

# Conceptual Model of Visual Analytics for Hands-on Cybersecurity Training

Radek Ošlejšek, Vít Rusňák, Karolína Burská, Valdemar Švábenský, Jan Vykopal, and Jakub Čegan

**Abstract**—Hands-on training is an effective way to practice theoretical cybersecurity concepts and increase participants' skills. In this paper, we discuss the application of visual analytics principles to the design, execution, and evaluation of training sessions. We propose a conceptual model employing visual analytics that supports the sensemaking activities of users involved in various phases of the training life cycle. The model emerged from our long-term experience in designing and organizing diverse hands-on cybersecurity training sessions. It provides a classification of visualizations and can be used as a framework for developing novel visualization tools supporting phases of the training life-cycle. We demonstrate the model application on examples covering two types of cybersecurity training programs.

**Index Terms**—Visual analytics, cybersecurity, hands-on training, classification, education.

## 1 INTRODUCTION

Our society is being exposed to an increasing number of cyber threats and attacks. The lack of a strong cybersecurity workforce presents a critical danger for companies and nations [1]. Hands-on training of new professionals is an effective way to remedy this situation. In our work, we use visual-based sense-making and reasoning to support participants in better and faster comprehension of attacks, threats, and defense strategies.

The ability to use visual-based analytical reasoning is essential in many fields, including biology [2], medicine [3], urbanization [4], and education [5]. The goal of this paper is to create a conceptual framework providing broader insight into the application of visual analytics (VA) principles [6] in hands-on cybersecurity training. Conceptual models like the one proposed in this paper help researchers design effective visual techniques in a given domain. To the best of our knowledge, the current literature for cybersecurity training lacks such a conceptual model.

There are several reasons for the absence of a conceptual model. Existing hands-on cybersecurity training is largely heterogeneous. Training sessions differ in content, organization, target audience, and technical means. Moreover, the cybersecurity domain represents a sensitive area similar to military or intelligence services, in which many sources are secret or restricted. Therefore, it is challenging to become familiar with this domain and clarify the terms and processes. Fortunately, we have the benefit of seven years of experi-

ence with the design and organization of training sessions. The results of this paper arise from close cooperation with domain experts who directly participate in the development and operation of the *KYPO Cyber Range* [7] – a sophisticated platform for cybersecurity training. Their knowledge and the survey of other existing approaches are essential for this work.

The two most widely recognized hands-on cybersecurity training activities are *Capture the Flag* (CTF) and the *Cyber Defense Exercise* (CDX). The main difference lies in their educational goals. While CTFs focus mainly on improving hard skills in the cybersecurity domain, CDXs target both hard and soft skills. CTF features a game-like approach [8]–[11]. Participants gain points for solving technical tasks that exercise their cybersecurity skills. Completing each task yields a text string called *flag*. In contrast, CDXs have been traditionally organized by military and governmental agencies [12] that emphasize realistic training scenarios that authentically mimic the operational environment of a real organization [13]. We deeply analyzed these types of training programs to distill a unified visual analytics model that fits the heterogeneous cyber-training events and is simultaneously instructive for the design of specialized visual analytics tools.

The major contributions of this paper are: (a) a definition of a unified training life cycle with user roles having clear responsibilities and requirements; (b) a proposal for a conceptual model of visual analytics for hands-on cybersecurity training that can be used as a framework for further research and for developing visualizations supporting particular life-cycle tasks; and (c) demonstrations of the applicability of the model using real examples and lessons learned from our long-term experience in designing and organizing hands-on cybersecurity training.

The paper is organized as follows: Section 2 introduces the related work. In Section 3, we discuss the generic life cycle of hands-on cybersecurity training sessions with user roles that delimit requirements put on analytical tasks

- R. Ošlejšek and K. Burská are with the Faculty of Informatics, Masaryk University, Brno, Czech Republic.  
E-mail: {oslejsek, xburska}@fi.muni.cz
- V. Rusňák, and J. Čegan are with the Institute of Computer Science, Masaryk University, Brno, Czech Republic.  
E-mail: {rusnak, cegan}@ics.muni.cz
- V. Švábenský and J. Vykopal are with the Institute of Computer Science and Faculty of Informatics, Masaryk University, Brno, Czech Republic.  
E-mail: {svabensky, vykopal}@ics.muni.cz

Manuscript received August 6, 2019.

and visualizations. Sections 4 and 5 provide classification schemes for data and analytical visualizations. A demonstration of the conceptual model is presented in Section 6. Section 7 summarizes the observations attained during our research. Section 8 outlines the direction for future research topics.

## 2 RELATED WORK

Our work is unique in its close interconnection of three areas: visual analytics, cybersecurity, and education. Publications dealing directly with the intersection of these fields are rare. Therefore, we have explored related work from several relevant points of view.

### 2.1 Visual Analytics in Cybersecurity

Many works have addressed the challenges related to the design or evaluation of cybersecurity tools and techniques [14]–[18]. A visual analytics approach to automated planning attacks has been discussed [19]. All the surveys have confirmed the importance of supporting analytic tasks by visual interfaces. However, they are aimed at the security-related focus only and do not tackle the educational aspect of the training of new experts. We took the challenges into account in our work, and we incorporated specific aspects of hands-on cybersecurity exercises.

### 2.2 Visual Analytics in Education and Training

Another perspective that considers visualizations in relation to cybersecurity emphasizes the educational aspect. There are distinct approaches to enhancing cybersecurity abilities that focus on training or teaching computer security [20]–[22]. However, these works again provide outputs of a narrow scope and often omit any profound conceptualization of their findings.

To help us comprehend the topic more thoroughly, we do not focus exclusively on the cybersecurity field; we also consider studies that relate to education and training from a broader view. A recent survey [23] introduces a literature classification in the field of interactive visualization for education with a focus on evaluation, and it lists common categories of educational visualizations from distinct fields. In this respect, our work is unique as it considers more than the educational theory. It also includes the application of hands-on training with practical and technical aspects that are an essential part of the learning process.

The issue of education has been approached from the opposite direction [24]. In this work, the authors focus on predictive models for teachers of higher education institutions. They confirm the need for insight for both the teachers and the students that exceed simple summative feedback.

### 2.3 Generic Models of Visual Analytics

Many generic design frameworks, models, and methods exist in the literature. These provide a structure and explanation of activities that designers perform when proposing suitable visualization tools [25]–[28]. However, the aim of this paper is not to discuss processes leading to the development of specific visualizations for cybersecurity training.

Instead, we provide a conceptualization of the domain so that our model can serve as a framework for discussion and the efficient application of existing design methods for specific training tasks.

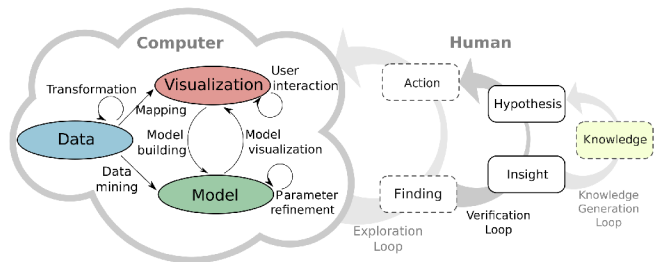


Fig. 1. Altered version of models by Keim [29] and Sacha [30] for insight retrieval based on visual analytics approaches.

Our solution builds upon Keim’s [29] and Sacha’s [30] conceptual models for the visual analytics process. The VA process is characterized by the interaction between data, visualizations, models of the data, and users discovering knowledge, as shown in Fig. 1. Keim emphasizes the computer-driven components of the VA process; Sacha extends the model with human reasoning. *Data* carries facts in structured, semi-structured, or unstructured form. The *model* captures the results of automated analysis methods. The interactive *visualizations* are the primary user interface presenting *data* and *models* in a comprehensible manner. The human-centered part consists of three loops. The *exploration loop* captures low-level visual interactions using actions and findings that are specific for individual visualizations and interests. The analysts then refine their hypotheses in the *verification loop*. The *knowledge generation loop* describes the transition from observations into generalized knowledge.

These two models form the foundations of our work. We utilize *data* and *visualization* components of Keim’s model and narrow our focus on the *verification loop* that plays a crucial role in building knowledge in any domain. The *model* component of the VA process represents the cross-cutting concern, which is out of the scope of this paper. Therefore, we do not provide a separate classification for it. Instead, we mention suitable *models* in our discussion of the classification of *visualizations* and *hypotheses*. The *exploration loop* and *knowledge generation loop* are omitted since they provide either too detailed or too generic concepts.

## 3 CYBERSECURITY TRAINING LIFE CYCLE

The human loops of Sacha’s VA model (see Fig. 1) reflect the needs of users who interact with the computer system. Based on the literature review, our experience, and the application of analytical methods, we distilled the following general life cycle that clarifies *who* is involved in the human loops, *what* they expect (at a high level of abstraction), and *when* they conduct their VA tasks. These pieces of information are later used for the detailed conceptualization of the “computer part” of the VA model by answering *what* (data and hypotheses) and *how* (visualizations) can be analyzed in the cyber training.

### 3.1 Phases

Based on the literature review and our experience, we distilled three generic phases (see Fig. 2) of the cybersecurity training life cycle. We performed a theory-driven qualitative coding method [31] on four key papers [32]–[35] that deal with organizational aspects of cybersecurity training. Using an open coding method helped us to structure the analysis and consolidate observations. Phases and outcomes discussed in the analyzed papers can slightly differ from our model. Nevertheless, the subtleties are rather negligible since the terminology in this domain is yet not established.

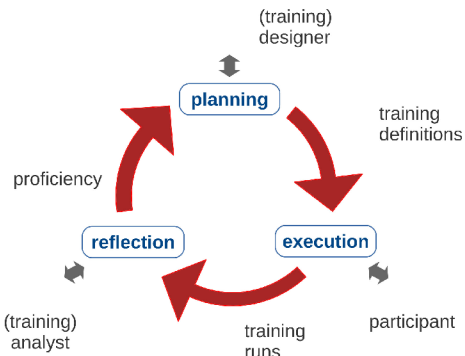


Fig. 2. Cybersecurity training life-cycle phases with corresponding user roles, and main outcomes of each phase.

**Planning** is the first phase of any new training. The goal is to formulate technical and educational requirements, set measurable objectives, and allocate necessary resources. The *training definition* – the main output – is a set of (more or less) formally defined configurations of the computer network and its nodes, specification of attacks, training tasks and objectives, scoring rules, expected skills of participants, and related configuration data of the training.

The **execution** phase represents a training session in which participants are physically involved. User activities and the state of the training infrastructure are monitored, and the data is stored for further analysis. We refer to the data from this phase as *training runs*.

During the **reflection** phase, *training definitions* and *training runs* are analyzed and evaluated. Reflection can be conducted at any time. *Analysts* usually explore the data after each training run to learn from it or provide feedback to involved people. However, they can also analyze the data before or during the planning phase of a new training session to gradually improve its quality. The reflection phase, therefore, helps to increase the *proficiency* in designing and organizing training events.

### 3.2 User Roles

The requirements put on visual analytic interfaces are affected by user roles. The basic roles emerged from the life cycle. They reflect individual phases captured in Fig. 2. For clarity, our roles are CAPITALIZED in the paper.

**TRAINING DESIGNERS** (**DESIGNERS** for short) are responsible for the design of training definitions during the *planning* phase. Multiple designers with different skills are usually involved in the preparation of new training content. Cybersecurity experts contribute primarily to the technical

aspects; education experts are responsible for defining the learning objectives and assessment criteria.

**PARTICIPANTS** represent everyone involved in the training event. Their analytical activities are associated with situational awareness and gaining insight into the training during the *execution* phase.

The **TRAINING ANALYST** (**ANALYST** for short) role covers all the people who conduct the post-training analysis of collected data. In our VA model, this role is used to capture the requirements of generic analytical interactions. Various people interested in the relevant data can take on this role, e.g., cybersecurity experts looking for talented participants.

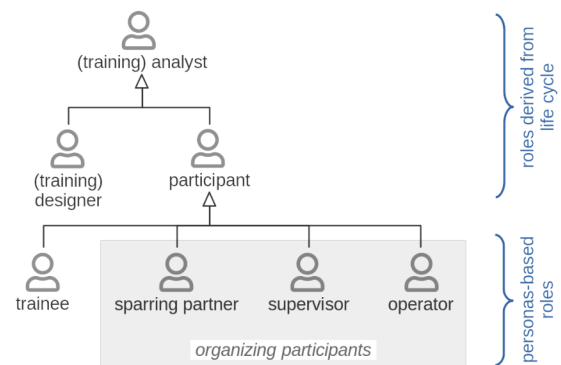


Fig. 3. Hierarchy of user roles participating in cybersecurity training.

These three roles are not independent. Arrows in Fig. 3 represent the inheritance of user roles as defined by requirements analysis methodologies in software engineering [36]. It means that **DESIGNERS** and **PARTICIPANTS** can conduct post-training analysis like other **TRAINING ANALYSTS**, e.g., to get feedback on completed training sessions. On the other hand, they can have a specific responsibility during the *planning* or *execution* phases, respectively.

The high-level roles that emerged from the life cycle proved to be too general to capture the fine-grained requirements of heterogeneous groups of people participating in real training events. Therefore, we employed the personas design method [37] to reveal archetypal users and further decompose user roles. We analyzed the same sources that we used during the conceptualization of the life cycle [32]–[35]. The observed personas are summarized in Table 1.

CTF training includes only two types of personas, which correspond to a teacher-student relation. The *student* (or *learner*) follows instructions defined by the *training definition* and performs the required tasks. The *instructor* facilitates the training session from the educational point of view. Moreover, the instructor is also responsible for the technical aspects of training and addresses any possible technical difficulties with the underlying infrastructure.

In CDXs, we identified seven personas. *Blue team* members are similar to *learners* of CTFs. They have to defend the entrusted network from the attacks of the *red team*. *White team* members are responsible for the organization and compliance with the “game rules” of a CDX. *Fictitious users* represent common users of the defended network. *Law enforcement officers* check whether the actions of the *blue team* are legal. *Journalists* request reports from the *blue teams*.

Finally, the *green team* is responsible for maintaining the infrastructure of the exercise.

By deeply analyzing the responsibilities and analytical goals of identified personas, we generalized them to four user roles. The mapping is captured in Table 1.

TABLE 1  
Mapping of CTF/CDX personas to fine-grained user roles.

user roles	CTF personas	CDX personas
<i>trainee</i>	student (learner)	blue team
<i>sparring partner</i>	–	red team white team fictitious user law enforcement officer journalist
<i>supervisor</i>	instructor	green team white team
<i>operator</i>	instructor	green team

TRAINEES solve tasks described in the *training definition*. Their activities are monitored and assessed. They can work either individually or in teams. For the sake of simplicity, we use the term “trainee” for both cases.

SPARRING PARTNERS represent individuals or teams involved in training sessions who actively compete with TRAINEES but who are not directly assessed. Sparring partners also follow the instructions from the *training definition*. However, their requirements for data analysis, feedback, and other educational aspects differ from the requirements for TRAINEES.

SUPERVISORS, unlike SPARRING PARTNERS, do not follow the exact rules of the *training definition*. They are responsible for overseeing the training session, enforcing rules, and other activities that are not exactly defined.

OPERATORS are responsible for the underlying (technical) infrastructure of the hands-on training. This role requires technical skills and a good knowledge of the underlying technologies. The work of operators can significantly affect the course of the exercise since any technical difficulties can devalue educational results regardless of how well the training session has been prepared.

All the roles distilled from personas represent participants directly involved in a specific training session. Therefore, they are defined as descendants of the PARTICIPANT role in the schema in Fig. 3. While TRAINEES are the primary subject of training sessions, SPARRING PARTNERS, together with SUPERVISORS and OPERATORS, represent backstage *organizing participants*.

## 4 DATA

Visualizations designed for operational cybersecurity deal with large data sets [15]. In contrast, training events are limited in time, resources, and the number of participants. As a result, the amount of data produced during the training sessions is also usually limited. However, the data is highly heterogeneous. Therefore, our classification has been developed iteratively together with the analysis of other parts of the VA model. The proposed scheme comes from the unified life cycle. Data categories reflect user roles and training phases during which the data is created. It enables us to clarify what data is available in each phase and define limitations to be considered in analytical visualizations.

**Technical scenarios (D<sub>1</sub>)** capture the technical aspects and predefined processes of a *training definition*. The technical aspects include, for example, the definition of the network topology, software running on individual network nodes (operating system, applications, services), and vulnerabilities injected in the network nodes. User procedures are defined as attack plans (attack vectors and their timing), TRAINEES’ tasks, hints, and other formalized steps.

**Assessment criteria (D<sub>2</sub>)** determine how to assess TRAINEES and how to measure whether learning objectives were achieved. Assessment criteria define metrics, indicators, and aspects of the training related to the evaluation of TRAINEES. Apart from that, the criteria can also include the definition of questionnaires for prerequisite testing of TRAINEES, assessment questions during the exercise, and post-training feedback surveys.

**User actions (D<sub>3</sub>)** are PARTICIPANTS’ actions monitored and collected during the *execution* phase. Examples include commands entered by TRAINEES, displayed hints, performed attacks or defenses and their results, intervention of SUPERVISORS, and other user-oriented events.

**Infrastructure data (D<sub>4</sub>)** represent the state of computer networks and the underlying technical infrastructure. The data encodes node availability, available services, packet flows, and the health of the infrastructure. The obtained information can be used for direct infrastructure surveillance, and the assessment of TRAINEES (e.g., TRAINEES can be penalized for the unavailability of required services).

**Assessment data (D<sub>5</sub>)** are related to the *assessment criteria* and determine the success rate of TRAINEES and their results in achieving learning objectives. The data encodes how successfully a particular user has solved a particular task (in percentages or as obtained penalties), time spent on tasks, answers to questionnaires, and other qualitative and quantitative indicators of the learning process. A great deal of quantitative data can be computed automatically by applying assessment criteria (D<sub>2</sub>) to monitored user actions and infrastructure data (D<sub>3</sub> and D<sub>4</sub>).

TABLE 2  
Data types mapping on life cycle phases, abstract data levels, and terminology from the paper.

	D <sub>1</sub> & D <sub>2</sub>	D <sub>3</sub> & D <sub>4</sub> & D <sub>5</sub>
<i>phase of creation</i>	planning	execution
<i>level of abstraction</i>	configuration data	operational data
<i>terminology</i>	training definition	training run

Mapping data categories to the planning and execution phases follows data abstraction as defined by Fowler for software systems [38]: D<sub>1</sub> and D<sub>2</sub> represent data from the *configuration level*. They are defined during the *planning* phase by DESIGNERS as a part of *training definitions*. D<sub>3</sub>–D<sub>5</sub> represent data from the *operational level*. They are acquired during the *execution* phase and we refer to them as *training runs*, as summarized in Table 2.

## 5 VISUALIZATIONS AND HYPOTHESES

According to the VA model of Sacha & Keim (see Fig. 1), requirements applied to visualizations are driven by hypotheses that people consider during their analytical activities.

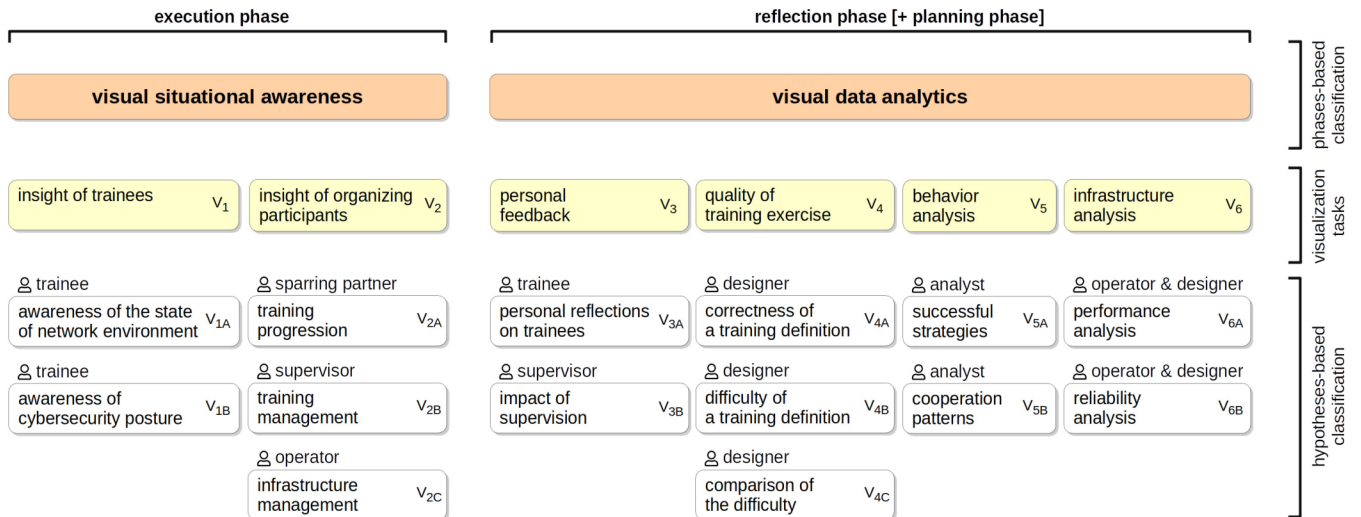


Fig. 4. Classification of visualizations and hypotheses in the context of hands-on cybersecurity training.

Therefore, we discuss and classify both visualizations and hypotheses together.

The classification shown in Fig. 4 was established iteratively by balancing two complementary directions. We broke down the top-level phases and roles of the training life cycle and, concurrently, we searched for low-level hypotheses that we organized into clusters. Balancing these two approaches, we concluded with a three-level classification scheme that, to the best of our knowledge, sufficiently covers the problem domain and emphasizes the design requirements of visual analytic tools. The low-level hypotheses were obtained from discussions with six domain experts (three of them are co-authors of this paper), each with more than six years of experience with organizing CTFs and CDXs. The final classification hierarchy was reached by consensus of the authors whose expertise includes cyber training design and organization as well as the design of analytical visualizations for KYPO Cyber Range [7]. The rest of this section is structured according to the proposed scheme as follows.

The top-level categories of *Visual Situational Awareness* and *Visual Data Analytics* in Fig. 4 represent distinct concepts using different data in different phases of the life cycle. They are discussed in two separate subsections. During conceptualization, we observed that the analytical tasks of TRAINING DESIGNERS represent a subset of activities associated with the *reflection* phase of TRAINING ANALYSTS. Hypotheses and visualizations of the *planning* phase are, therefore, covered by the *Visual Data Analytics* category.

Classification at the second level defines key visualization tasks  $V_1$ – $V_6$  that are detailed later in this section. They differ in the roles involved in the visual analysis, analytical goals, and other aspects. Discussion is primarily focused on visual requirements and justification for the third-level classification of hypotheses  $V_{1A}$ – $V_{6B}$ .

Providing an exhaustive list of hypotheses for each task  $V_{1A}$ – $V_{6B}$  is impossible; they emerge continuously as users conduct analyses and gain insights into the solved problem. Instead, we discuss an abstraction used for the classification and propose several hypotheses as examples.

## 5.1 Visual Situational Awareness

Existing theoretical concepts of situational awareness distinguish between *perception*, *comprehension*, and *projection* corresponding to the three levels of the well-known Endsley model [39]. However, the significance and meaning of the levels can differ in the context of cybersecurity training depending on users' roles and their goals. This is because providing comprehensive insight into cybersecurity events during the *execution* phase can be undesirable in certain circumstances. This aspect is reflected in our classification, as discussed in what follows. Table 3 summarizes visualizations and hypotheses for situational awareness.

**Insight of Trainees ( $V_1$ )** visualizations support TRAINEES in keeping track of what is happening at the moment and understanding the training content. The view on the data should be strictly person-centered and adapted to the history and performance of each particular TRAINEE so that they can concentrate on the development during the training session from their perspective.

The level of detail provided to TRAINEES has to be carefully considered when designing visualizations. A visual storytelling approach to learning can provide comprehensive guidance of TRAINEES throughout the training session. Using event-based visualizations emphasizing important actions and events that appeared during the *execution* phase can help the TRAINEES grasp the main ideas of the training content. However, this approach is rather exceptional, and visual guidance is usually intentionally restricted. A typical goal of hands-on cybersecurity training is just to exercise the *perception*, *comprehension*, and *projection* skills of TRAINEES; a subtle visual run-time support better mimics real-world conditions. The visual-based comprehension is often left for the *personal feedback* ( $V_4$ ) tools in the *reflection* phase (discussed later in this Section).

The clustering of hypotheses revealed two fields of TRAINEE interest. *Awareness of the state of the network environment* ( $V_{1A}$ ) covers hypotheses relevant to overseeing the state of the training network maintained by a TRAINEE. It is used to infer knowledge of hidden cyber events and actions

TABLE 3

Visual Situational Awareness: Visualization tasks  $V_1$  and  $V_2$  are further divided into two ( $V_{1A}$ – $V_{1B}$ ), and three ( $V_{2A}$ – $V_{2C}$ ) categories. Each category is accompanied by sample hypotheses formulated as prerequisites for verification (“I suppose that ...”).

<b><math>V_1</math> – Insight of Trainees</b>
<b>Awareness of the state of network environment (<math>V_{1A}</math>):</b> As a <i>trainee</i> , I suppose that ... ... the web running at host $X$ is accessible for users. ... the host $X$ is accessible for me via SSH. ... the external network (including internet) remains accessible.
<b>Awareness of cybersecurity posture (<math>V_{1B}</math>):</b> As a <i>trainee</i> , I suppose that ... ... server $X$ I am defending is now under attack. ... my previous attack actions were successful. ... I have successfully protected server $X$ against the DDoS attack.
<b><math>V_2</math> – Insight of Organizing Participants</b>
<b>Training progression (<math>V_{2A}</math>):</b> As a <i>sparring partner</i> , I suppose that ... ... the trainee $X$ completed task $Y$ , a prerequisite for task $Z$ . ... the DDoS attack against host $X$ defended by trainee $Y$ was successful. ... trainee $X$ fixed the vulnerability allowing a DDoS attack at host $Y$ .
<b>Training management (<math>V_{2B}</math>):</b> As a <i>supervisor</i> , I suppose that ... ... all trainees completed task $Y$ , a prerequisite for task $Z$ . ... trainee $X$ solved the task successfully. ... trainee $X$ is in trouble (working on task longer than $Y$ min).
<b>Infrastructure management (<math>V_{2C}</math>):</b> As an <i>operator</i> , I suppose that ... ... service $X$ at host $Y$ is up and running. ... service $X$ at host $Y$ is inaccessible longer than $Y$ min. ... network of trainee $X$ is connected to the rest of exercise infrastructure.

from the *infrastructure data* ( $D_4$ ). *Awareness of cybersecurity posture* ( $V_{1B}$ ) is related to the understanding of cyber events and actions defined as education goals in *training definitions*.

**Insight of Organizing Participants ( $V_2$ )** visualizations support SPARRING PARTNERS, SUPERVISORS, and OPERATORS in gaining insight into the state and progress of training sessions. Views are usually shared across all participants of the same role, providing them a view of the training progression, score, solved tasks, and other milestones and assessment data related to planning and timing. However, the views have to be adapted to each organizing role.  $V_2$  is, therefore, divided into three categories of hypotheses according to organizing roles. *Training progression* ( $V_{2A}$ ) is used by SPARRING PARTNERS who need to know the current state of the TRAINEES’ networks and services so that they can coordinate their actions and perform them in proper order and time. *Training management* ( $V_{2B}$ ) of SUPERVISORS should be able to identify troubles of TRAINEES as soon as possible. *Infrastructure management* ( $V_{2C}$ ) is intended for OPERATORS who have to monitor the unreliable infrastructure of the cyber range to detect technical problems.

Regardless of the specific role, the supervising activities of all organizing participants force them to *perceive* the current state of the training, to *comprehend* the situation, and to *project* the future status so that the training progresses smoothly and efficiently. In contrast to the *Insight of Trainees* ( $V_1$ ), analytical visualizations of organizing participants should fully support all these levels of awareness.

## 5.2 Visual Data Analytics

Our classification combines user roles of the cybersecurity training life cycle (see Fig. 2) and data categories (Section 4). Table 4 summarizes the classification of hypotheses that are explained in the remainder of this section.

TABLE 4

Visual Data Analytics: Visualization tasks  $V_3$ – $V_6$  are further divided into several categories (e.g.,  $V_{4A}$ – $V_{4C}$ ). Each category is accompanied by sample hypotheses formulated either as a prerequisite for verification (“I suppose that ...”), or as working empirical hypothesis that is assumed to be explaining certain fact about phenomena (“I wonder ...” and “I search for ...”).

<b><math>V_3</math> – Personal Feedback</b>
<b>Personal reflection of trainees (<math>V_{3A}</math>):</b> As a <i>trainee</i> , I wonder ... ... what I did wrong in the task $X$ . ... where I lost the most points and why. ... how I performed compared to other trainees.
<b>Impact of supervision (<math>V_{3B}</math>):</b> As a <i>supervisor</i> , I wonder ... ... if I intervened in time. ... if I intervened properly. ... if I overlooked some troubles.
<b><math>V_4</math> – Quality of Training Exercise</b>
<b>Correctness of a training definition (<math>V_{4A}</math>):</b> As a <i>designer</i> , I suppose that ... ... all tasks are relevant to learning objectives. ... task $X$ of the training definition $Y$ is solvable. ... the training definition $X$ is solvable as a whole (no logical flaws in connections and dependencies of individual tasks).
<b>Difficulty of a training definition (<math>V_{4B}</math>):</b> As a <i>designer</i> , I suppose that ... ... prerequisite skills of trainees were well-defined. ... the training definition $X$ is suitable for beginners/experts/... ... teams of trainees were well-balanced (there were no extreme differences in skills of each team).
<b>Comparison of the difficulty (<math>V_{4C}</math>):</b> As a <i>designer</i> , I suppose that ... ... the training definition $X$ is more difficult than definition $Y$ . ... tasks in the training definition $X$ require more time to finish than tasks in definition $Y$ . ... assessment criteria of the training definition $X$ were of lower quality than assessment criteria of definition $Y$ .
<b><math>V_5</math> – Behavior Analysis</b>
<b>Successful strategies (<math>V_{5A}</math>):</b> As an <i>analyst</i> , I suppose that ... ... limiting network access is a better strategy than fixing individual vulnerabilities in the network. ... dividing responsibility for defending individual hosts between team members is more efficient than ad-hoc defense.
<b>Cooperation patterns (<math>V_{5B}</math>):</b> As an <i>analyst</i> , I suppose that ... ... closer cooperation between team members leads to more effective protection against attacks. ... the team $X$ had a strong leader who communicated with the rest of the team significantly more often.
<b><math>V_6</math> – Infrastructure Analysis</b>
<b>Performance analysis (<math>V_{6A}</math>):</b> As an <i>operator</i> or <i>designer</i> , I search for ... ... the most utilized links/nodes/CPU's in the infrastructure for training definition $X$ . ... the peak memory usage of individual network nodes in training definition $X$ .
<b>Reliability analysis (<math>V_{6B}</math>):</b> As an <i>operator</i> or <i>designer</i> , I search for ... ... the mean time to failure of nodes in the infrastructure. ... unstable custom network services in the infrastructure.

**Personal Feedback ( $V_3$ )** to PARTICIPANTS has a significant positive impact on the learning process [40, p. 480]. A good post-training visual feedback should explain the pros

and cons of the chosen approach and indicate the areas for further improvement.

Effective person-centered feedback should occur as soon as possible, during or right after the *execution* phase when the TRAINEES remember details of their behavior, decisions, and conducted actions. Deploying such immediate visual feedback requires automated data processing and automatically generated personalized views for individual TRAINEES.

Our classification scheme is divided according to roles that benefit from timely feedback: *personal reflection of trainees* ( $V_{3A}$ ) and *impact of supervision* ( $V_{3B}$ ).

Personal feedback is crucial for the TRAINEES to learn from the exercise as much as possible. Nowadays, the feedback is often restricted to providing a simple scoreboard with very limited informal comments from SUPERVISORS (a so-called “hot wash-up” session). There might be an additional debriefing later when SUPERVISORS manually process the data. However, the analysis is laborious, and the delayed presentation of findings might reduce the impact on TRAINEES [32]. They should receive a view of their behavior during the training session as well as comparison with other TRAINEES. Moreover, the data analysis should be automated to provide in-depth feedback right after the training session. Feedback visualizations have to be well-designed and intuitive. Using common techniques would be necessary because TRAINEES usually do not have time to familiarize themselves with complex tools. A low number of easy-to-decode charts (bar/line charts, scatter plots, etc.) should be favored over the complex VA tools. The user interface should motivate users to explore the data and learn from their mistakes. Applying the methods of user-centered design [26], [41] is, hence, a must.

SUPERVISORS can also benefit from personalized feedback after a training session since their interventions influence TRAINEES. The visualizations should provide an overview as well as detailed per-trainee data. This allows SUPERVISORS to analyze the impact of their interventions and learn from their possible mistakes in managing the training session.

Feedback for SPARRING PARTNERS and OPERATORS is rare, since the main objective of the training is to teach TRAINEES. This is why we omitted these two roles from the classification.

**Quality of Training Exercise ( $V_4$ )** reflects the usefulness of training sessions for TRAINEES. The main motivation is to improve future training programs by reviewing collected data by DESIGNERS, i.e., experts with educational skills, who are responsible for the training content. The quality can be measured and compared by various qualitative attributes that capture individual features of training sessions. *Correctness*, for example, can express the ability of TRAINEES to solve required tasks considering properties of the underlying infrastructure, the logical consistency of tasks, or availability of meaningful instructions. *Difficulty* can be expressed as the time required to finish the training session or minimal skills required of TRAINEES. DESIGNERS can study either results of individual *training runs* of the same *training definition* or compare *training definitions* mutually.

Our classification scheme divides  $V_4$  hypotheses according to qualitative attributes and the multiplicity of

involved training runs: *Correctness of a training definition* ( $V_{4A}$ ), *difficulty of a training definition* ( $V_{4B}$ ), and *comparison of the difficulty* ( $V_{4C}$ ). Other qualitative attributes, apart from correctness or difficulty, can be considered. However, not all combinations are meaningful. For example, correctness typically represents a binary value (correct or incorrect) and then mutual comparison does not make sense.

The quality of a training session is primarily affected by three mutually connected factors:

- Training content defined by *technical scenario* ( $D_1$ ). Ambiguous or illogical tasks and their extreme difficulty or simplicity can discourage TRAINEES from proceeding, rendering the training session useless.
- Assessment defined by *assessment criteria* ( $D_2$ ). They affect achieving educational goals. Unbalanced assessment (too lax or strict) can lead to bypassing tasks or demotivate TRAINEES.
- Proficiency and motivation of TRAINEES. The lack of knowledge, skills, or motivation can prevent TRAINEES from finishing the training. Knowledge and skills are usually measured as part of prerequisite testing using questionnaires or small practical tasks.

Visual analytics can help to balance these factors by providing different views on the triplet and enabling DESIGNERS to study their mutual interactions and dependencies so that the impact of training is maximized for a given group of TRAINEES. Techniques of multiple coordinated views [42] can be used to support this exploratory analysis effectively.

**Behavior Analysis ( $V_5$ )** can help in discovering relevant facts about TRAINEES, their skills, or behavioral patterns under stress. The observations can either reveal issues or inconsistencies in training definitions or identify general patterns applicable in practical cyber defense. For instance, visualization of users’ actions can reveal patterns of successful cooperation or successful attack/defense strategies.

*Successful strategies* ( $V_{5A}$ ) and *cooperation patterns* ( $V_{5B}$ ) are two primary categories of analytical hypotheses directly related to cybersecurity education where visual perception can significantly help. The former analyzes defense and attack strategies, e.g., completely cutting off the defended network on the firewall vs. selective suspension of services being under attack. The analysis of cooperation patterns can be considered a part of the strategy analysis. However, it focuses more on people, their cooperation tactics, and how they influence the results of the training. The classification scheme can be extended to reflect other requirements of cybersecurity experts.

The raw data  $D_3 - D_5$  of *training runs* has usually a form of time-stamped events. Reconstruction, visualization, and analysis of user processes that produced the data are possible by employ techniques of process mining [43], [44]. Analysis of behavioral aspects can also be supported by specific statistical, knowledge discovery, or machine learning *models* incorporated into the VA process (see Fig. 1). For example, methods related to the node centrality in social networks [45] can be used to identify skilled leaders in team-based training sessions. Anomaly detection algorithms [46] can identify strong/weak skills of *trainees*, for instance.

These data can also serve to measure learning. [47] proposes several metrics for measuring performance that are applicable in cybersecurity training. These include tracking the time spent on tasks, observing the usage of specific tools in logs, or automatically checking properties of the virtual environment, such as uptime of services. A concrete example in the context of CDXs is presented in [48]: the evaluators measure the time of the attack, compromise, detection, mitigation, and restoration. In [49], also non-technical aspects are measured, such as team behavior.

**Infrastructure Analysis ( $V_6$ )** represents another essential activity that can affect the results and impact of cybersecurity training. Any technical difficulties or malfunctions can negatively influence TRAINEES. Related visualizations should support OPERATORS and DESIGNERS in exploring *training definitions* and their requirements on the infrastructure and provide them with a “backstage” view on the operational data captured in the *execution* phase.

As opposed to the *infrastructure management* ( $V_{2c}$ ) in situational awareness, this category relates to the feasibility of the underlying infrastructure to serve according to the prescription of the *training definitions*. For example, if a heavily used server is allocated on a shared virtual node in the cyber range, then its response time can be prohibitively slow. This can hinder TRAINEES in fulfilling the tasks.

Suitable visual tactics strongly depend on features and possibilities that are specific for technology used to implement the underlying infrastructure. Our classification, therefore, uses qualitative aspects that delimit generic requirements on the infrastructure: *performance analysis* ( $V_{6A}$ ) and *reliability analysis* ( $V_{6B}$ ). The performance deals with the utilization of resources at various levels of granularity (CPU, memory, network nodes). Reliability is related to the failure rate of individual facilities. However, these two qualities represent only an example.

## 6 DEMONSTRATION

In this section, we illustrate the application of our conceptual model on the KYPO Cyber Range platform, which is being developed by the cybersecurity team at our university since 2013. From the beginning, KYPO was designed with an emphasis on user-friendliness and support for providing interactive visual insight into cybersecurity and learning processes. It represents a comprehensive system suitable for demonstrating the applicability of our model. As the KYPO visualizations were designed on the fly without a conceptual view towards the application domain, this section aims to demonstrate how the model fits the existing design of a complex cyber range and to reveal the undersupported parts of the training life cycle. The presented visualizations only illustrate possible approaches to the design of specific visual analysis tools.

To the best of our knowledge, other cyber ranges and cybersecurity training tools focus primarily on the training content, providing only limited visual insight. Nevertheless, we aim to discuss other approaches when the KYPO does not provide a suitable example.

### 6.1 Training Life Cycles and Data in KYPO

The **KYPO Cyber Range** [7] is a highly flexible and scalable cloud-based platform. Its core functionality is to emulate

computer networks with full-fledged operating systems and network devices that mimic real-world systems. Its primary use is hands-on cybersecurity training, especially *attack-only* capture the flag games and cyber defense exercises. It is also used in other cybersecurity applications, such as forensic investigation. The platform provides tools for the automated collection of various data that can be further analyzed. These include network flows, computer logs, user commands, and user actions from GUI (e.g., mouse clicks or submitted forms).

The main user interface is a web application called the *KYPO portal*. We gradually extend the set of available visualizations and visual analytics tools integrated into the *KYPO portal* using the participatory design process. Nine cybersecurity experts (two specializing in cybersecurity education who are co-authors of this paper) closely collaborated in the design and evaluation of novel visualizations and the improvement of their features.

**Capture the Flag** games consist of tasks divided into consecutive levels where access to the next level is conditioned by completing the previous one. Players can use hints or skip entire levels. These actions (taking hints and skipping or completing a level) are penalized or rewarded by scoring points. The final scores of individual TRAINEES within the same session are mutually comparable and can be used for their evaluation. A typical session lasts for one to two hours. Several SUPERVISORS facilitate a group of up to 20 TRAINEES working as individuals or in pairs.

DESIGNERS of CTF games are experts from the cybersecurity incident response team of our university or undergraduate students of a one-semester course on designing cybersecurity games [50]. They produce *training definitions* that describe both *technical scenarios* ( $D_1$ ) and *assessment criteria* ( $D_2$ ). The training definition is a set of (plain text) documents that include: a description of the network environment and the configuration of individual network nodes (including vulnerabilities to be exploited in the game levels); a common background story and task descriptions (for each level); definition of hints, worked-out solutions and penalty points for taking hints (for each level); the TRAINEE’s prerequisites, educational objectives and further assessment criteria. *Designers* can interactively prepare content and allocate resources required for training sessions through the KYPO portal.

The produced *training definitions* are used for creating training sessions in the *execution* phase. The KYPO Cyber Range automatically logs TRAINEES’ *user actions* ( $D_3$ ). Some of the *training definitions* contain pre- and post-game questionnaires for assessing TRAINEE knowledge (i.e., *assessment data* ( $D_5$ )), which is stored as well. So far, *infrastructure data* ( $D_4$ ) collection is not supported in CTF games.

**Cyber Czech** is a series of technical cyber defense exercises for up to six *blue teams* (3–4 members). The TRAINEES must protect their infrastructure against various attacks from the *red team* and fulfill requests from other SPARRING PARTNERS, as defined in Sec. 3.2. The exercise spans two days. During the first day, the TRAINEES familiarize themselves with the virtual environment. The second day is devoted to the actual training session, which lasts 6 hours. A brief (up to 30 minutes) personalized feedback session follows right after the exercise. Finally, there is another



feedback session approximately two weeks later, in which organizers elaborate on the strengths and weaknesses of each team. From each exercise, we collect network flows, computer logs, user commands, and automatic and manual scoring records.

The variability and complexity of CDXs are substantially bigger than in CTFs. The preparation of a new training run of Cyber Czech exercise takes tens of person-months. A unique training definition is created almost from scratch each year and is only repeated a few times. Only a GUI for the *execution* and *reflection* phases are currently supported in the KYPO Portal, both to a limited extent.

The *technical scenario* ( $D_1$ ) is comprised of the infrastructure of nearly 200 computer nodes in multiple local networks, scheduled attacks and respective vulnerabilities, and configuration of monitoring tools for both trainees and organizers. Multiple iterations make the preparation very laborious. Each Cyber Czech exercise series is framed with a unique story and additional non-technical tasks. The *assessment criteria* ( $D_2$ ) include several dozen automatically scored network services (e.g., availability of web server or database) and up to 30 manually scored tasks (e.g., penalties for individual attacks, communication with the SPARRING PARTNERS from the *white* team or *fictitious users*), and requests for reverting malfunctioned network nodes. Complex dependencies in which one network service (e.g., active directory) depends on other services (such as DNS) often exist. All this complicates the design and implementation of a unified data scheme and corresponding front-end tools. Correctness and the estimation of difficulty of training definitions are addressed by so-called “dry runs” in which the whole exercise is tested by volunteers. However, the approach is costly and can be misleading because the readiness of testers may significantly differ from the readiness of target learners.

## 6.2 Visual Analytics of Capture the Flag Games

**Insight of Trainees ( $V_1$ ).** TRAINEES gain insight into the game content through the web-based KYPO portal, which provides them with task descriptions, hints, and solutions for each level and also shows information about the current level and remaining time of the training session. The *Network Topology* visualization (Fig. 5) mediates remote access to individual hosts via a web browser and provides situational awareness by decorating a simple network graph with various semantic symbols. For example, it is possible to support  $V_{1A}$  by coloring network links depending on current throughput, and  $V_{1B}$  by glyphs distinguishing logical roles of hosts (attacker, victim), or events captured in hosts (e.g., received mails). The importance and quantity of this semantic data differ between training definitions, and they also vary in time. Combining them meaningfully and showing them at the right time so that the TRAINEES are not overburdened is a challenging task.

**Insight of Organizing Participants ( $V_2$ ).** Since we currently support attack-only CTFs without SPARRING PARTNERS, no special visualizations for  $V_{2A}$  exist in KYPO.

1. We provide a full-page version of the visualization in Supplementary Materials at <https://www.kypo.cz/media/3197111/tvcg19-supplemental-materials.pdf>

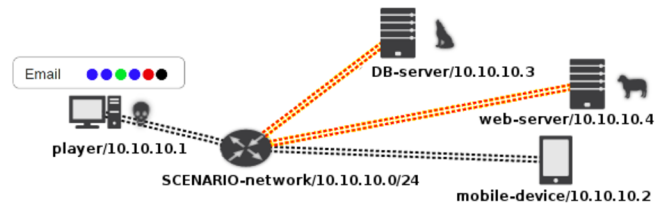


Fig. 5. Network Topology with glyphs supporting situational awareness.<sup>1</sup>

SUPERVISORS use *CTF Training Session Overview* visualization (Fig. 6) that displays the progress of TRAINEES throughout the CTF game. Each row captures the training session of individual TRAINEES, who can start at slightly different times. Colored bars represent levels. Dots represent user events (e.g., taking a hint), vertical lines show expected level duration. SUPERVISORS use this view to actively manage the training session ( $V_{2B}$ ) by looking for TRAINEES in trouble (e.g., those stuck in a level for too long, those repeatedly trying to guess the flag to pass the level instead of solving the task, or those about to quit without trying, which is signaled by displaying all the hints and the solution shortly after each other).

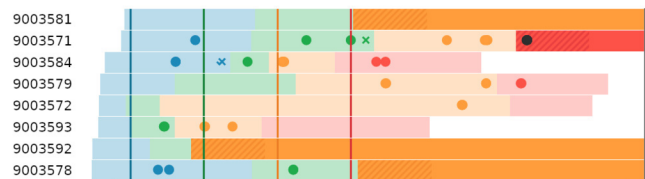


Fig. 6. CTF Training Session Overview shows the progress of individual trainees during the training session.<sup>1</sup>

Since our CTFs are executed in the complex cloud-based KYPO Cyber Range, dealing with technical issues is delegated to specialized *operators* managing this infrastructure. They gain insight into the infrastructure state ( $V_{2C}$ ) via off-the-shelf OpenNebula Sunstone dashboard (see supplemental materials<sup>1</sup>).

**Personal Feedback ( $V_3$ ).** At the end of a session, TRAINEES receive a *CTF Feedback Dashboard* [51] supporting  $V_{3A}$  with two complementary views (Fig. 7). The left view provides the final score overview for comparison with other TRAINEES. The lengths of the bars show the time of the slowest trainee; different color intensity provides information about the average time. The right side of the dashboard displays the individual score development in time throughout the game. The width of striped areas represents time spent in levels. Dots represent user events. A very similar dashboard is used by SUPERVISORS ( $V_{3B}$ ) who, in addition, can plot multiple TRAINEES into the score development time series chart for comparison.

**Quality of Training Exercise ( $V_4$ ).** Qualitative aspects of CTF *training definitions* are supported in KYPO by simple statistical visualizations, e.g., histograms and boxplots capturing the distribution of scores gained by TRAINEES. The *CTF Feedback Dashboard* (Fig. 7) from *personal feedback* ( $V_3$ ) can be also used to identify weak parts of the training, e.g. levels where TRAINEES spend a long time. However, deeper

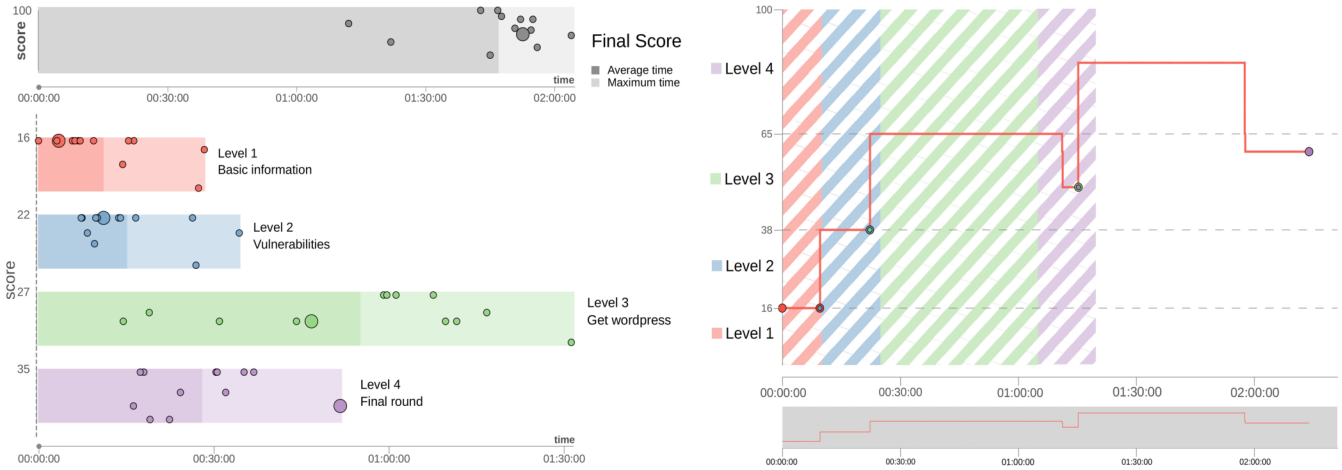


Fig. 7. CTF Feedback Dashboard providing individual view on TRAINEE's score results and development in time.<sup>1</sup>

research and the design of narrowly focused visualizations for quality-related analysis is a future work opportunity.

**Behavior Analysis ( $V_5$ ).** Behavior in connection with cybersecurity is often linked to attack graphs and estimation of weak points in the network. A study [52] introduced a method for analyzing computer network security. The method operates with attack paths that represent a linkage of individual nodes with conditions of compromised network security. The output is an attack graph with behavior prediction, and the authors propose the use of their method for incident response training. As for CTF games, the method could also bring insight to the trainee's actions and help the instructor to monitor progress or strategies.

**Infrastructure Analysis ( $V_6$ ).** The already mentioned off-the-shelf dashboard provided by OpenNebula Sunstone is currently used also for the basic qualitative evaluation of the underlying cloud infrastructure of the KYPO Cyber Range. However, its utilization for these tasks is not very effective, as it is a universal cloud management tool.

### 6.3 Visual Analytics of Cyber Czech

**Insight of Trainees ( $V_1$ ).** Since Cyber Czech is mainly a technical exercise, awareness of the network state  $V_{1A}$  and cybersecurity posture  $V_{1B}$  are intentionally restricted to resemble real-world settings, as discussed in Section 5. TRAINEES interact with a network topology visualization similar to Fig. 5. However, the network infrastructure is more complex, and there are no semantic decorations. Instead, the TRAINEES use a standard monitoring tool (Nagios) showing the status of the network services they are trying to protect. Further, they can infer the consequences of their actions only from the real-time *CDX Scoreboard* (Fig. 8) displayed during the exercise. The scoreboard shows the current total score as well as per-category scores and penalties of all *blue teams*, allowing them to compare themselves. The use of a restricted table-based view is intentional, as we aim to simulate real conditions during the CDX with only limited real-time feedback.

**Insight of Organizing Participants ( $V_2$ ).** Training progression ( $V_{2A}$ ) of the *red team* is supported by *CDX Attack Plan* (Fig. 9) showing the interactive plan of individual

Cyber Exercise Score						
Team Name	Services	Attacks	Injects	Users	VNC	Total Score
Blue Team 1	91,843	-8,500	9,000	-1,100	0	91,243
Blue Team 4	74,518	-11,000	6,650	0	-4,000	66,168
Blue Team 3	85,756	-12,000	2,475	-1,700	-9,500	65,031

Fig. 8. CDX Scoreboard shows the current scores of all *blue teams*.<sup>1</sup>

attacks and their state (inactive/ongoing/completed). The green color stands for successful attacks; red stands for unsuccessful ones (i.e., the *blue team* has defended themselves). Attack type abbreviations and given penalty points are shown within each block. Clicking on an attack block reveals further details (e.g., additional comments or screenshots). The *green team* uses the *Nagios* service monitoring system to watch the infrastructure ( $V_{2C}$ ), to detect when the trainees (un)intentionally blocked some of the monitored and scored services, and to provide brief advice ( $V_{2B}$ ). Visual insight of other organizing participants is not currently supported.

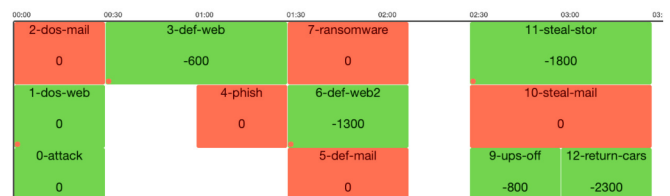


Fig. 9. CDX Attack Plan displays scheduled attacks of the *red team* at the end of a 6-hour long training session.<sup>1</sup>

**Personal Feedback ( $V_3$ ).** During the hot-washup session, organizers give immediate verbal feedback to TRAINEES. *Personal reflections on the trainees ( $V_{3A}$ )* are supported by presenting them the *CDX Attack Plan* (Fig. 9) that was hidden from the TRAINEES during the exercise. TRAINEES are also provided with the *CDX Personalized Feedback* [53] (Fig. 10) that shows the score development of their *blue team*. Dots include details about penalties entered by *red*, *white*, and *green teams*. Each dot is associated with a short

feedback poll used for gathering further information from TRAINEES. The data is used in the follow-up analysis. The impact of supervision  $V_{3B}$  is not currently supported.

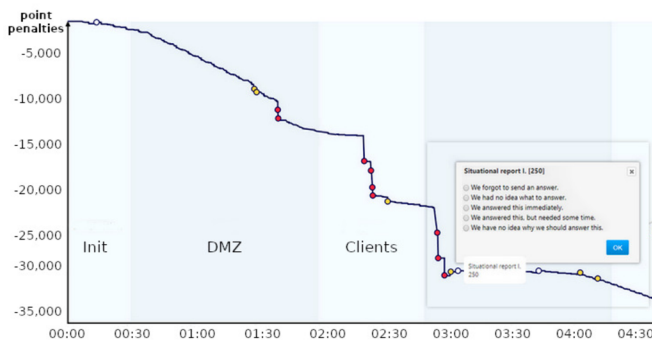


Fig. 10. CDX Personalized Feedback shows the score development throughout the training session of a single blue team.<sup>1</sup>

**Quality of Training Exercise ( $V_4$ ).** Vorobkalov and Kamaev [54] describe an approach to the quality estimation of e-learning systems. Their learning process model is based on an extended stochastic Petri net. The method has been implemented in an automated system, and it focuses on helping the expert to perform e-learning process analysis and to deduce learning course mistakes. However, it covers only systems based on net models. For CDX training, the model would not reflect the closely related state of the operational environment. Furthermore, when we consider the unstructured nature of CDX, the model would have to be very sophisticated and extensive.

**Behavior Analysis ( $V_5$ ).** The above-mentioned method by Bassett and Gabriel [52] can also be applied to the CDX use case. In this embodiment, the method could be utilized in the form of an attack tool to execute or simulate the events and conditions in the attack graph. The trainee would then receive the output, helping them identify attacks they were facing and allowing them to learn from the events retrospectively (since in CDX, we don't usually want to give them any instant feedback). However, such output would have to be further transformed into a visual form suitable for this type of training.

**Infrastructure Analysis ( $V_6$ ).** The support for this type of visual analysis is essentially non-existent at the moment. Although the KYPO platform collects some types of relevant data (e.g., system logs and commands entered by blue teams at individual network nodes), the data is processed ad-hoc and manually or not at all. This is usually done for a debriefing meeting of the organizing participants about a week after the training session. The attendees summarize their observations backed by collected data (e.g., feedback forms from the TRAINEES, analysis of the score development). To support the discussion, we are developing an analytical tool for CDX evaluation that will provide a timeline visualization of automatic and manual logs together with the communication threads among the *blue team* and corresponding *white team* members (Fig. 11).

## 7 DISCUSSION

In this section, we emphasize four key observations we attained and present the challenges for future visualization

research in the domain.

The current visualization tools support only situational awareness during the execution phase. The main focus of training sessions is on the execution phase. Therefore, visualizations are designed to provide insight both to trainees ( $V_1$ ) and organizing participants ( $V_2$ ). The reflection phase, in contrast, is vastly unsupported, with the exception of personal feedback ( $V_3$ ) for trainees.

Organizers have limited insight into the educational impact on learners. The design of cybersecurity training sessions is driven mainly by technical aspects. Training sessions often aim at mastering a particular cybersecurity technique or procedure without focusing on broader learning goals. To overcome this issue, the top-down approach of designing the training must be applied, starting from defining learning goals and going down to a selection of particular techniques. Visual measuring and comparing the quality of learned skills, which is largely overlooked, could help in this process. There is a broad unexplored research area in training quality ( $V_4$ ) and behavior ( $V_5$ ) analysis.

Organizers underestimate infrastructure monitoring and analysis. CTF and CDX depend heavily on customized monitoring and management tools for the underlying infrastructure ( $V_{2C}$ ). However, these tools are lacking. Low-level monitoring tools and other general-purpose solutions, which do not provide a complex overview of the situation, are preferred to customized ones. Analytical tools for post-event infrastructure analysis ( $V_6$ ) are also lacking.

Data collection is not a problem; data processing is. It is possible to collect large amounts of multivariate data either from the emulated network environment (e.g., network flows, computer logs, commands entered) or from the user interfaces of the cyber range (e.g., mouse tracking, and clicks). The bottleneck lies in data processing and presentation, as we point out in the demonstrative examples. Especially in CDX, data correlation is a difficult task. With rising interest in the quality of training exercise ( $V_4$ ), a behavior analysis ( $V_5$ ) could accelerate the demands on the use of the data.

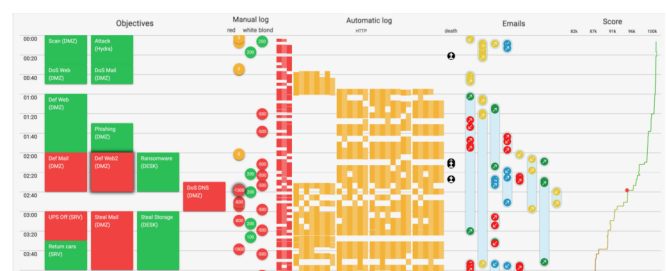


Fig. 11. Prototype of CDX Analytical Dashboard.<sup>1</sup>

Challenges for the visualization community are a reflection of the absence of tools. Table 5 summarizes users who benefit from the six visualization tasks, as revealed by the conceptual model in Section 5. Each bullet represents a visually-analytical use case. However, only a few use cases are somehow covered in current practice. For the post-exercise analysis, the main challenge is to find meaningful uses of the collected data to improve the SUPERVISORS' understanding of TRAINEES skill development as well as to provide insight into the training processes for DESIGNERS.

TABLE 5  
The mapping of the low-level roles on the visualization tasks.

	trainee	sparring partner	supervisor	designer	operator
V <sub>1</sub>	•				
V <sub>2</sub>		•	•		•
V <sub>3</sub>	•		•		
V <sub>4</sub>				•	
V <sub>5</sub>	•	•	•	•	•
V <sub>6</sub>				•	•

Another challenge is to design and develop VA tools to help the DESIGNERS and ORGANIZERS test their hypotheses. Last but not least, it is necessary to revisit the tools for situational awareness of participants during the exercise and provide them with timely individual feedback.

## 8 CONCLUSION AND FUTURE WORK

Hands-on cybersecurity training is crucial in educating the future workforce. However, measuring the effectiveness of the training process, using either technical or educational indicators, remains largely unexplored. Our work is motivated by a desire to improve these aspects by applying visual analytics. To the best of our knowledge, this paper is the first attempt to describe the application of VA models to hands-on cybersecurity education.

We used software engineering methods to describe the training life cycle and formalize user roles involved in cybersecurity training sessions. The foundations of our work lie in the existing generic VA models. We systematized the visualizations and hypotheses into six categories and demonstrated the application of the VA model on two classes of cybersecurity training hosted at the KYPO Cyber Range platform. The main limitation is the lack of details from other cyber ranges and training sessions. However, we assume that they are on a similar level of maturity. We back this claim with the experience of our university cybersecurity team members from their participation in events similar to the Cyber Czech exercise series.

Each of the six visualization tasks of the presented conceptual model deserves further investigation. The definition of specific guidelines that can help VA designers and researchers build visual tools is out of the scope of this paper. However, this paper aims to serve as a framework for such guidelines, providing researchers relevant use cases where the application of VA is demanding. We hope that our work will help to establish the agenda for advancing the state of the art and motivate other visualization researchers to explore the domain in which the research areas of education, cybersecurity, and data visualization intersect.

## ACKNOWLEDGMENT

This research was supported by ERDF “CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence” (No. CZ.02.1.01/0.0/0.0/16\_019/0000822). Computational resources were provided by the European Regional Development Fund Project CERIT Scientific Cloud (No. CZ.02.1.01/0.0/0.0/16\_013/0001802).

## REFERENCES

- [1] D. Restuccia, “Job Market Intelligence: Cybersecurity Jobs,” Burning Glass Tech, Tech. Rep., 2015.
- [2] M. Krone *et al.*, “Visual Analysis of Biomolecular Cavities: State of the Art,” *Computer Graphics Forum*, 2016.
- [3] K. Lawonn *et al.*, “A Survey on Multimodal Medical Data Visualization,” in *Computer Graphics Forum*, vol. 37, no. 1. Wiley Online Library, 2018, pp. 413–438.
- [4] X. Huang *et al.*, “TrajGraph: A Graph-based Visual Analytics Approach to Studying Urban Network Centralities Using Taxi Trajectory Data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 160–169, 2016.
- [5] S. Govaerts *et al.*, “The Student Activity Meter for Awareness and Self-reflection,” in *CHI’12 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2012, pp. 869–884.
- [6] P.C. Wong and J. Thomas, “Visual Analytics,” *IEEE Computer Graphics and Applications*, no. 5, pp. 20–21, 2004.
- [7] J. Vykopal *et al.*, “KYPO Cyber Range: Design and Use Cases,” in *Proceedings of the 12th International Conference on Software Technologies – Volume 1: ICSoft*, v.S.M.C.E. Cardoso J., Maciaszek L., Ed. Madrid, Spain: SciTePress, 2017, pp. 310–321.
- [8] G. Vigna *et al.*, “Ten Years of iCTF: The Good, The Bad, and The Ugly,” in *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education*. San Diego, CA: USENIX Association, 2014.
- [9] A. Davis *et al.*, “The Fun and Future of CTF,” in *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*. San Diego, CA: USENIX Association, 2014.
- [10] J. Werther *et al.*, “Experiences in Cyber Security Education: The MIT Lincoln Laboratory Capture-the-flag Exercise,” in *Proceedings of the 4th Conference on Cyber Security Experimentation and Test*, ser. CSET’11. USENIX Association, 2011.
- [11] A. Doupé *et al.*, “Hit ‘Em Where It Hurts: A Live Security Exercise on Cyber Situational Awareness,” in *Proc. of the 27th Annual Computer Security Applications Conf.* ACM, 2011, pp. 51–61.
- [12] W.M. Petullo *et al.*, “The Use of Cyber-Defense Exercises in Undergraduate Computing Education,” in *2016 USENIX Workshop on Advances in Security Education (ASE 16)*. Austin, TX: USENIX Association, 2016.
- [13] C. Eagle, “Computer Security Competitions: Expanding Educational Outcomes,” *IEEE Security & Privacy*, vol. 11, no. 4, pp. 69–71, 2013.
- [14] D. Staheli *et al.*, “Visualization Evaluation for Cyber Security: Trends and Future Directions,” in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*. ACM, 2014, pp. 49–56.
- [15] D.M. Best, A. Endert, and D. Kidwell, “7 Key Challenges for Visualization in Cyber Network Defense,” in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*. ACM, 2014, pp. 33–40.
- [16] A.E. Attipoe *et al.*, “Visualization Tools for Network Security,” *Electronic Imaging*, vol. 2016, no. 1, pp. 1–8, 2016.
- [17] A. D’Amico *et al.*, “Cyber Operator Perspectives on Security Visualization,” in *Advances in Human Factors in Cybersecurity*. Springer, 2016, pp. 69–81.
- [18] C.N. Adams and D.H. Snider, “Effective Data Visualization in Cybersecurity,” in *SoutheastCon 2018*. IEEE, 2018, pp. 1–8.
- [19] J. Yuen, B. Turnbull, and J. Hernandez, “Visual Analytics for Cyber Red Teaming,” in *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*. IEEE, 2015, pp. 1–8.
- [20] D. Schweitzer and W. Brown, “Using Visualization to Teach Security,” *Journal of Computing Sciences in Colleges*, vol. 24, no. 5, pp. 143–150, 2009.
- [21] X. Yuan *et al.*, “Visualization Tools for Teaching Computer Security,” *ACM Transactions on Computing Education (TOCE)*, vol. 9, no. 4, p. 20, 2010.
- [22] E. Fouh, M. Akbar, and C.A. Shaffer, “The Role of Visualization in Computer Science Education,” *Computers in the Schools*, vol. 29, no. 1-2, pp. 95–117, 2012.
- [23] E.E. Fratt and R.S. Laramée, “Towards a Survey of Interactive Visualization for Education,” *EG UK Computer Graphics & Visual Computing*, 2018.
- [24] L.P. Macfadyen and S. Dawson, “Mining LMS data to develop an “early warning system” for educators: A proof of concept,” *Computers & education*, vol. 54, no. 2, pp. 588–599, 2010.
- [25] M.R. Endsley, *Designing for Situation Awareness: An Approach to User-centered Design*. CRC press, 2016.

- [26] S. McKenna, D. Staheli, and M. Meyer, "Unlocking User-centered Design Methods for Building Cyber Security Visualizations," in *2015 IEEE Symp. on Vis. for Cyber Security (VizSec)*. IEEE, 2015, pp. 1–8.
- [27] M. Sedlmair, M. Meyer, and T. Munzner, "Design Study Methodology: Reflections from the Trenches and the Stacks," *IEEE Trans. on Vis. and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.
- [28] L.C. Koh *et al.*, "Developing and Applying a User-centered Model for the Design and Implementation of Information Visualization Tools," in *15th Int. Conf. on Information Vis.* IEEE, 2011, pp. 90–95.
- [29] D. Keim *et al.*, Eds., *Mastering the Information Age: Solving Problems with Visual Analytics*. Goslar : Eurographics Association, 2010.
- [30] D. Sacha *et al.*, "Knowledge Generation Model for Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1604–1613, Dec 2014.
- [31] J. Saldaña, *The Coding Manual for Qualitative Researchers*. Sage, 2015.
- [32] J. Vykopal *et al.*, "Lessons Learned from Complex Hands-on Defence Exercises in a Cyber Range," in *Frontiers in Education Conference (FIE)*. IEEE, 2017, pp. 1–8.
- [33] U.J. Staff, "Joint Training Manual for the Armed Forces of the United States (CJCSM 3500.03 D)," *Washington, DC: Joint Chiefs of Staff*, 2012.
- [34] J. Kick, "Cyber Exercise Playbook," MITRE Corp., Bedford, MA, Tech. Rep., 2014.
- [35] H. Alliance, "CyberRX 2.0 Level I Playbook Participant and Facilitator Guide," HITRUST Alliance, LLC, Tech. Rep., 2015.
- [36] G. Schneider and J.P. Winters, *Applying Use Cases: A Practical Guide*. Pearson Education, 2001.
- [37] B. Hanington and B. Martin, *Universal Methods of Design: 100 ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers, 2012.
- [38] M. Fowler, *Analysis Patterns: Reusable Object Models*. Addison-Wesley Professional, 1997.
- [39] M.R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.
- [40] G. Petty, *Teaching Today: A Practical Guide*. Nelson Thornes, 2009.
- [41] D.A. Norman and S.W. Draper, *User Centered System Design: New Perspectives on Human-computer Interaction*. CRC Press, 1986.
- [42] J.C. Roberts, "State of the Art: Coordinated & Multiple Views in Exploratory Visualization," in *Fifth Int. Conf. on Coordinated and Multiple Views in Exploratory Visualization*. IEEE, 2007, pp. 61–71.
- [43] A. Weijters and W.M. van der Aalst, "Process Mining: Discovering Workflow Models from Event-based Data," in *Belgium-Netherlands Conf. on Artificial Intelligence*. Citeseer, 2001.
- [44] S. Kriglstein *et al.*, "Visual Analytics in Process Mining: Classification of Process Mining Techniques," in *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2016.
- [45] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths," *Social networks*, vol. 32, no. 3, pp. 245–251, 2010.
- [46] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.
- [47] S. Måses *et al.*, "Obtaining Better Metrics for Complex Serious Games Within Virtualised Simulation Environments," in *European Conference on Games Based Learning*, 2017, pp. 428–434.
- [48] K. Maennel, R. Ottis, and O. Maennel, "Improving and Measuring Learning Effectiveness at Cyber Defense Exercises," in *Nordic Conference on Secure IT Systems*. Springer, 2017, pp. 123–138.
- [49] D.S. Henshel *et al.*, "Predicting Proficiency in Cyber Defense Team Exercises," in *MILCOM 2016-2016 IEEE Military Communications Conference*. IEEE, 2016, pp. 776–781.
- [50] V. Švábenský *et al.*, "Enhancing Cybersecurity Skills by Creating Serious Games," in *Proc. of the 23rd Annual sConf. on Innovation and Technology in Computer Science Education*. ACM, 2018, pp. 194–199.
- [51] R. Ošlejšek *et al.*, "Visual Feedback for Players of Multi-Level Capture the Flag Games: Field Usability Study," in *2019 IEEE Symposium on Visualization for Cyber Security (VizSec)*. IEEE, 2019.
- [52] G. Bassett, "System and Method for Cyber Security Analysis and Human Behavior Prediction," Mar. 22 2016, uS Patent 9,292,695.
- [53] J. Vykopal *et al.*, "Timely Feedback in Unstructured Cybersecurity Exercises," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. ACM, 2018, pp. 173–178.
- [54] P. Vorobkalov and V. Kamaev, "Quality Estimation of e-Learning Systems," *Supplement to International Journal Information Technologies and Knowledge*, vol. 2, pp. 25–30, 2008.



**Radek Ošlejšek** received his Ph.D. degree in informatics from Masaryk University in Brno, the Czech Republic, in 2004 for the application of software engineering methods to the area of computer graphics. He is an assistant professor with the Faculty of Informatics, MU Brno. His current research interests include interactive visualizations, knowledge modeling, and exploratory data analysis.



**Vít Rusňák** is a researcher at the Computer Security Incident Response Team at Institute of Computer Science, Masaryk University. He received a PhD degree in Informatics from Masaryk University in Brno, the Czech Republic in 2016. His research interests include the user-centered design of interactive visualizations and collaborative user interfaces.



**Karolína Burská** is currently a PhD student of computer science at Masaryk University in the Czech Republic. In her research, she aims at visualization in the context of cybersecurity education. As a member of a team of Masaryk University called KYPO, which focuses on simulation and mitigation of cybernetic threats, she focuses on interactive techniques in scientific visualization and exploratory analytics within cybersecurity.



**Valdemar Švábenský** enjoys teaching, so it is no surprise that he researches how to train new cybersecurity experts effectively. Specifically, he analyzes data from KYPO cybersecurity games to provide personalized feedback to learners who practice their offensive security skills. He actively participates in computing education conferences and received the Masaryk University award for the best teachers.



**Jan Vykopal** received the PhD degree from Masaryk University, Brno, in computer systems and technologies in 2013 for network-based intrusion detection in high-speed networks. His current research interest is cybersecurity education, particularly active learning using cyber ranges and virtual environments. Jan has been designing and organizing various cybersecurity games and exercises, including the Czech national defense exercise, since 2015.



**Jakub Čegan** is KYPO Cyber Range Platform and Cyber Defence Exercise (CDX) project manager. His area of interest is the development of meaningful and engaging CDX and training and providing them to customers.