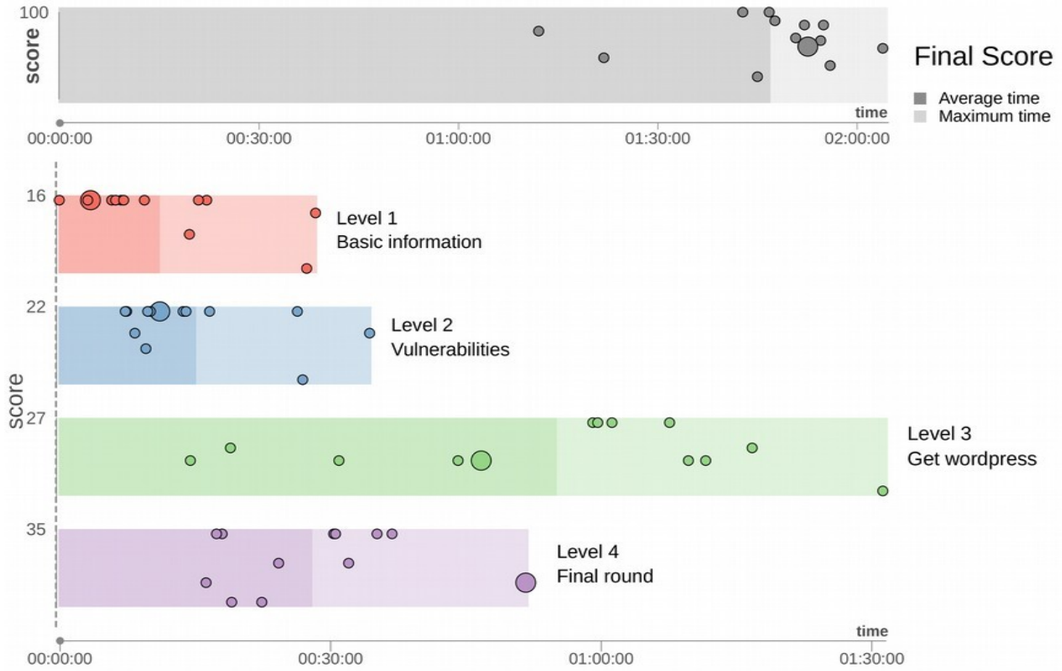
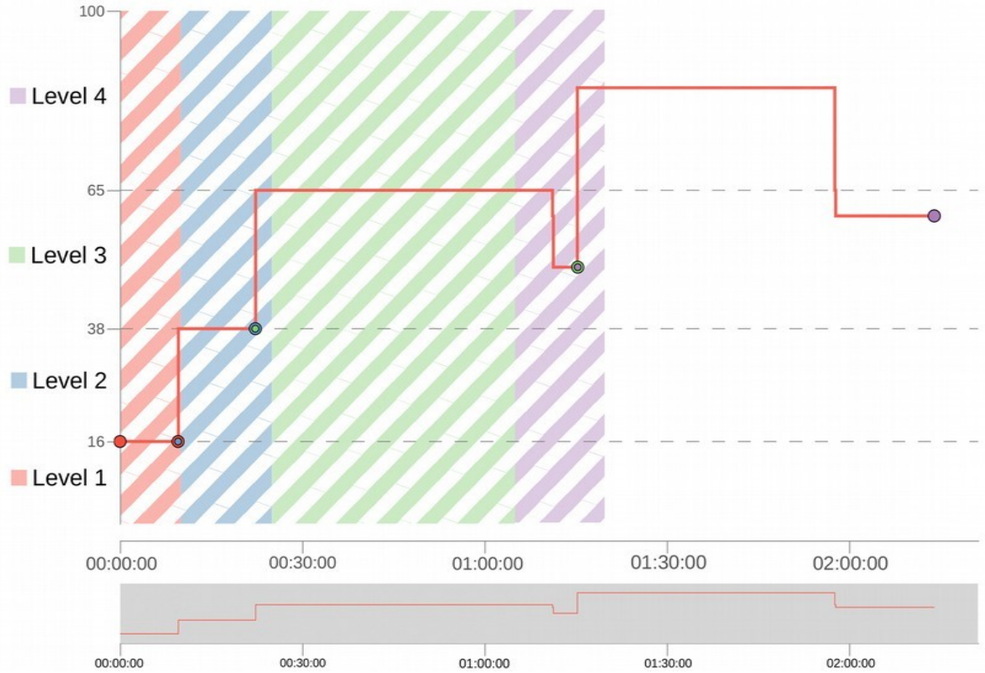
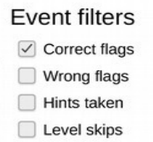




Player	Level 1 Score	Level 2 Score	Level 3 Score	Level 4 Score	Final Score
9003575	16	22	12	10	60
9003576	16	22	27	35	100
9003579	16	22	12	20	70
9003580	16	22	12	35	85
9003582	16	22	0	-	38
9003584	0	15	12	20	47
9003585	16	15	27	0	58
9003590	16	22	12	35	85
9003593	16	10	17	35	78
9003570	8	22	27	10	67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



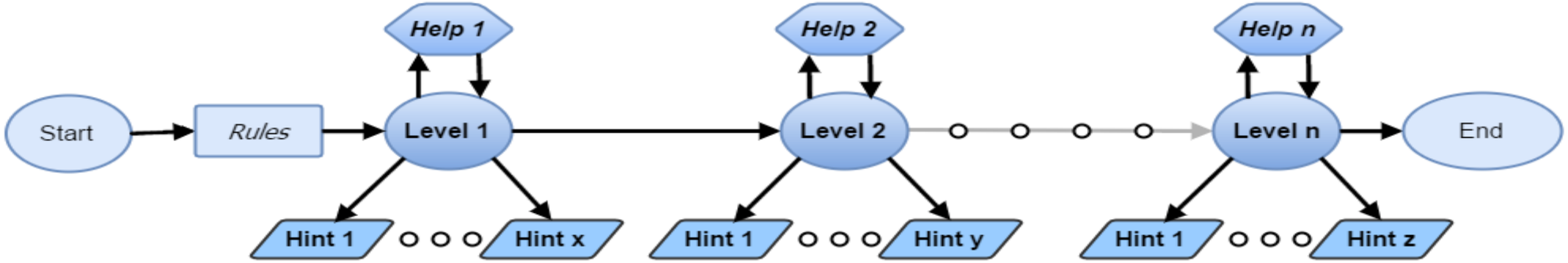
Visual Feedback for Players of Multi-Level Capture the Flag Games: Field Usability Study

Radek Ošlejšek, V. Rusňák, K. Burská, V. Švábenský, J. Vykopal

Masaryk University, Brno, Czech Republic

MUNI
FACULTY
OF INFORMATICS

Multi-level Capture the Flag Games



- In each level, a specific cybersecurity task has to be solved, e.g.,
  - Scan the network and find a vulnerable server
  - Use SQL injection to get into the server, ...
- Flag = secret code obtained at the end of a successful level used to proceed to the next level.
- Players can take hints or display a help (solution) to finish the level.
- Scoring data: points for correct flags, penalties for wrong flags, hints, ...

Research Goal

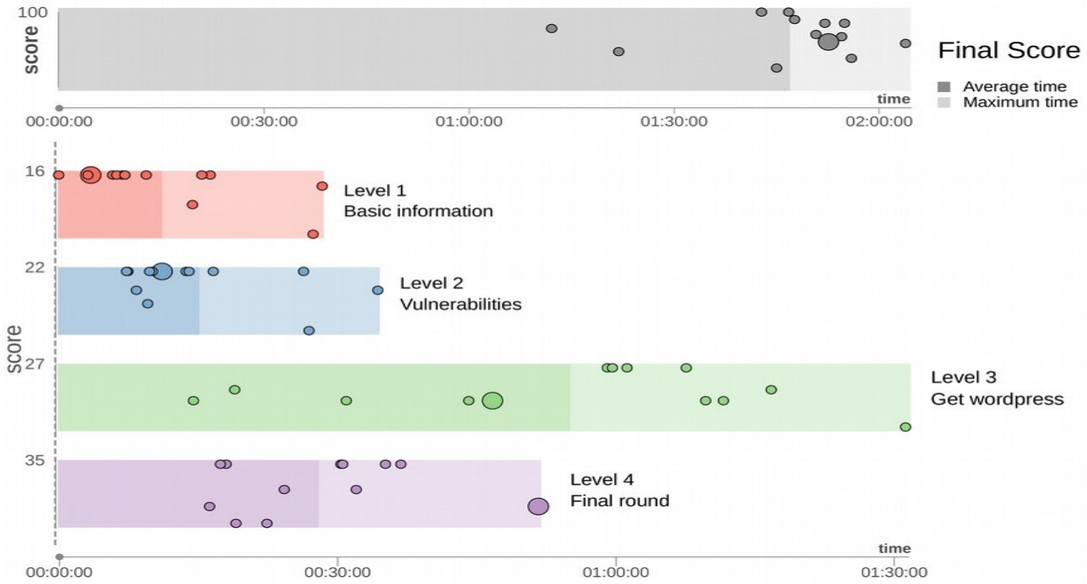
- **Problem:** Learning from doing is not sufficient
  - Trainees do not see various ways how to do things, e.g., multiple ways how to find a vulnerable server in the network.
  - Feedback provided by supervisors is OK, but is informal. Moreover, we want to support on-line plying without supervision at any time.
- **Solution:** Improving impact of CFT games by providing valuable feedback right after the exercise so that the players can learn from their mistakes and behavior.
  - **Automatically** generated from data
  - Easy to decode, i.e. interactive **visual** feedback

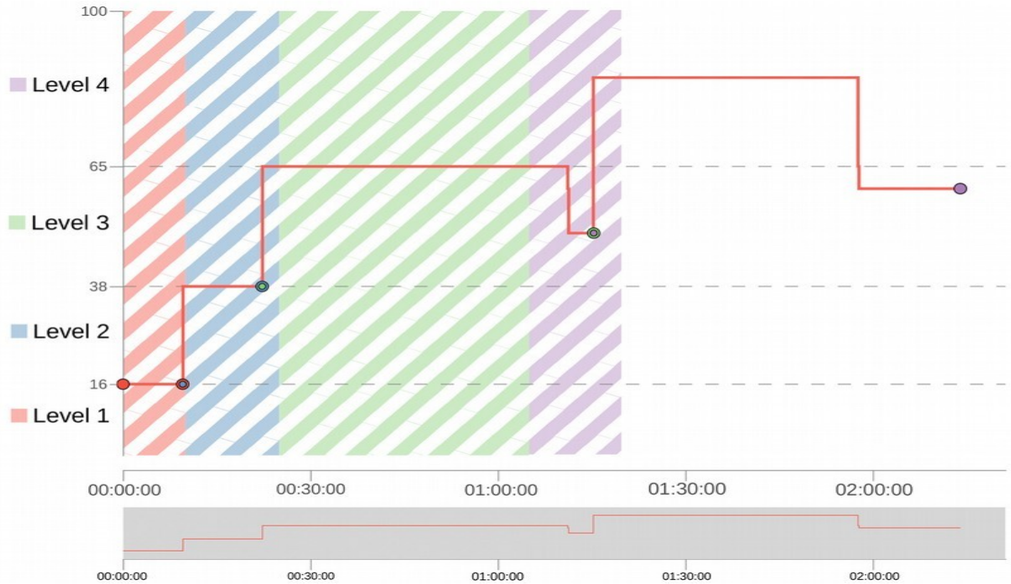
Feedback Requirements: To provide ...

- R1: ... **personalized feedback**
  - Person-centric view. Identification of well-done and problematic parts of my gameplay.
  - ”In which level did I lose the most points and why?”
- R2: ... **comparative feedback**
  - “Where I was better or worse than other players?”
  - Assessment of player’s abilities within a group, typically in a competition.
- R3: ... a brief **overview of the overall game results and features**
  - Insight into the game difficulty and other aspects that enable players to put their personal and comparative findings into the context of a particular game.
  - Useful mainly in situations when a user plays multiple games.

Design

- Two complementary visualizations
  - An easy to decode **overview of the score and time** spent to reach the score
  - **Detailed** view on **score development**





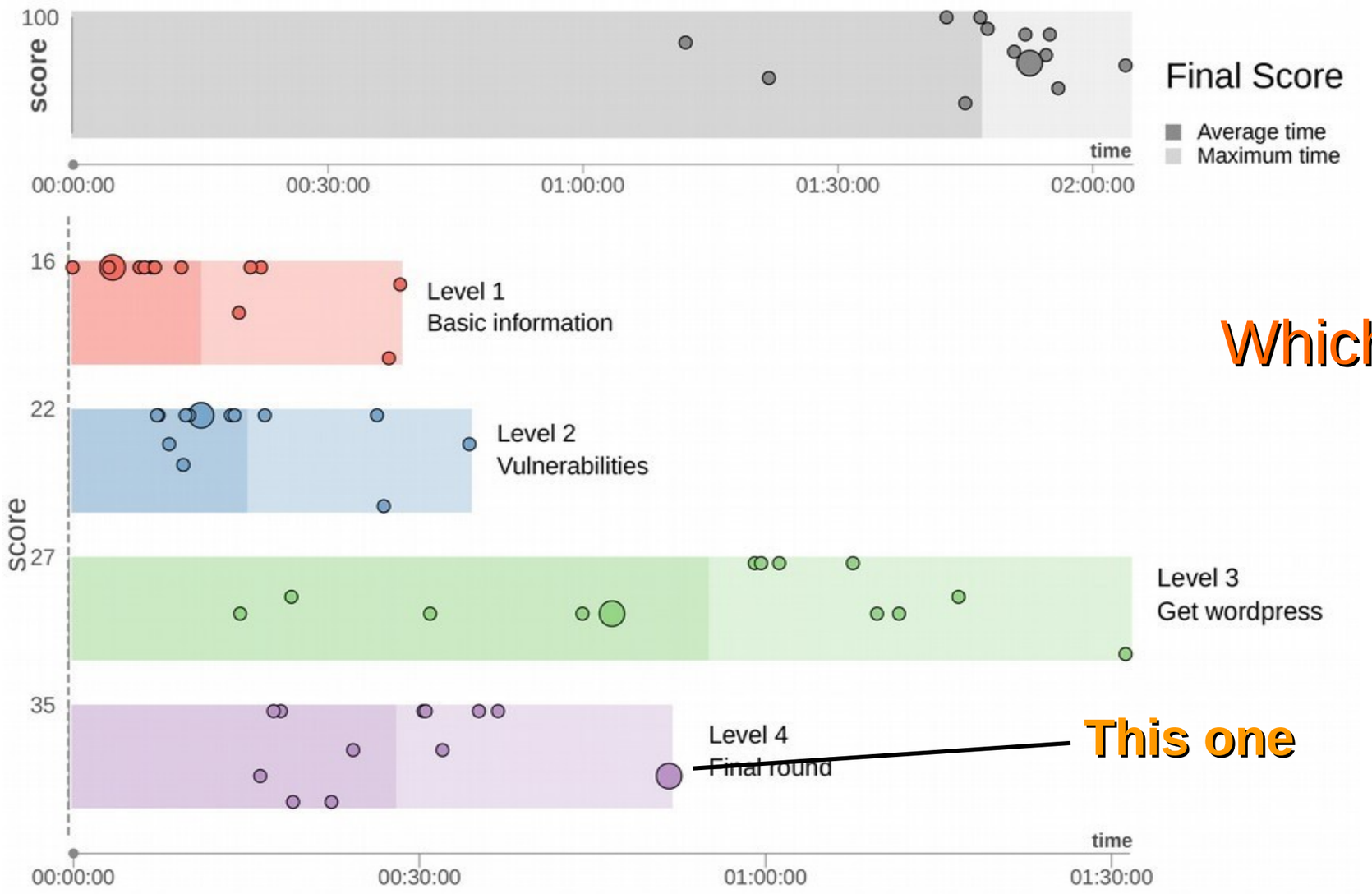
Player	Level 1 Score	Level 2 Score	Level 3 Score	Level 4 Score	Final Score
9003575	16	22	12	10	60
9003576	16	22	27	35	100
9003579	16	22	12	20	70
9003580	16	22	12	35	85
9003582	16	22	0	-	38
9003584	0	15	12	20	47
9003585	16	15	27	0	58
9003590	16	22	12	35	85
9003593	16	10	17	35	78
9003570	8	22	27	10	67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

Estimated time

Event filters

- Correct flags
- Wrong flags
- Hints taken
- Level skips

An overview of the score and time

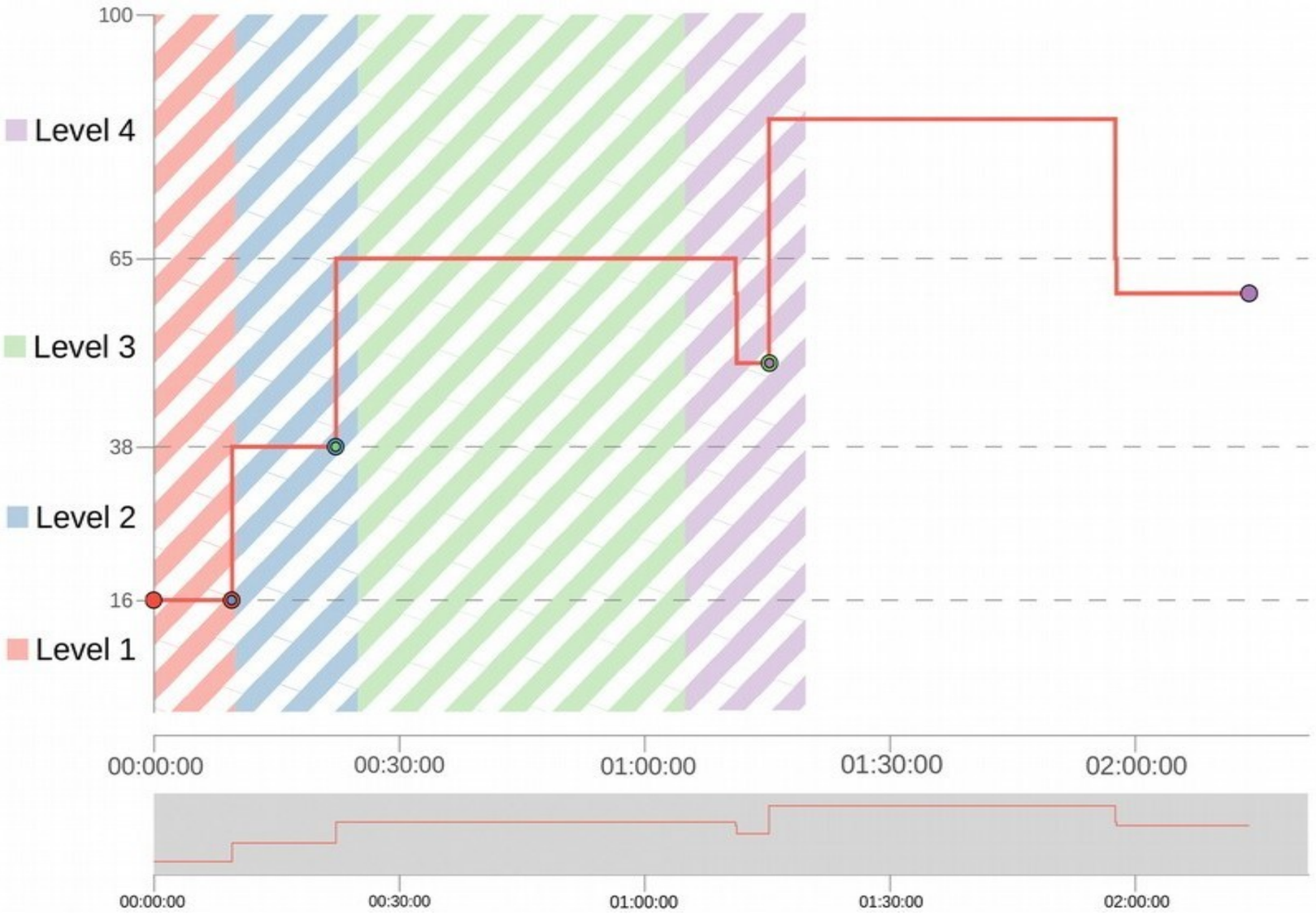


Which level was my worst?

This one

video

Detailed view on score development



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

Estimated time

Event filters

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

video

Evaluation

- **Hypotheses:**
  - H1: Requirements R1 – R3 are really meaningful and useful for players.
  - H2: The visual feedback is useful in providing insight into R1 – R3.
  - H3: Some visualizations or their parts are more useful for specific tasks or requirements R1 – R3 than others.
- **Setup** of the usability study:
  - 16 attendees of a summer school, 12 of them senior high-school students, all of them between 16 – 19 years old.
- **Procedure:**
  - CTF game play (~ 1 hour).
  - Familiarization with visualizations (~ 10 min).
  - Solving 12 tasks covering R1 – R3 (~ 15 min).

Results

- H1: Design requirements are correct, and the tasks reflect user interests. But ...
  - Players prefer exploration of *personal results* to the *overall game results* and comparison with others.
  - Unexpectedly, *comparative feedback* is the least useful.

- H2: Visualizations support trainees in the understanding of results.
  - Players were able to complete given tasks correctly, and tasks were easy to solve with the visual feedback.

Results (cont.)

H3: We did not find that any of the visualizations would better support either the personal results, comparative feedback, or the overall game results. But ...

- We identified specific tasks (across the requirements) for which one of the visualizations might be more appropriate.
- The results are uncertain due to data limitations, and further inspection is needed.

Conclusion and Future Work

- This preliminary usability study proved that our visual feedback is meaningful and worth of further development.
- We already extended the visualizations according to findings and integrated them in into the gaming environment.
- Extended data including *bash history* (commands executed by the players).
  - Reconstruction of user process from logs using process mining approaches.
  - Visual analysis of players behavior for educators and game designers.

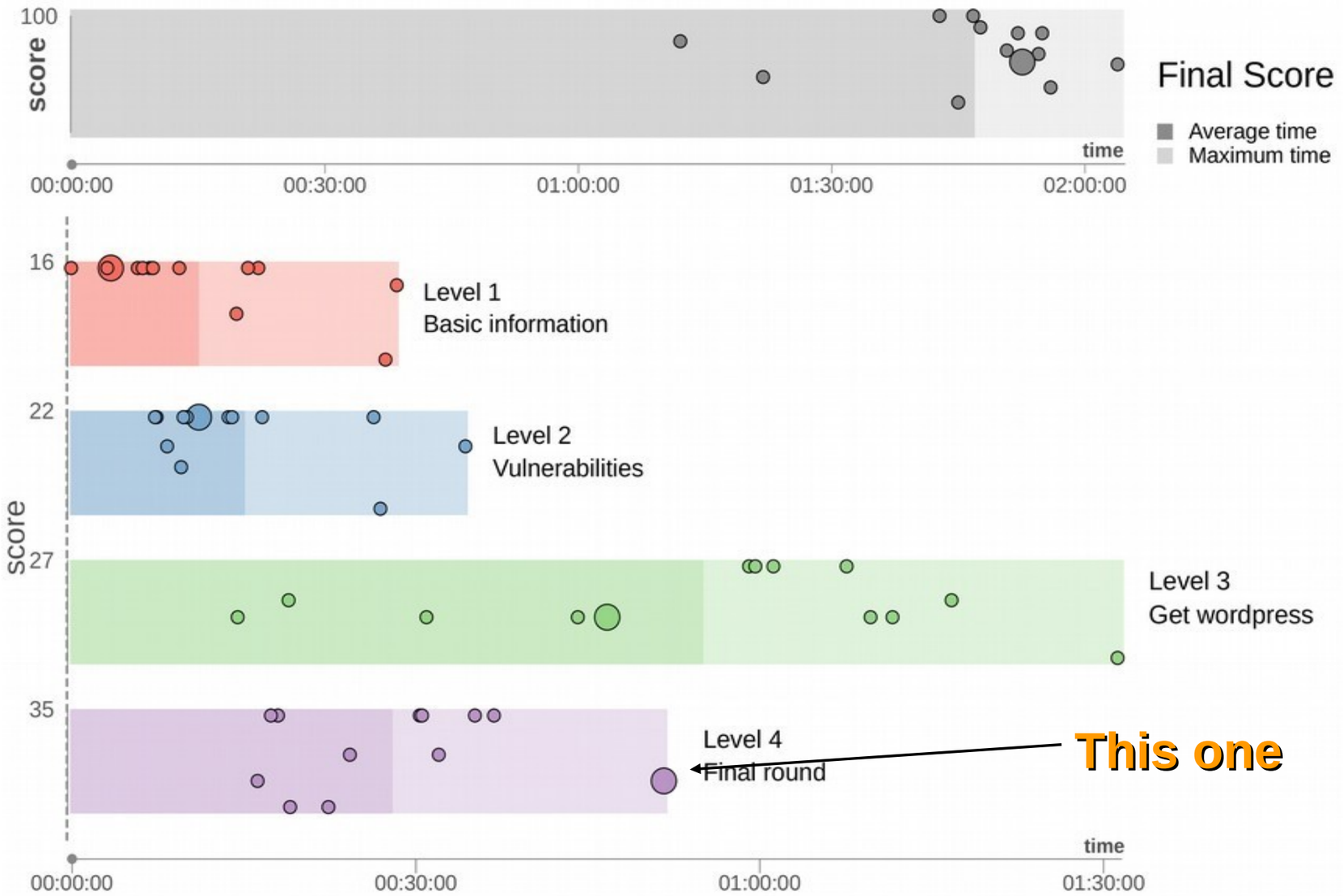
Thank you! Questions?

Online demo: www.radek-oslejsek.cz/it/supp-material/
Contact: oslejsek@fi.muni.cz

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Which level was my worst?



Which level was my best?