Sons. Google-Books-ID: YiHcAgAAQBAJ.

[37] Simone. 2020. Threats Manager Studio. https://threatsmanager.com/ (accessed on 2024-05-04).

[38] Mohammad Tahaei, Kami Vaniea, Konstantin (Kosta) Beznosov, and Maria K Wolters. 2021. Security Notifications in Static Analysis Tools: Developers' Attitudes, Comprehension, and Ability to Act on Them. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* ACM, Yokohama Japan, 1–17. https://doi.org/10.1145/3411764.3445616

[39] Izar Tarandach. 2024. izar/pytm. https://github.com/izar/pytm (accessed on 2024-05-04).

[40] Ronald Thompson, Madeline McLaughlin, Carson Powers, and Daniel Votipka. 2024. "There are rabbit holes I want to go down that I'm not allowed to go down": An Investigation of Security Expert Threat Modeling Practices for Medical Devices. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association. (in press).

[41] ThreatModeler. [n. d.]. Automated Threat Modeling Solution. https://threatmodeler.com/ (accessed on 2024-05-04).

[42] Anton V. Uzunov, Katrina Falkner, and Eduardo B. Fernandez. 2015. A comprehensive pattern-oriented approach to engineering security methodologies. *Information and Software Technology* 57 (2015), 217–247.

[43] Stef Verreydt, Koen Yskout, Laurens Sion, and Wouter Joosen. 2024. Threat modeling state of practice in Dutch organizations. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX Association. (in press).

[44] Imano Williams and Xiaohong Yuan. 2015. Evaluating the effectiveness of Microsoft threat modeling tool. In *Proceedings of the 2015 Information Security Curriculum Development Conference (InfoSec '15)*. 1–6. https://doi.org/10.1145/2885990.2885999

[45] Wenjun Xiong and Robert Lagerström. 2019. Threat modeling – A systematic literature review. *Computers & Security* 84 (July 2019), 53–69. https://doi.org/10.1016/j.cose.2019.03.010

Appendices

A INTERVIEW QUESTIONS

The main interview questions used in the study are presented below. Note that these questions are presented as a guide only and participants' responses may have led to other related questions:

- Do you think that the information presented on the potential threats is actionable and of satisfactory quality? Why (not)?
- Do you think that the information presented on the suggested countermeasures (if available) is actionable and of satisfactory quality? Why (not)?
- From the threat report, do you know exactly where in your system the vulnerabilities lie (e.g., if you were to pick a threat at random)? Why (not)?
- Are all the details about each threat clearly presented (e.g., do you need further security expertise to interpret the description of some of the terminologies)? Why (not)?
- With limited resources (e.g., budget, time, manpower), do you know which threats to prioritize to facilitate further security analysis (e.g., risk analysis)? How/Why (not)?
- If the threat report is all the information that you have, can you that information to make a security decision (e.g., about system security requirements, or potential mitigations, or where to spend your resources to protect critical system functionality)? Why (not)?
- On a scale of 1 (not confident) – 5 (highly confident), what would be your level of confidence in making the decision? Why?
- Do you consider the threat report to be complete? Why (not)? If not, what else would you like to see to be included in the report?

Note that, Clarification for terminologies used in the questions and/or the threat reports included the following definitions and resources used for the definitions:

- Threats and threat modeling [36].
- Data Flow Diagrams (DFD) [10].
- STRIDE - Spoofing, Tampering, Repudiation, Information Disclosure, Denial-of-Service, and Elevation of Privilege [24].

B CASE STUDY SYSTEM DESCRIPTION AND THE CORRESPONDING DATA FLOW DIAGRAM

A high-level overview of the online immunization system (OIS) Architecture is shown in Fig. 3. Users can use the OIS mobile client or web client to interface with the OIS server to (1) create a user account (2) record or update vaccination status, and (3) access up-to-date immunization records for themselves and their families. Individual records are stored and retrieved from the OIS database. A synchronization engine synchronizes immunization records across the OIS database and a regional health repository. The OIS system components and interactions within the trust boundary are inherently trusted as they are physically secured and are expected to be unaffected by attacks.

Table 4 presents details on the primary components of a Data Flow Diagram (DFD), i.e., external entities, processes, data stores, and data flows, and the associated notations used to create the system models for threat modeling tools. We created two DFDs for the OIS using the built-in threat model generation feature of Microsoft TMT and OWASP TD – the two tools selected for this study. Fig. 4 shows the DFD generated by Microsoft TMT.

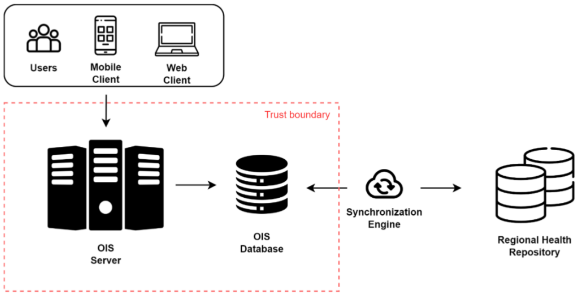


Fig. 3. Architecture of the online immunization system (OIS).

Table 4. Primary components of a DFD and a brief description of their notations

DFD Component	Notation Description
External Entity	*External entity (rectangle)*: An external system or actor that communicates with the system by sending and/or receiving data. Examples are people or code outside of the system's control, e.g., end users and external software.
Process	*Process (circle)*: Any function that modifies input data to produce an output. Examples include any running code such as read/write operations and logical computations.
Data Store	*Data store (two parallel lines)*: Any form of repository or file that can preserve information for later use. Examples of data stores are databases and forms.
Data flow →	*Data flow (arrow)*: The path through which data moves between external entities, processes, and data stores, typically represented as a unidirectional arrow. The bidirectional flow of data is represented with two separate unidirectional arrows.

C PARTICIPANTS' DEMOGRAPHICS

Table 5 presents the demographic information collected using our post-study online questionnaire, including participants' sex, age group, job title, education level and background, self-reported security expertise on a 10-point scale (1: Novice – 10: Proficient), and familiarity with the two threat modeling tools on a 6-point scale (0: Unfamiliar – 5: Expert).

Table 5. Participants' Demographics

ID	Sex	Age Group	Job Title	Education Level	Educational Background	Security Expertise	Familiarity with Microsoft TMT	Familiarity with OWASP TD
P1	Male	18 - 25	Software Engineer	Bachelor's degree	Cybersecurity	8	No data	No data
P2	Female	18 - 25	Security Expert	Master's degree	Cybersecurity	8	No data	No data
P3	Male	18 - 25	Software Engineer	Master's degree	Security Engineering	8	No data	No data
P4	Male	26 - 35	Software Engineer	Bachelor's degree	Computer Science	7	No data	No data
P5	Male	18 - 25	Software Engineer	Master's degree	Computer and Software Engineering	8	No data	No data
P6	Male	18 - 25	System Architect	Master's degree	Systems Engineering	7	2	3
P7	Male	18 - 25	Software Engineer	Bachelor's degree	Software Engineering	7	0	2
P8	Female	26 - 35	Security Expert	Ph.D. or higher	Computer systems engineering	7	3	0
P9	Male	18 - 25	Software Engineer	Bachelor's degree	Software Engineering	4	0	2
P10	Male	26 - 35	System Designer	Master's degree	Electrical and Computer Engineering	7	2	0
P11	Female	26 - 35	Security Expert	Master's degree	Criminology	7	3	2
P12	Female	18 - 25	Security Expert	Bachelor's degree	Computer Science	8	4	0
P13	Female	36 - 45	Software Engineer	Ph.D. or higher	Computer Engineering	6	3	0
P14	Female	26 - 35	Software Developer	Ph.D. or higher	Computer Science	5	1	0
P15	Male	18 - 25	Software Developer	Bachelor's degree	Electrical and Electronic Engineering	5	0	1

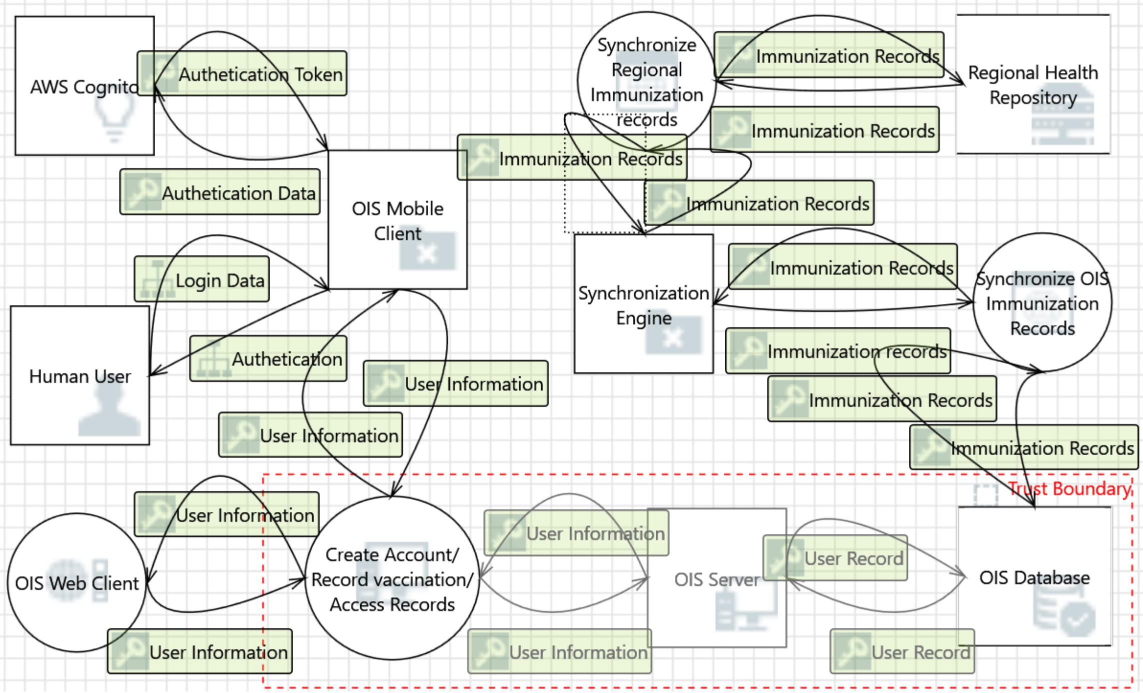


Fig. 4. Data flow diagram of the OIS created in Microsoft TMT.

Received 31 May 2024; revised 22 July 2024; accepted 6 August 2024