UNIVERSITÀ DI TRENTO

Cyber Security for Europe

# An Experimental Approach for Estimating Cyber Risk: a Proposal Building upon Cyber Ranges and Capture the Flags

**Giorgio Di Tizio**

joint work with Fabio Massacci, Luca Allodi, Stanislav Dashevskyi, Jelena Mirkovic

IEEE EuroS&P CACOE’20

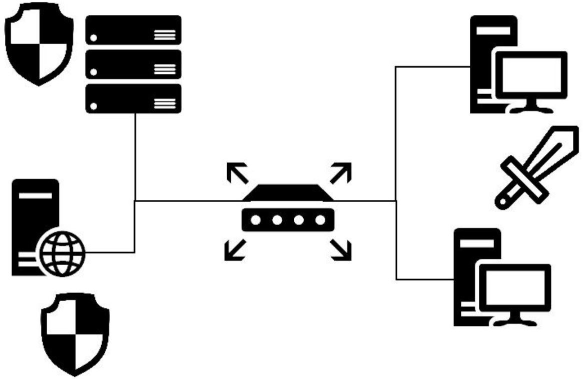
This work is part of CyberSec4Europe (cybersec4europe.eu), a project that has received funding from the European Union’s Horizon 2020 research and innovation programme H2020-SU-ICT-03-2018 under grant agreement No. 830929.

# Why we need an experiment to estimate the risk

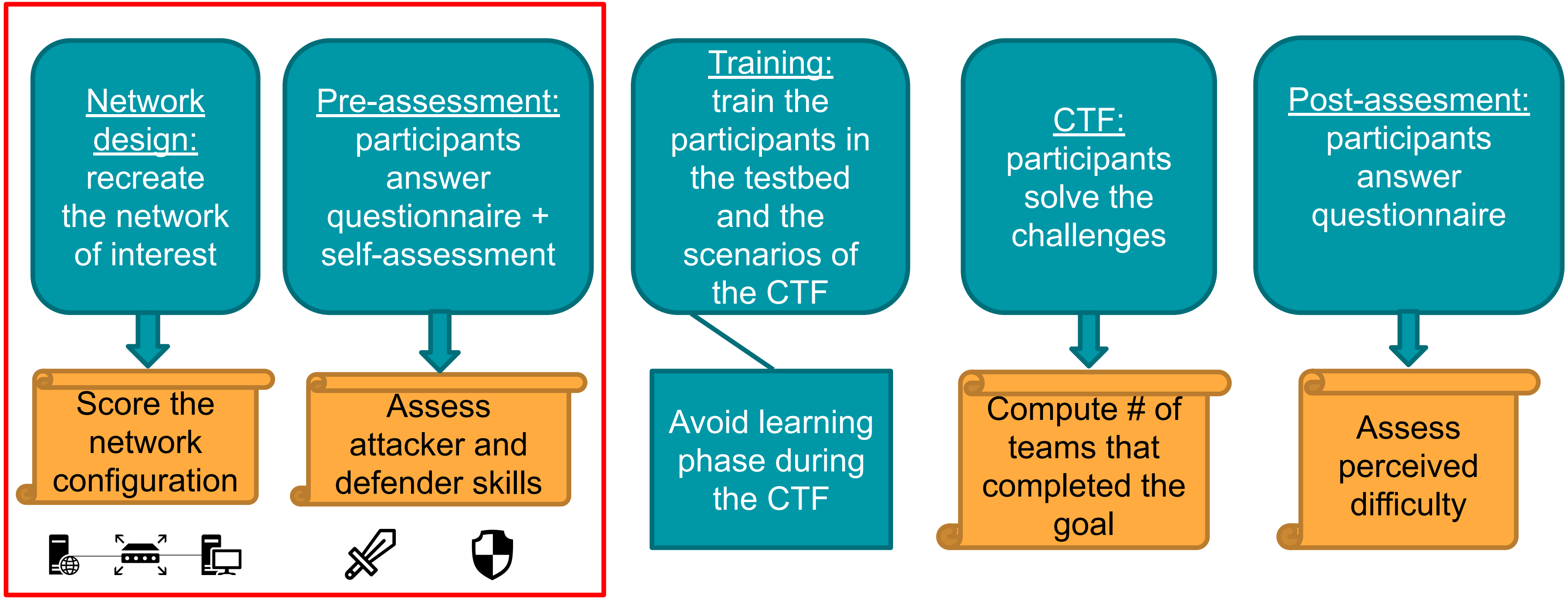
- Qualitative risk assessment is widely used in cyber-security
  - E.g. NIST risk matrices
- But it is expert dependent → causes wrong risk prioritization
- $\text{Risk} = \text{Impact} \times \Pr(\text{Attack}) \times \boxed{P(\text{Compr}|\text{Attack})}$
- $P(\text{Compr}|\text{Attack})$ cannot be computed using security data exhaust → data must be generated with an experiment

# Estimate P(Compr|Attack) using Capture the Flags

- Capture the flags → information security competitions where participants exploit and patch security vulnerabilities
- $P(\text{Compr}|\text{Attack}) = \dfrac{\#\ \text{teams completed the goal}}{\#\ \text{of teams in the CTF}}$
- Several factors influence this probability:
  - Attacker skills
  - Defender skills
  - Network configuration



# Experiment timeline



**Network design:** recreate the network of interest

Score the network configuration

**Pre-assessment:** participants answer questionnaire + self-assessment

Assess attacker and defender skills

**Training:** train the participants in the testbed and the scenarios of the CTF

Avoid learning phase during the CTF

**CTF:** participants solve the challenges

Compute # of teams that completed the goal

**Post-assesment:** participants answer questionnaire

Assess perceived difficulty

# How to design the network?

- Recreate the network we want to assess in a virtual environment
- Design decision (Control *vs* Realism):
  - More control: only one vulnerability is present, less machines in the network,...
  - More realism: multiple vulnerabilities are present, more machines in the network,...
- Available platforms to virtualize networks:
  - DETERLab, PlanetLab, GENI,...
- Containers-based frameworks can simplify the set-up
  - Labtainers, TestRex,...

# Scoring the network configuration

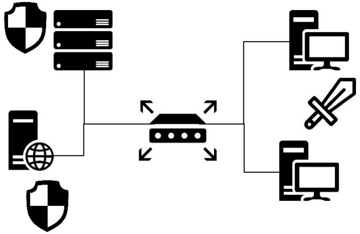
- How can we assign a score to each vulnerable network configuration?
- Existing security metrics are based on:
    - Attack graphs and network diversity -> precise but too complex for a fast assessment
    - **CVSS -> approximate but easy to compute for further statistical analysis**
        - **Open framework for communicating the characteristics and severity of software vulnerabilities**
        - **Used in the industry (e.g. Payment Card) and by federal governments**
            - **E.g. PCI-DSS compliant organizations must not have vulnerabilities with CVSS score >= 4.0**

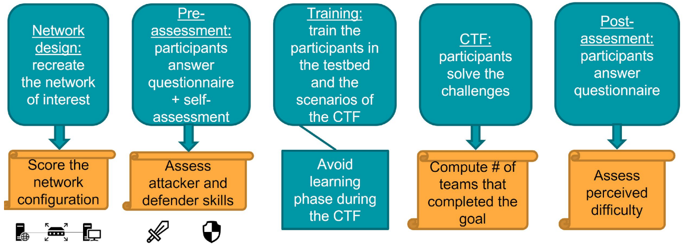
# Scoring the red and the blue team

- How can we measure the skills of the red and blue teams?
  - SANS questionnaire or war games -> precise but take too much time 🕒
  - Self-assessment or security certifications -> fast but unreliable ?
  - **Self-assessment validated with some SANS-style questions -> trade-off between 🕒 and ?**
- Impact of the blue team:
  - Jeopardy CTF: the system is automatically managed by experimenter’s scripts □ no impact
  - Attack-Defense CTF: defender will change the network configuration -> compare the results of the CTF with and without the blue team

# Conclusion and Future Works

- We propose a methodology to experimentally estimate risks using Capture the Flags





Network design: recreate the network of interest

Score the network configuration

Pre-assessment: participants answer questionnaire + self-assessment

Assess attacker and defender skills

Training: train the participants in the testbed and the scenarios of the CTF

Avoid learning phase during the CTF

CTF: participants solve the challenges

Compute # of teams that completed the goal

Post-assesment: participants answer questionnaire

Assess perceived difficulty

- Future works:
  - How to reduce the impact of the human factors?
    - E.g. Darpa Cyber Grand Challenge
  - How to estimate long-term attacks carried by APTs?
  - How to know if the red teams are representative of the criminal population?