MASARYK UNIVERSITY

Visual Analytics in Cybersecurity Education

Karolína Burská

September 15, 2018

KYPO

Motivation

- Cybersecurity exercises and training events **can be** a suitable instrument to provide sufficient knowledge to the participants
- We need a suitable environment for the exercises to have the desired effect on the participants
- **How to increase the impact of the exercises?**

Cybersecurity & Visual Analytics

- Connection of cybersecurity and visual analytics approaches
- **Interaction is the key**
- The primary goal is **development of new visual analytics tools** and creating a platform for **efficient organization** of the cybersecurity training events and exercises.

The Main Aims

- Categorization of the heterogeneous information of the exercises
- Suitable model for utilization of the exercises data
- Design of exploratory visualizations
- Creation of a base for iterative refinement of the new visualizations

Visual Analytics

- *VA is the science of analytical reasoning faciliated by interactive visual interfaces.*
- Uses data to gather information and acquire the required knowledge.
- The intersection of education and cybersecurity has gaps that can be explored with VA
- Keim et al. designed a VA process for this purpose, which has been further extended.

Visual Analytics – Model

An extended model by Sacha is based on interplay of the base by Keim et al. and Gestalt theories. It enables complex processes between human and computer ("*human-is-the-loop*").

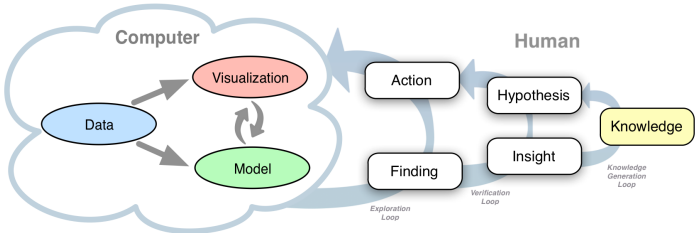


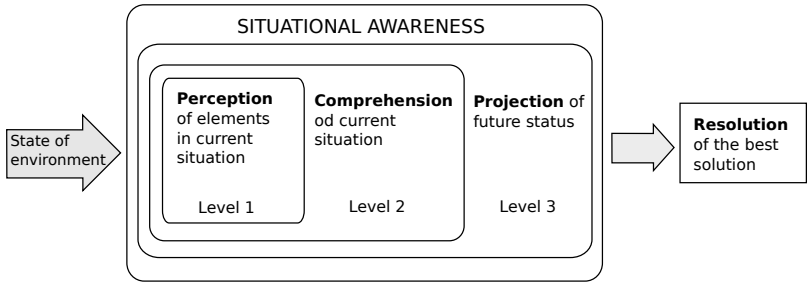
Figure: A knowledge generation model by Sacha

Cybersecurity exercises

- Many types of educational exercises
- Large-scale exercises -> long-term training programs / courses -> short games
- Variety of data to process - logs, runtime infrastructure data, scenario-specific data, etc.

Cybersecurity exercises

- Need for awareness over the current situation during the exercise
- The new tools help to gain an insight to both learners and organizers.
- Collection of cybersecurity-related data to iteratively build a knowledge base



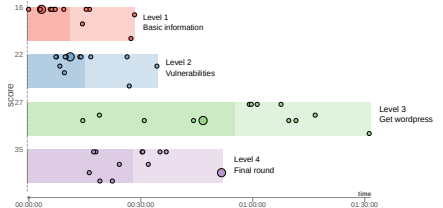
Interactive visualizations

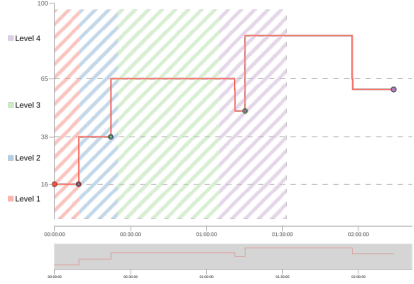
Visual data representation follows the Schneiderman mantra *"Overview first, zoom and filter, then details on demand"*.

- **Presentation** – presents the facts, what is known in advance, what needs to be visualized.
- **Hypothesis analysis** – users hypothesize based on the data and evaluate the analyzed findings.
- **Exploratory analysis** – not much knowledge of the data, nor any initial hypothesis. A multi-stage process that does not a priori define the way how to deal with the data. Typically, a taxonomy is needed.

What needs to be done?

- Mapping of the data to knowledge via unified taxonomy -> their utilization in form of OWL (Web Ontology Language) ontologies
- Utilization of an abstract model for knowledge generation. A base for data analysis, which distills the knowledge to refine exploratory visualizations
- Design of new methods and analytic visualizations





Estimated time

Player	Level 1 Score	Level 2 Score	Level 3 Score	Level 4 Score	Final Score
9003575	16	22	12	10	60
9003576	16	22	27	35	100
9003579	16	22	12	20	70
9003580	16	22	12	35	85
9003582	16	22	0	-	38
9003584	0	15	12	20	47
9003585	16	15	27	0	58
9003590	16	22	12	35	85
9003593	16	10	17	35	78
9003570	8	22	27	10	67
9003571	13	0	12	0	25
9003572	16	22	17	35	90
9003574	16	22	27	35	100

Event filters

- [x] Correct flags
- [ ] Wrong flags
- [ ] Hints taken
- [ ] Level skips

Summary

- The main motivation – very **limited and/or delayed feedback** in cybersecurity exercises
  - -> limited opportunity to learn from the mistakes
- **Players are seeking for assessment** regardless of their achieved score
- **Benefits for instructors** which lie in exploration of the data from the exercises