Tools Make Me Snore: A Next-Gen Framework for Training SOC Analysts Non-Perishable Skills

Francis Hahn, Spencer Cherry, Kumar Shashwat, Laura Buldrini, Daniel Lende, Xinming Ou
University of South Florida, Tampa, FL, USA
Emails: {fhahn, sferrini, kshashwat, lauraaraujobuldrini, dlende, xou}@usf.edu

***Abstract*—While the work force for the field of cybersecurity grows, the supply of trained and experienced individuals lags behind the demand. This issue coupled with a lack of emphasis on secure software design has led to a growth in opportunity for adversarial actors as evidenced by the consistent occurrence of headline-making cyber threat incidences such as data breaches and supply chain attacks. This paper describes the rationale behind a research effort to discover and improve the quality and efficiency of cyber training pedagogies. The development and testing of these pedagogies was guided by initial discussions with practitioners who work in a SOC (Security Operations Center) and had different levels of work experience and responsibilities. These discussions indicated that both critical thinking and technical skills matter to being successful within a SOC. Technical skills were viewed as "perishable", given how security tools and specific types of attack change over time and how companies use different systems and proprietary programs. Critical thinking skills, in comparison, are viewed as "non-perishable" since they persist despite the changing threat and technology landscape. In the subsequent development of our Mock SOC training scenarios for students, we focus on how critical thinking matters for successfully analyzing and mitigating threats. We perform a case study review of real-world cyber threat incidents to design, build, and collect synthetic incident and attack data. We identify and eliminate where tool-based analysis is needed, thus reducing the need to draw on perishable knowledge during the Mock SOC investigation. Our training scenarios thus emphasize critical thinking in how to analyze and address security breaches. Research on this scenario-based training blends computer science and anthropology expertise to better understand how particular scenarios engage students and how students problem solve within a scenario. We use grounded theory to analyze the scenario data and to refine our hypotheses for what works and what doesn't through multiple rounds of scenario-based training. Based on these results, we are designing a framework for building scenario-based training modules based on accumulated insights into what is and what is not effective for developing non-perishable critical analysis skills. The overall aim is to be able to train students for industry positions by providing them critical skills that are useful in any given organization's technology stack. This paper details how we have designed our framework and used it to conduct human-subject research on building effective scenario-based trainings utilizing the concept of a Mock SOC. We discuss preliminary findings behind our initial training sessions using the scenarios designed based on this framework.**

Workshop on Security Operation Center Operations and Construction (WOSOC) 2025
24 February 2025, San Diego, CA, USA
ISBN 979-8-9919276-0-4
https://dx.doi.org/10.14722/wosoc.2025.23015
www.ndss-symposium.org

I. Introduction

In this work we design a framework for building scenario-based training used to facilitate researching our hypotheses on how to build effective scenario-based trainings for teaching non-perishable skills. The design comes from the use of concepts from computer science and methods from anthropology. Our scenario building is based on discussions with SOC team members and a comprehensive case-study review of real-world attacks to build a near authentic experience and interactive environment for our training subjects. The use of interactive training in cybersecurity has been discussed [7] as an effective model for improving the practical proficiency of trainees, as it provides them experience and insight on how to apply theoretical knowledge. Our research focuses specifically on how student participants use and apply critical analysis skills in scenarios designed to challenge their analytic thinking. To gather data, we run three training sessions using a single scenario, with unique trainees in each session. Our goal throughout the data analysis is to begin identifying themes that match or go against our initial hypotheses, for the sake of identifying "what does and does not work in a cybersecurity themed scenario-based training." Our first goal was "Can we build a condensed training scenario which includes a short primer and a Mock SOC investigation for facilitating learning real-world skills in a hands-on capacity?" We will show in this work that the preliminary findings indicate that such a scenario-based training can be built. In addition to our goal, we look to investigate some of the following hypotheses:

- Will the primer be sufficient for giving the trainees a base level of understanding?
- Can we strip the investigation of the need for specific cybersecurity tools and still build a meaningful scenario?
- Can a Mock SOC training scenario teach the trainees the critical thinking skills required for comprehensive investigations that is often gained from on-the-job training?

The analysis of the data, the recognition of themes, and feedback received from the trainees help guide the direction of future session design decisions, allow us to form new hypotheses, and allow us to begin forming data-driven opinions on our initial hypotheses.

II. Motivation

There is a unique challenge in designing a scenario-based training that blends the aspects of human processes and technical processes without the use of any industry specific tools or, as we call it, perishable knowledge. To accomplish this we need a thorough understanding of SOC procedures and the primary building blocks [16] of a SOC to best build a rich experience for the trainees. Our discussions with industry professionals indicated that more senior SOC analysts tend to have the knowledge and experience to think critically and creatively throughout their investigation to uncover root causes and suggest appropriate remediations. In contrast, entry-level SOC analysts are often challenged by the demands of an investigation; they have to "drink from the firehose" to deal with the high-pressure environment when incidents happen. These two observations motivated us to consider approaching the problem from the perspective of wanting to understand where are the short-comings in a SOC analyst's training that can promote or inhibit this continual growth once entering this position. To do this we decided on the Mock SOC. The Mock SOC aims to use real-world data to model the events of actual cyber-incidents for the trainees to respond to. In developing a Mock SOC we can gain insight into the issues an entry level SOC analyst experiences by simulating the events of an actual SOC environment through the careful orchestration of information flow and incident design. In providing training this way, we as researchers could witness first hand their actions, struggles, and growth. Our design philosophy borrows from various discussions on topics such as inductive learning [13] and simulation-based learning [3] for laying out our Mock SOC scenario and giving the trainees a reason to find out, on their own, the need for learning the material presented in the primer.

III. Scenario Design

Our combination of computer science and anthropology focuses on collecting data that is not easily obtained through conventional means such as questionnaires, surveys, and interviews. The data we aim to collect is that which is uncovered in the moment through interactive observation. We designed this training for participants preparing to enter the workforce, just starting their careers, or who are looking to gain additional experiences, i.e., those working in entry level positions within a SOC or students approaching graduation on the cusp of their interviews. In our approach to designing an effective scenario-based training we had to consider what it means to learn and what are the outcomes when one is engaged in such institutional training programs. We had to consider, when in an environment designed to teach specific skills or concepts what is directly learned and what is indirectly learned [2]. In our scenario we consider what is directly learned to be explicitly known skills which are documented and externalized by experts or educators, what is indirectly learned are the skills or knowledge gained when one internalizes and discusses what they've learned directly [12]. This led us to two questions: what skills do not transfer from one organization to the next and what amount of directly learned knowledge is necessary to induce the formation of tacit knowledge. From these considerations we formed the concept of "perishable knowledge" which is used to classify knowledge which does not persist from one environment to the next. Next we looked to design our scenario-based training using three primary data sources: the expertise of SOC practitioners, case studies, and the curriculum for a university cybersecurity program. From our discussions with practitioners we identified concepts which help the more senior SOC analysts, those who have made a career out of this line of work, perform more effectively as a skill set or a set of experiences we wanted the trainees to walk away with. We then attempted to identify core issues entry-level SOC analysts were faced with: repetitious tasks, reliance on playbooks, and a high pressure to learn job critical procedures on-the-fly. We then looked at our own university's curriculum to identify what could be used as an effective medium for deploying our scenario-based training and we found that Microsoft's Windows Active Directory service would work well for this as it has a high volume of use among organizations, the curriculum's saturation of taught material for it is low and it is rich in security design flaws for us to work with. With the aforementioned in mind we began to explore case studies of incidents related to the Active Directory service and decided on the Golden Ticket attack. This attack provided us with numerous opportunities for indicators of compromise to appear in the data provided to the trainees for investigation. After setting up the network for attack and enabling the necessary logging, we set up our scenario-based attack in a way near to what we had found from our case-study review of other similar attacks, the high-level attack chain can be seen in Figure 1. This approach allowed for the logs to contain both benign and attack data for trainees to think upon during their investigations. We performed various forms of processing on the data to alleviate the need for the trainees to know or use any industry specific SIEM or analysis tools. The scenario-based training is designed to induce a sense of the pressures or the demands of the SOC environment itself, but without the fear or pressures to perform while in a typical job setting. The low-stress simulation of a work environment aimed to facilitate the trainees ability to think more freely or creatively in their investigations to stimulate learning. A primer was included at the start of each session to ensure all trainees had a clear base line of understanding for us to make assumptions on and build hypotheses around. The primer was used to address some central issues in our training with regards to the diversity of participants with variable amounts of experience, training, and a priori knowledge. The primer allowed us to put at the forefront of the trainees memory concepts which would appear in the investigation allowing them to make sufficient progress without "starting from scratch" in a relatively short time period to work. The content of the primer was determined by the methods, technologies, and attack procedures present in what we call, tooling free investigation data, which were the result of our pre-processing efforts on all of the logs and data obtained from building and attacking the network and after determining where perishable knowledge would appear in a typical investigation. The primer is followed by the actual investigation event, where trainees are given a service ticket, a network diagram, and an employee inventory with specific details pertaining to the mock company under attack, the purpose of these given items are to act as bread crumbs to push the trainees in a direction on what data to begin asking for to start their investigation. The approaches the trainees take and data they request are critical to determine what has occurred in the attack and what we aim to collect in this research. These discoveries are reliant on the trainees' ability to critically think and synthesize a path forward based on their a priori experiences and knowledge of what they've gained from the primer.

A. Perishable Knowledge

We knew the scenario had to be a short one-shot event unlike the table-top exercises we drew inspiration from [1]. To this end we took our own experience with security practices and our discussions with the community of SOC practitioners and recognized a lot of time spent during an investigation relates to the tooling. This led us to begin analyzing and preprocessing the data to alleviate the need to use any tools. This includes using python scripts to process the Windows event logs and extracting the necessary data pertaining to date/time, eventID, user name or SID, and domain, in the case of an executable event we provided the execution path from the log as well. We also provided the trainees with raw slack logs and injected into it a url which would appear as malicious when ran through VirusTotal or other similar analysis websites.

B. Synthetic Attack Data

By building real environments and performing real attacks, we're able to synthesize organizational data which mimics some of what a SOC would see. Due to the limitations of getting real attack data which portrays the threat we decided to build our Mock SOC around, we went with the approach of building our own network using a collection of VMs running on a local server hosting VMware ESXI. We implemented the VMs in a way that modeled a small organizational environment with two domain controllers and three workstations for modeling employee use. For the domain controllers we used Microsoft Windows Server 2018 and for the employee workstations Microsoft Windows 10 Pro. To perform the attack we used a somewhat mock approach in itself to allow for the indicators of compromise. From the initial point of compromise we used powershell scripts to disable various services on the machines including real-time detection, Windows Defender Firewall we used the powershell scripts to install the chrome browser, and download an attacker's github repository which contained a copy of the mimikatz tool encrypted to avoid signature detection. The shell scripts were also used to download the various .dll files needed to run the attacker's encryption program. Mimikatz [4] is a hacking tool built for Window's Active Directory security interrogation, the tool is written in the C programming language and exploits issues with Microsoft's implementation of the Kerberos authentication protocol. Mimikatz was used on a workstation and on the domain controller to login and access a remote file share hosted on another workstation with access control policies in-place to prevent access from users other than the workstation owner account and the admin account. All of the event logs and data generated from these actions are what we call synthetic attack data.

C. The Primer

A short presentation of concepts critical to the investigation that trainees may or may not be aware prior to this event. For our scenario this included a light introduction to the Kerberos protocol, a high-level overview of Windows Active Directory, Mimikatz, the Golden Ticket Attack, and the Windows Event Log Manager.

D. Tooling Free Investigation Data

A result of recognition of this concept of perishable knowledge led to what we designed and call, Tooling Free Investigation Data. This includes both the data we present to the trainees and what we expect them to request throughout their investigation. This is our processed synthetic attack data which had two categories: relevant and irrelevant. The relevant data was all of the processed event log or other data files which directly contained indicators of compromise relevant to the attack. The irrelevant data looked similar to processed data for machines on the attacked network that were not relevant to the attack and thus only contain benign events or material.

E. Data Collection

The data we are trying to collect in this research are the thought processes of the trainees during their investigations. To collect this type of data we recognized the trainees needed an environment that offered them ample opportunities for making decisions. To facilitate this we designed a story with critical points that were substantiated by the evidence. Every data request attempt made by the trainees gives us insight into their train-of-thought and their thought process that led to the uncovering of the investigation's overall story. By giving them a story to uncover this gives us the opportunity to have conversations with them during their investigation as they will not be engaged in technical tasks but engaged in thought. Through our conversations during this process we are able to discover and collect their thoughts through our own questioning.

IV. EXPERIMENTATION

The study was designed to be a three-to-four-hour event consisting of the primer. Followed by the investigation event where the trainees were given some initial investigation data to begin with and were then left to perform the investigation by requesting more information based on questions that informed them that there were varying possibilities for how a solution could have been formed. We conclude the event by gathering the trainees and the research team for a round-table group discussion to collect their thoughts and ideas regarding the event, inquiring from them various questions about the scenario. Throughout the scenario we collected data from the

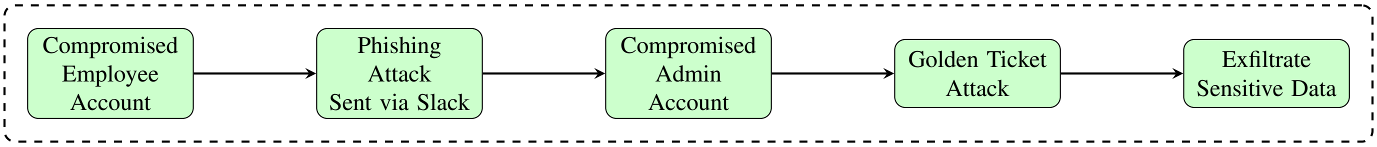

Compromised Employee Account
Phishing Attack Sent via Slack
Compromised Admin Account
Golden Ticket Attack
Exfiltrate Sensitive Data

Fig. 1: High-Level Cyber Incident Flowchart

trainees for their thoughts and rationale behind the actions they take in their investigation. The research team performs this function by casually walking around and having dialogue with the trainees asking them how their investigation is going and why they're currently pursuing specific lines of investigation. The purpose of the field note process is to uncover, in the spur of the moment, a snapshot of their thought processes. The research team also asks questions involving feedback, asking their opinion of the scenario itself. This field note approach is to gain a holistic understanding of the scenario from the perspective of the trainees who are our models or windows into understanding the effectiveness of such a scenario-based education approach.

A. Recruitment

The recruitment was performed by mass-email campaign targeted at our University's student body across the colleges of Engineering, Arts and Sciences, and Business. The students from these colleges, after filling out the necessary consent information, received a survey which asked them questions regarding what degree they were seeking, the current progress of their degree, what courses they have taken, if they have ever had industry experience, and what relevant extra-curricular activities they participate in. The aim of this is to allow an adequate pool of applicants with the necessary background skills which models the target skill set of entry-level SOC analyst. From this campaign we were able to recruit 16 students across 3 groups, from diverse backgrounds including Business and Cyber, all aiming for careers in cybersecurity. The demographics from our recruitment campaign is shown in Figure 2.

B. The Agenda

The Mock SOC investigation is the orchestration of events which is detailed in an agenda shown in Figure 3.

C. The Scenario

We start the scenario with providing the trainees with three articles of information: a service ticket, a network diagram, and an employee inventory and policy map. The service ticket details the events which led to the need for an investigation, what is being made known to the trainees, and what their tasks are. The network diagram describes the company's enterprise network setup; it shows which computers are domain controllers, which computers are workstations, an example of who should have access to the machines. The employee inventory and policy map lists all of the employees who are with the company, the department they work in, the machine they have access to, and access control policies. These three starting documents are supplied to act as breadcrumbs for the trainees to begin their investigation, no other documents are supplied and for them to receive more they must request more through critical thinking about what they know and what they've been given. When a data request is made we inform the trainee about whether the document is available or not and what their reason for requesting the document is. Throughout the scenario the trainees are asked questions by the research team on their current progress, thoughts on the scenario, the reasons and rationale on their current investigation approaches. Once the investigation ends we ask the trainees to produce an incident report detailing their findings with an emphasis on their thought process for all things investigated, this is to help us measure how they performed, measure the effectiveness of the primer, and measure critical points within the scenario which were and were not effective. After all of the reports are collected we provide the trainings with a possible solution, when we say possible we mean that given the agency of being able to request data there are many solutions depending on how one traverses the data. The solution shows the trainees one path of data requests and we go through all of the relevant documents observing where the indicators of compromise appear.

D. Round-Table Discussion

We conclude the event with a round-table discussion, the purpose of this is to provide the research team and the trainees an open forum for everyone to discuss the event. The research team aimed to gather specific information in the moment by asking questions such as the trainees biggest take-away, what they felt like they could have done better, how the approached the investigation, and if they could see themselves doing this kind of work full-time. The trainees are encouraged to respond however they like and as candidly as possible.

V. Framework

Given this work will span across three years, we wanted to begin our scenario-based training design in a way that set us up with a uniform methodology which allowed us to systematically reflect and refine upon as our research progressed. To allow for this process of iterative improvement, we designed a framework for building a scenario-based training. Here we will discuss our high-level design methodology followed by our rationale on key components of the design. Our design starts with a problem of interest and working environment, this lays the ground work for designing what we aim to gain or impart upon the trainees when performing the scenario-based training. With an idea for what our

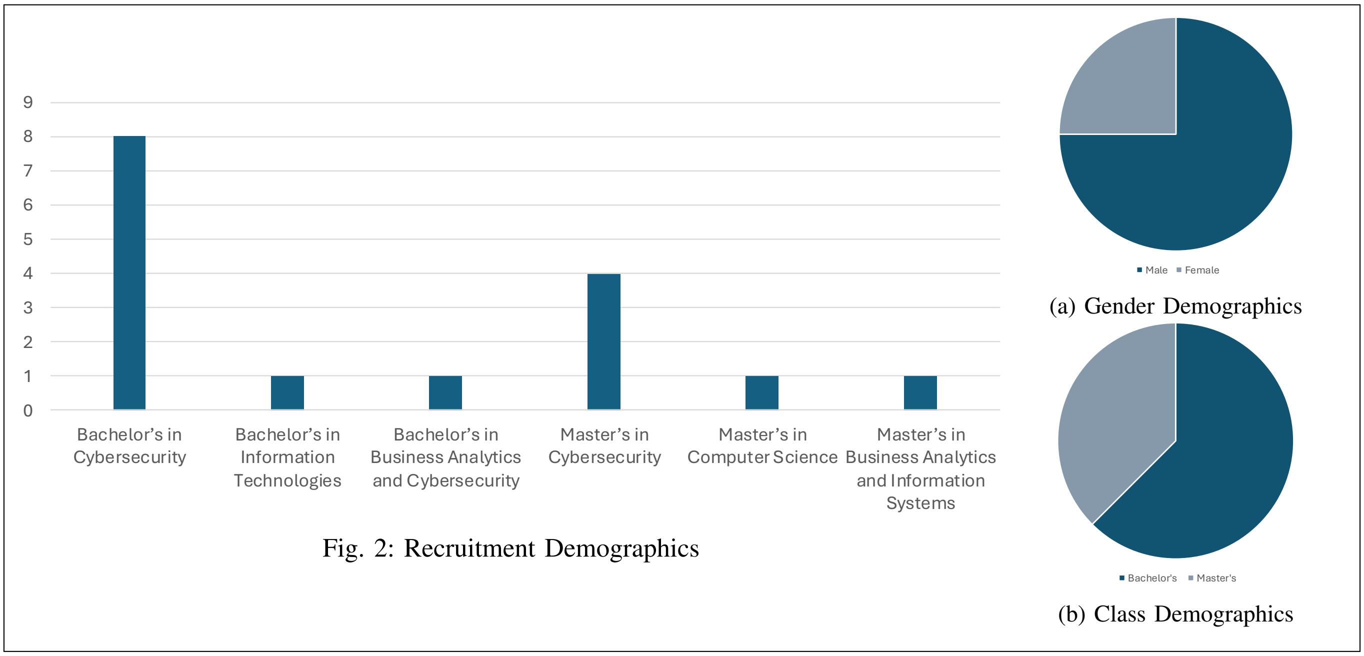
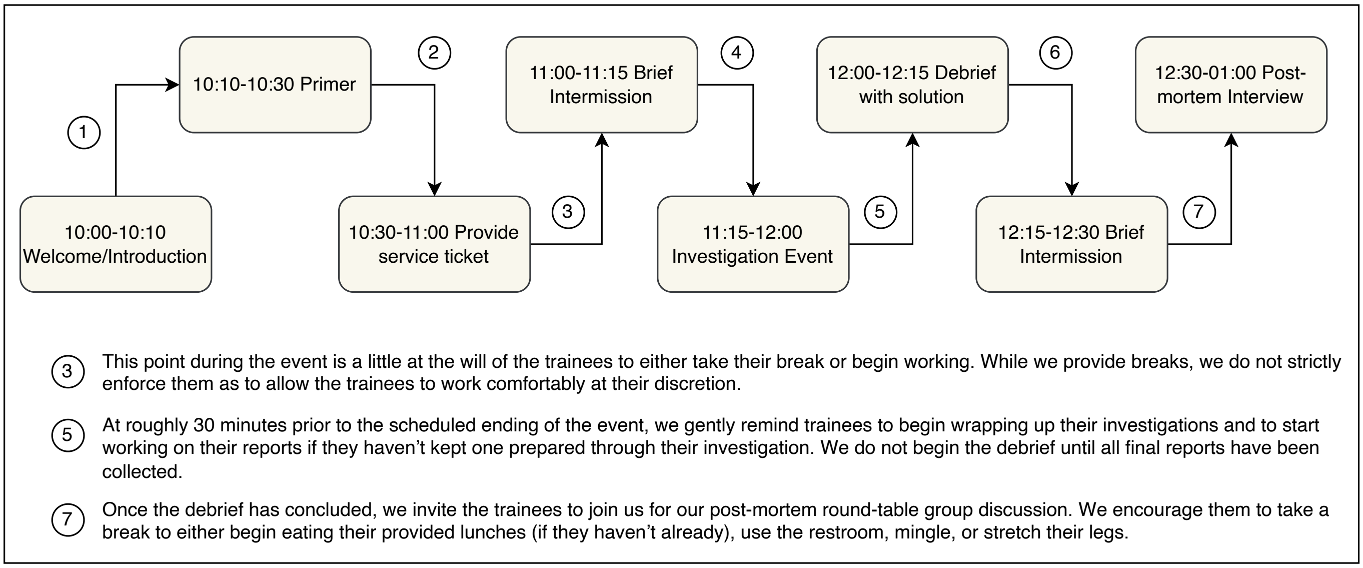


Fig. 2: Recruitment Demographics

(a) Gender Demographics

(b) Class Demographics

Fig. 3: Overview of Recruitment, Gender, and Class Demographics.


10:00-10:10 Welcome/Introduction
1
10:10-10:30 Primer
2
10:30-11:00 Provide service ticket
3
11:00-11:15 Brief Intermission
4
11:15-12:00 Investigation Event
5
12:00-12:15 Debrief with solution
6
12:15-12:30 Brief Intermission
7
12:30-01:00 Post-mortem Interview

3 This point during the event is a little at the will of the trainees to either take their break or begin working. While we provide breaks, we do not strictly enforce them as to allow the trainees to work comfortably at their discretion.

5 At roughly 30 minutes prior to the scheduled ending of the event, we gently remind trainees to begin wrapping up their investigations and to start working on their reports if they haven't kept one prepared through their investigation. We do not begin the debrief until all final reports have been collected.

7 Once the debrief has concluded, we invite the trainees to join us for our post-mortem round-table group discussion. We encourage them to take a break to either begin eating their provided lunches (if they haven't already), use the restroom, mingle, or stretch their legs.

Fig. 4: Agenda Flowchart

working environment will look like and what we're trying to accomplish with the scenario-based training we want to design the story. The story is in essence the vehicle which drives everything. The story dictates what our IoCs should look like the types of technologies we plan to setup, the attacks we intend to implement, and how we will introduce the scenario to the trainees. Once we have our story in place, we move on to building the environment and implementing the attacks. Having the environment configured correctly (or incorrectly) will enable us to generate and collect the data necessary when performing the attacks. This allows us to decide on both the location of the IoCs and where perishable knowledge exists when parsing the data. Knowing the IoCs and where our perishable knowledge exists allows for the creation of the tooling-free investigation data. The tooling-free investigation data should be curated in a way that includes a sufficient amount of IoCs with the understanding that the trainees may not find all but should be able to find at least one. Thus this requires careful consideration on the balance of malicious and benign data present in any one data file. Once all of the tooling-free investigation data is collected we then decide on how we want to prepare the primer, this is a critical point where careful consideration must be made to include the right amount of information that appropriately exposes your trainees to the concepts so that they have a high-level understanding but not going into too much depth to prevent any one concept dominating the rest. Finally you can conclude the design with setting up your agenda in a way that allows for organizing your event that starts with the primer and leads to the investigation. Given sufficient training in the technologies and procedures used within a scenario's design, using this framework we found we can reduce the time to design a scenario-based training from a 2 to 3 month timespan to a matter of weeks.

A. The Environment, Problem Statement, and Hypothesis

The design of the scenario-based training started with our problem of interest or target environment at the core, for us it was the SOC. Once we knew we wanted to work in an environment which modeled something a SOC analyst would investigate we then decided what our target take-away from the training is going to be or what it is about the environment we're trying to observe, for us it is the idea of burnout and experience deficits. Once laying out the fundamental concepts we identified how to implant our hypotheses into the training, in our case this was our idea of perishable knowledge being an inhibiting factor to critical thinking for entry-level analysts and the benefits of a scenario-based training.

B. The Story

For our problem, once we knew who it was were were trying to train and what it was we were modeling, we were able to build a story. This story acted as the guidelines or our map towards network design, IoC usage, and tooling requirements. The story building was a critical-point for us, it led us towards which case-studies to review and what information from our practitioner discussions could be used to refine our scenario-based training towards being as authentic to the real-world as possible.

C. Environment Design

Having our problem statement, hypotheses, and story all set, the next steps were to build and setup a network which was to act as an environment that resembles that which is portrayed in our story. This was important as when we wrote our service ticket it largely took from our story, thus it was critical to have an environment that closely resembles the story as we did not want to confuse our trainees with fabricated inconsistencies during the proceedings.

D. Attack performance, Data Collection, and Locating your IoCs

Once we had our network setup we began our attack procedures to collect as much data relevant to the attack as possible, for us this was performing the necessary steps to enact a golden ticket attack on our Windows Active Directory based network. Once we collected sufficient data, via the Window's Event Log system, of the attack we then began to analyze for the location of the IoCs. In our case these were erroneous login times, execution of suspicious programs, events pertaining to suspicious ticket delegations, and slack logs containing a phishing link.

E. Identifying your Perishable Knowledge and Building your Tooling-Free Investigation Data

While we were analyzing the data for IoCs we were concurrently considering what tools are required to locate the IoCs which are subject to training programs or prerequisite knowledge for effective usage. We used python scripts to parse the event log data, however we recognize in large scale operations tools such as Splunk are used. Since one of our research questions is to determine if the removal of perishable knowledge can still allow for an effective training program, we decided to remove any need for such tooling. At this point is where our tooling-free investigation data is created. Since we've identified where the IoCs are located and what tools are required to extract them, we perform these actions and generate a reasonable quantity of articles for each relevant eventID and slack conversation, while maintaining a reasonable amount of noise which are benign or normal actions. Maintaining this balance was critical to ensure that the IoCs aren't glaringly obvious but aren't impossible to locate throughout the file due to an oversaturation of noise given we were removing the use of tools.

F. Putting It All Together

Once we had our tooling-free investigation data we proceeded to design our primer to only include what was necessary for the trainees to know and to be delivered in 20 to 30 minutes. This necessity based design was to ensure that the trainees all had a base level of understanding by uniformly exposing them to the same material. We then decided on what ”bread crumbs” or initial investigation data to provide the trainees at the start of the event. How we considered what would be effective was to begin with the details of our service ticket. The service ticket included the initial signal to the SOC that an event needing investigation occurred and data informing the trainee of who, within the mock organization, is suspected to be involved. Because this story includes various actors, both relevant and irrelevant, we believed an employee list which included the personnel's permissions and roles within the organization would give the trainees a sufficient amount of information to begin forming questions of who could be involved. We also decided to provide the trainees with a map of the organizations network design so that the trainees could better visualize the information portrayed in both the employee inventory and service ticket. The primer, service ticket, employee inventory, and network map are designed to give the trainees just enough information to begin forming questions that would lead to data requests allowing them to uncover the story's unknown factors in why the attack was possible, how it unfolded, what the attacker accomplished, and provide remediation recommendations.

VI. Preliminary Findings

Data from the training scenario comes from multiple sources – an initial demographic survey, interactions with participants during the scenario, the requests they make for logs and other information during the scenario, their end reports describing how they tackled the scenario and discovered problems, and a round-table discussion at the end with all study participants and research team members. Research team members also wrote up field notes after the scenario which described their interactions with participants, including what participants were working on, their verbal descriptions about what they were doing and why, and how they used things to problem solve (these included using the provided handouts, hand written notes covering the primer and their own work, documents on their laptops to track relevant information, and using different data tools that ranged from Ctrl-F, google searches, AI summaries on particular topics, to checking suspicious links on sites like VirusTotal). The analysis of data happens in an iterative process – initial ideas developing during writing individual team member field notes post-scenario, collation of data from various sources around specific participants (e.g., the survey, the data requests, the end of investigation reports, and the participants' differing interactions with team members over the scenario), and research team discussions around patterns observed, emerging themes in the analysis, and metrics around how participants did during the scenario.

A. Overview

The following describes an overall reporting of the preliminary findings generated from our initial analysis of the data retrieved from the first three sessions. Key observations about the trainees include:

- They experienced challenges understanding where to get started.
- They were favorable to the challenge of not being provided data up front.
- There was no difference in "success" in a collaborative vs solo approach. We also noted that individuals that collaborated often had the similar responses to our questions in the round-table sessions.
- Making connections was important to problem solve.

The following three subsections give more detailed information about each session. These reportings are formatted to show the overall performance of the trainees and is meant to show some individual or differential observations found between each cohort.

1) *Session 1:*

- 3 of 5 or 60% of participants showed a level of full scope understanding of scenario.
- 2 of 5 or 40% of participants performed all required tasks as directed.
- Most trainees started with slack logs and moved to event logs afterwards.
- Trainees in this session had a handful of self-imposed limitations that were perceived from their training environment. We observed and recorded them mentioning that they were in the mindset similar to that of a closed-book examination which caused them to limit the usage of resources available to them.

2) *Session 2:*

- 3 of 6 or 50% of the participants showed a level of full scope understanding of scenario.
- 3 of 6 or 50% of the participants performed all required tasks as directed.a
- Most trainees focused on time-of-incident.
- Trainees that did not complete only found the phishing link.
- Some trainees used various 3rd party tools (i.e. VirusTotal).

3) *Session 3:*

- 1 of 5 or 20% of participants showed a level of full scope understanding of scenario.
- 2 of 5 or 40% of the participants performed all required tasks as directed.
- Most trainees focused on the timeline of the incident.a
- Everyone started at computer logs, but needed to be directed towards the Slack logs.

B. What the Trainees Had to Say...

The following lists some of the quotes we recorded (by hand) during our various interactions with the trainees during the investigation and round-table discussions. These quotes are in a sense a way to gauge how the trainees reasoned through their investigation and what they thought about the scenario. These quotes are used as the window into the thinking rationale of the trainees and are a critical way for us to interpret the effectiveness of the scenario-based training and to learn how to improve.

1) ***Trainee A:*** We observed during various points in the investigation trainees deeply engaged in problem solving, this can be seen from the dialogue with trainee A,

> "What was in the slack message?"

trainee A was then observed questioning themselves on their initial inquiry with questions such as

> "Why did he change the password?"
>
> "What made him do that?"
>
> "Why didn't he report that?"

2) ***Trainee B:*** Another trainee, trainee B, was engaged in dialogue during their investigation of the Windows Active Directory Event Logs. Their initial question was

> "This Tom Jones guy, what is he doing?"

trainee B was observed trying to reason around their core question of

> Where to start?

trainee B began thinking deeper and more critically by asking,

> "What did he click on?"
>
> "What was the source of the initial compromise?"

this line of thinking through the dialogue was able to capture the trainee then, through their own questioning, pivoting to the Slack logs to further their investigation.

*3) **Trainee C:*** During our round-table discussion, when asked what the trainees thought of the scenario, one of the trainees responded with a quote that resonated with the group.

> "Though I have worked within a SoC, this pushed me to think outside the box"

This was an observation that gave us an indication that our scenario-based training was, in some aspects, successful in engaging the trainees to think critically. Signaling that our scenario promotes critical thinking by providing an environment for them to think freely and creatively to solve the investigation. Which from previous research is a critical factor to what inhibits the growth of an entry-level SOC analyst. This conversation with our trainee also gives us a perspective that we can begin using to relate our training with that of an organizational SOC environment and begin making comparisons.

VII. Related Work

All prior works reviewed discuss a lack of research [8] in the area of SOCs. While the issues the SOC faces are many, we aim to approach what we see as a critical issue, that being the people. A comprehensive review of the work available shows the problem or pain points for the human-factor [11] of the SOC. Prior work [14] investigated the idea of SOC burnout, based on the "vicious cycle" of the work. Drawing from economics, they proposed a solution where analysts were considered "Human Capital" who need training and support to effectively do their jobs. Our work builds upon this by taking the human capital model and applying our scenario-based training to improve the skills of trainees by putting them in an environment where they must think creatively in a low-stress and low-risk environment and regardless of their performance they walk away with insight and knowledge not previously known. Another work [15] investigated the use of computer science students trained in anthropological methods to observe and contribute to a software development team with the intent to create a culture focused on secure coding practices. Our work builds upon this approach by observing the roles of SOC analysts of various career stages, through discussion sessions to better understand the culture which appears to be unchanged in over a decade. This allows us to build near authentic scenario-based training modules which can be used by others to administer or build upon for their own use. However, this work is not possible without performing the methods found in research [9] involving the co-creation of solutions alongside the analysts, the managers, the researchers and the trainees who enable this work. In the area of human-centered cybersecurity there has been an observed research-practice gap problem [5] [6] between practitioners and researchers, this has been seen to reduce the effectiveness of research by a two-fold issue of a lack of interest by practioners and an lack of understanding of when to interact by the researchers. Our work aims to document an approach to overcoming this problem through continuous engagement to stay informed, allowing us to provide authentic experiences in our scenario-based training and by providing usable documented training modules for SOC practitioners that are guided by practitioners. In our work we consider the results or outputs of software-based tools as data [10]. Because we consider tooling skill sets to be perishable knowledge we separate the user, our trainee, from the need to produce the results (data) themselves. Thus we alleviate the need to use the tools during an investigation and this allows us to provide our trainees a tooling-free investigation environment. Our work uses this method to enable the teaching of non-perishable skills observed in senior SOC analyst that are discussed to be lacking in the more junior SOC analysts. A SOC analyst armed with experience, technical skill, and the ability to critically think empowers them with the capability to think freely and objectively to generate meaningful paths for incident investigation.

VIII. Ethics and Considerations

This human-subjects research was approved by the university IRB where the research was conducted, and all participants provided informed consent. As part of informed consent, participants received information on the benefits and risks of participating in the research before deciding to voluntarily participate in the research. Benefits included learning more about cybersecurity. Risks included potential violation of confidentiality and possible embarrassment, due to the nature of the training study where multiple subjects participate in one session. Both risks were considered low. During the data collection, participants were reminded that their participation was voluntary, and they were not required to answer questions and could exit the study without consequence. All data collected was handled in a secure manner and stored using a password protected repository where all researchers had enforced multi-factor authentication in-place. Anonymization was performed during data processing and maintained during write-up and publication.

IX. Conclusion and Future Work

In this work we've provided a framework for building a scenario-based training around the concept of a Mock SOC. To design our scenario-based training and framework we've engaged a substantial amount practitioners in discussion sessions to gain insightful information on how to build an authentic training environment for the trainees. To test the efficacy of our scenario-based training we've recruited numerous skillful and capable students trained in cybersecurity with some form of extra-curricular or work experience to participate in our scenario-based trainings. To gain insight into the trainees thought processes to better inform us on how to further improve our scenario-based trainings and framework, we spoke with the trainees during the training and in a round-table discussion at the end of each scenario training session. Participants indicated the training sessions challenged them in ways that typical classwork did not, and forced them to problem solve on their own terms without being given the answer. This finding points towards the efficacy of our scenario-based training from our trainees and interest from our practitioner collaborators. To further improve our framework and scenario-based training design we plan to perform a more comprehensive analysis of the data collected from our training sessions. At the time of writing, we have performed two additional scenarios which have yielded data not yet processed by the research team as well as an improved scenario-based training design and an additional attack chain story for the trainees to investigate which is more complex and rich in IoC and advanced cybersecurity concepts. Overall, our approach shows that scenario-based training can effectively engage participants in problem solving and critical thinking in ways that match up with the demands of being a SOC analyst. Our future research will aim to further distill key elements to successful scenario design and examine in-depth how study participants handle the challenges of Mock SOC training.

X. Acknowledgements

Raj Rajagopalan and John McHugh provided invaluable suggestions in the formulation of the Mock SOC training ideas. We thank Rapid7 SOC practitioners for spending time with us providing feedback and expertise that helped us design and refine the mock-up scenarios used in the research. We thank our many volunteer trainees for their time and effort to make this research possible. This work is supported by the National Science Foundation under award no. 2235102, and Office of Naval Research under award no. N00014-23-1-2538. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these agencies.

References

[1] Giddeon N Angafor, Iryna Yevseyeva, and Ying He. Game-based learning: A review of tabletop exercises for cybersecurity incident response training. *Security and privacy*, 3(6):e126, 2020.

[2] Susan D Blum. Why don't anthropologists care about learning (or education or school)? an immodest proposal for an integrative anthropology of learning whose time has finally come. *American anthropologist*, 121(3):641–654, 2019.

[3] Olga Chernikova, Nicole Heitzmann, Matthias Stadler, Doris Holzberger, Tina Seidel, and Frank Fischer. Simulation-based learning in higher education: A meta-analysis. *Review of educational research*, 90(4):499–541, 2020.

[4] Alva 'Skip' Duckwall and Benjamin Delpy. Abusing microsoft kerberos: Sorry You Guys Don't Get It. Black Hat USA 2014 Briefings, August 2014. https://www.blackhat.com/docs/us-14/materials/us-14-Duckwall-Abusing-Microsoft-Kerberos-Sorry-You-Guys-Don%27t-Get-It.pdf.

[5] Julie Haney, Clyburn Cunningham, and Susanne Furman. Towards bridging the research-practice gap: Understanding researcher-practitioner interactions and challenges in human-centered cybersecurity. *practice*, 24:41, 2024.

[6] Julie Haney, Clyburn Cunningham, and Susanne M Furman. Towards integrating human-centered cybersecurity research into practice: A practitioner survey. In *Symposium on Usable Security and Privacy (USEC)*, 2024.

[7] Marcus Knüpfer, Tore Bierwirth, Lars Stiemert, Matthias Schopp, Sebastian Seeber, Daniela Pöhn, and Peter Hillmann. Cyber taxi: A taxonomy of interactive cyber training and education systems. In George Hatzivasilis and Sotiris Ioannidis, editors, *Model-driven Simulation and Training Environments for Cybersecurity*, pages 3–21, Cham, 2020. Springer International Publishing.

[8] Faris Bugra Kokulu, Ananta Soneji, Tiffany Bao, Yan Shoshitaishvili, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. Matched and mismatched socs: A qualitative study on security operations center issues. In *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, pages 1955–1970, 2019.

[9] Daniel Lende, Alexis Monkhouse, Jay Ligatti, and Xinming Ou. Co-creation in secure software development: Applied ethnography and the interface of software and development. *Human Organization*, 82(1):13–24, 2023.

[10] Daniel H Lende. Software and human systems: An academic and applied introduction. *Human Organization*, 82(1):1–12, 2023.

[11] Calvin Nobles. Stress, burnout, and security fatigue in cybersecurity: A human factors problem. *HOLISTICA–Journal of Business and Public Administration*, 13(1):49–72, 2022.

[12] lkujiro Nonaka, Hirotaka Takeuchi, and Katsuhiro Umemoto. A theory of organizational knowledge creation. *International journal of technology Management*, 11(7-8):833–845, 1996.

[13] Michael J. Prince and Richard M. Felder. Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2):123–138, 2006.

[14] Sathya Chandran Sundaramurthy, Alexandru G Bardas, Jacob Case, Xinming Ou, Michael Wesch, John McHugh, and S Raj Rajagopalan. A human capital model for mitigating security analyst burnout. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*, pages 347–359, 2015.

[15] Anwesh Tuladhar, Daniel Lende, Jay Ligatti, and Xinming Ou. An analysis of the role of situated learning in starting a security culture in a software company. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 617–632. USENIX Association, August 2021.

[16] Manfred Vielberth, Fabian Böhm, Ines Fichtinger, and Günther Pernul. Security operations center: A systematic study and open challenges. *IEEE Access*, 8:227756–227779, 2020.