SPIDER
5G CYBER RANGE

SPIDER: ML Applied to 5G Network Cyber Range

Authors: Stanislav Vakaruk (Universidad Politécnica de Madrid), Alberto Mozo (Universidad Politécnica de Madrid) and Antonio Pastor (Telefónica I+D)
Presenter: Stanislav Vakaruk

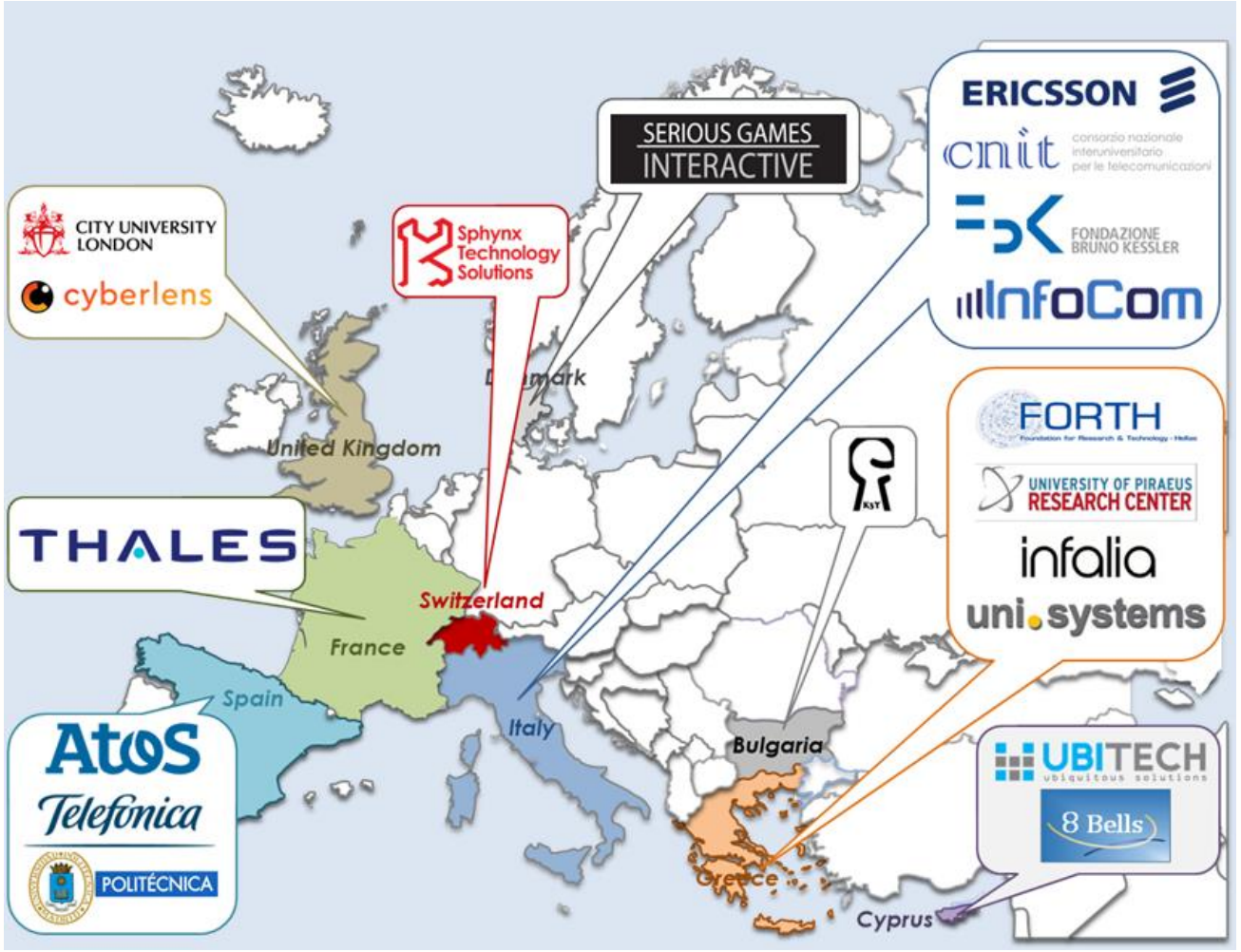
Index

Introduction: SPIDER Cyber Range

SPIDER Consortium

19 partners from **9 European countries** (high diversity)

- 5 x Large Industries
- 6 x Research Institutes and Universities
- 8 x SMEs



SPIDER Cyber Range

- Cyber ranges are well defined controlled **virtual environments** used in cybersecurity training as an efficient way **for trainees to gain practical knowledge** through hands on activities.
- **5G infrastructure** relies on the **latest virtualization technologies, increasing** the **exposition to** cyber-security **attack** vectors.
- The vision of H2020 SPIDER project is to deliver a next-generation, extensive, and replicable cyber range platform for the telecommunications domain.
  - Training has become extremely important:
    - SPIDER does not restrict the target group to ethical hackers/experts, that aim to leverage their competences, but to an increased audience covering risk assessors and non-expert users.
    - SPIDER aims to cover holistically the cyber security niche requirements of the 5G domain.

Why do we need ML in a Cyber Range

- Machine Learning (ML) **impacts** in cybersecurity in 2 dimensions:
  - ML based **tools** :
    - Anomaly detection
    - Identification of attacks (as spam, malware, phishing, …)
    - and others …
  - ML based **attacks***:
    - Leverage ML to **improve** malicious activities
      - Malware: obfuscate from antivirus, avoid spam filters, use cloud ML services …
      - Penetration test: password guessing, vulnerability scans, …
    - Use ML to **deceive** ML
      - Manipulate data sources:
        - » Adversarial networks (ML against ML -> Resilient ML)

*https://www.europol.europa.eu/publications-documents/malicious-uses-and-abuses-of-artificial-intelligence

Why do we need ML in a Cyber Range

SPIDER
5G CYBER RANGE

- Security experts needs to ***learn*** **how ML impact** in their job
  - Use ML based **tools to detect and mitigate** attacks:
    - **Understand the results** of a ML tool:
      - Confidence levels, False positive, True negatives, …
    - **Parametrize** ML tools (hyperparameters, confidence levels, …)
    - **Compare** different tools (ML or classical tools)
    - **Is not** an **infallible** but supplementary tool
  - Learn to live with **ML based attacks**

Ambition in SPIDER

- **Integrate ML tools** in SPIDER Cyber range
  - Infrastructure to **train and test** customized ML models
  - Use cases: ML-based attack detectors integrated in toolboxes to be utilised in **Cyber-exercises**

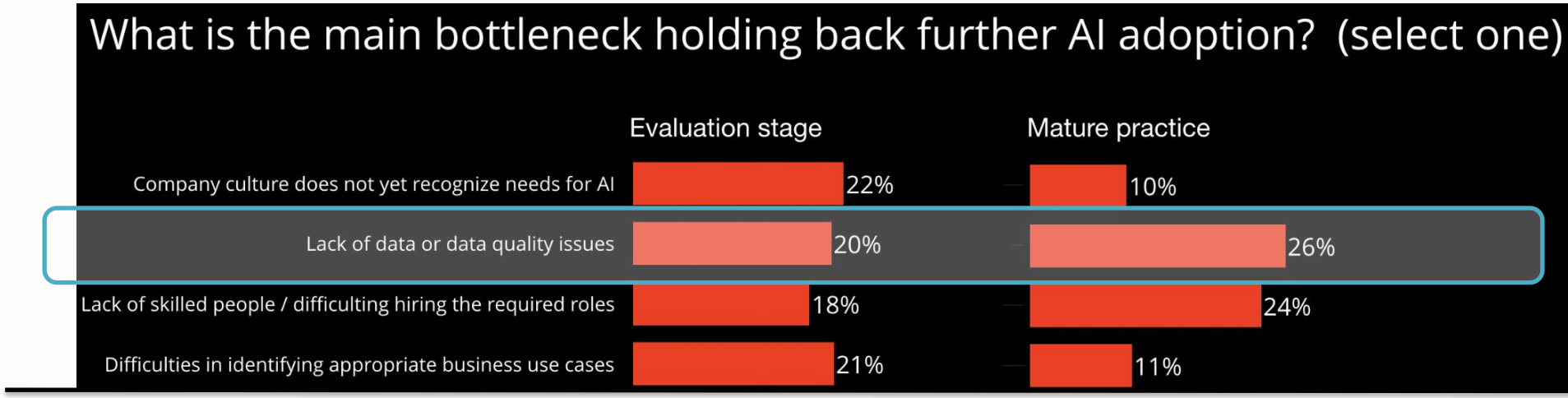
- Provide ML tools **exercises**:
  - Define and create some ML related **toolbox components**
  - Define and create some ML related **attacks**

SPIDER
5G CYBER RANGE

Traffic Generation with Mouseworld

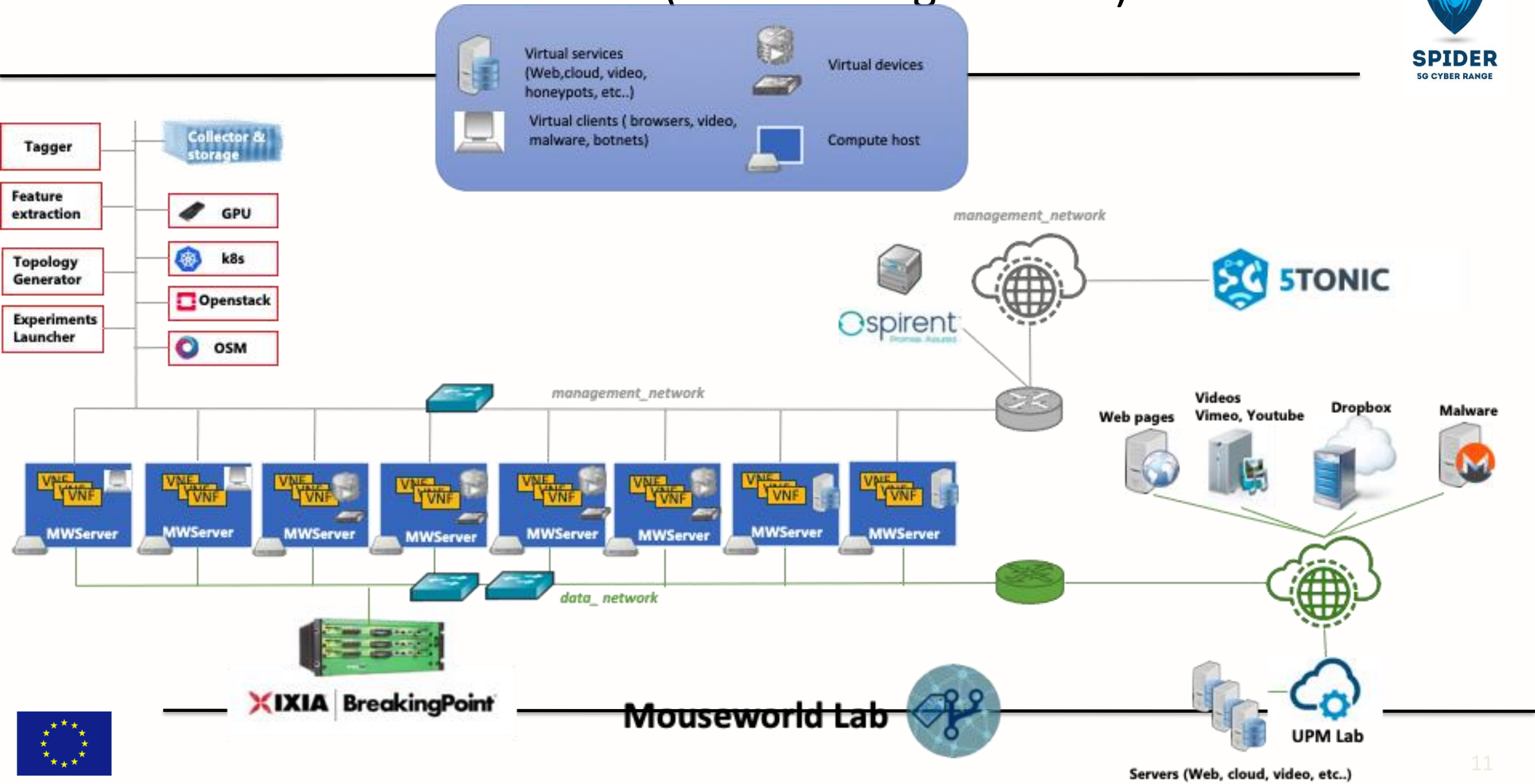
ML and data Thirst

- There is a serious **lack of training datasets**
  - Data as an asset ($$)
  - Privacy, regulatory concerns
  - Business interest
- There is a serious **Lack of LABELLED** data
  - Needed for ML algorithms
    - Supervised: training + validation
    - Unsupervised: validation



Source: https://www.oreilly.com/data/free/ai-adoption-in-the-enterprise.csp

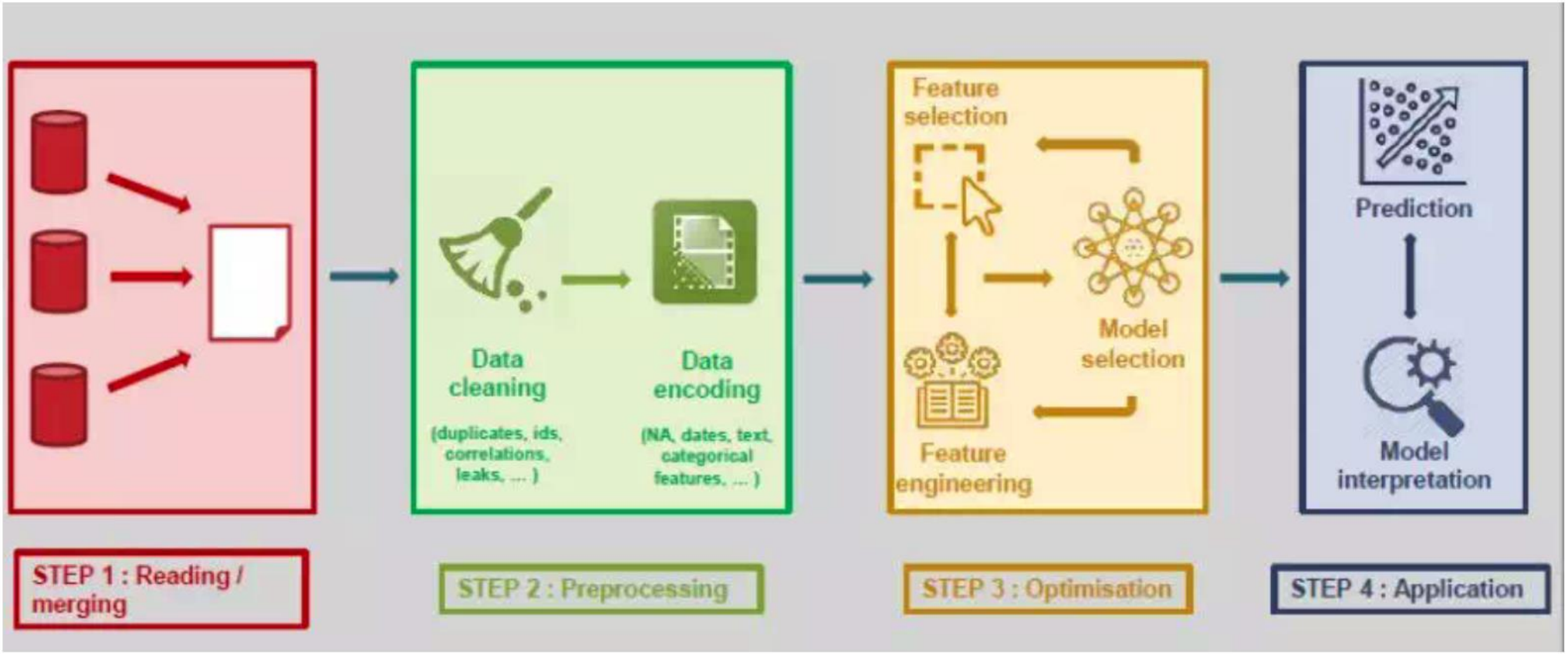
Mouseworld infrastructure (Network Digital Twin)



SPIDER
5G CYBER RANGE

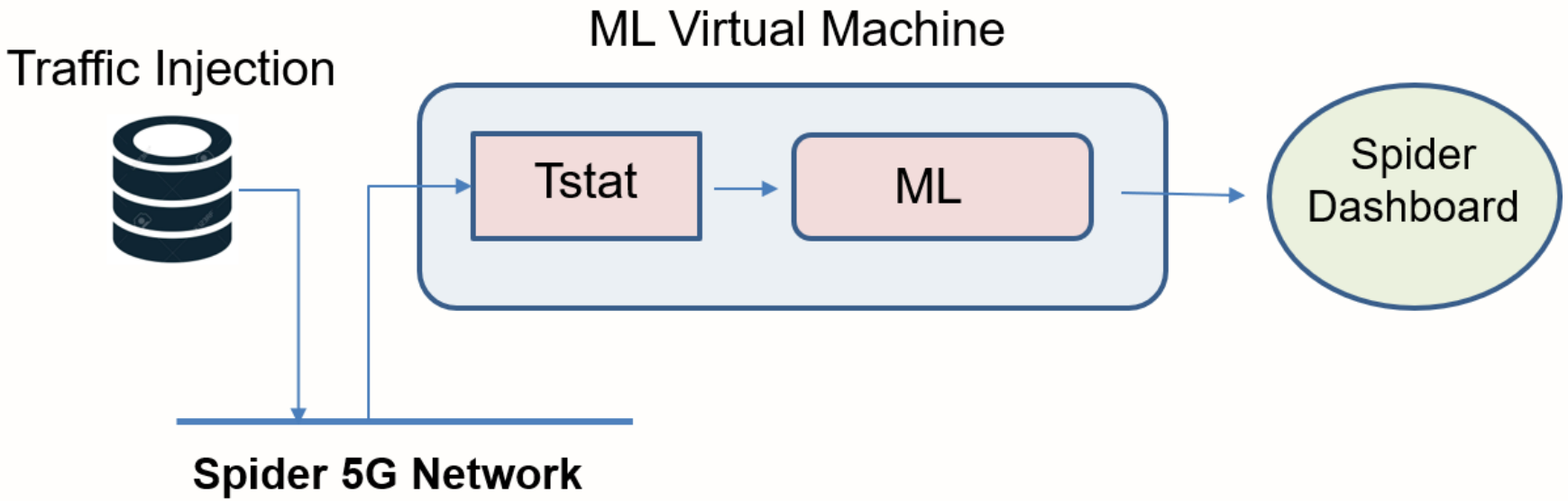
Machine Learning Attack Detector

Machine Learning model life cycle



Source: https://www.slideshare.net/AxeldeRomblay/mlbox-082-178177773

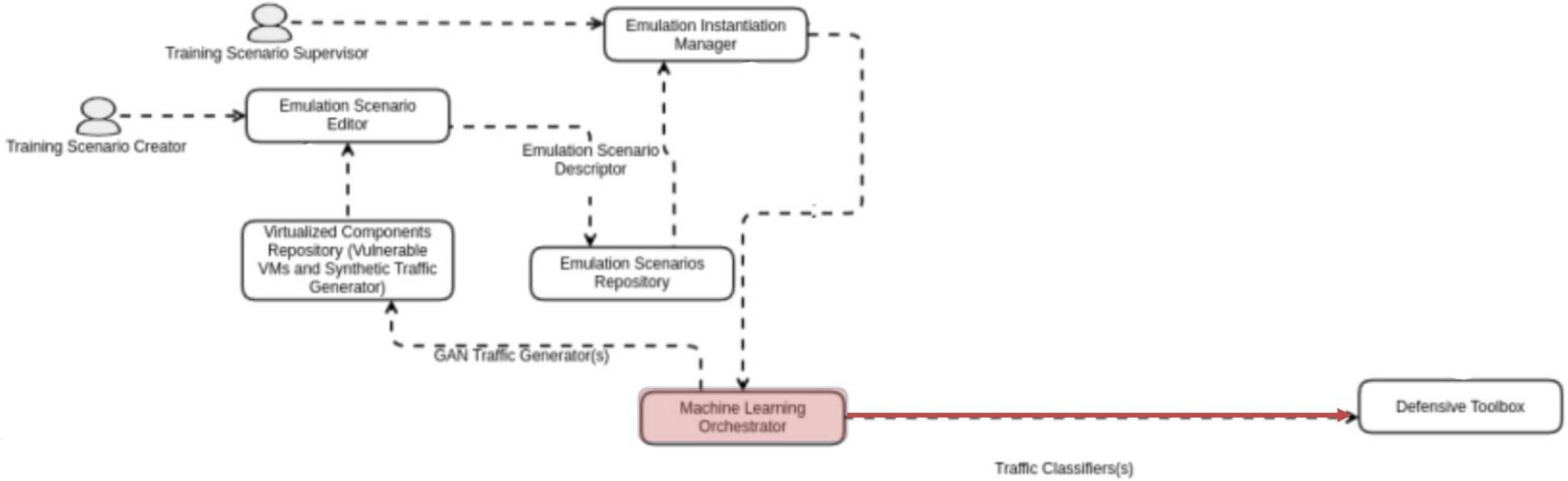
Machine Learning Attack Detector in Spider



- **Traffic is previously generated** in the Mouseworld laboratory
- The **traffic is mirrored** into the ML Virtual Machine
- Tstat captures and extracts traffic **flow-based information**

Integration into SPIDER Architecture

ML in Spider Architecture



Blue Team Training Process

Blue Team Training Process

- In SPIDER exercise scenarios:
  - The **trainees** who exercise with ML defensive tools are called **Blue Team**
  - The **experts** who run the attacks are called **Red Team**

- Blue Team is able to:
  - Select a **specific dataset** to be injected
  - Select a **specific ML model** to be deployed into the ML VM
  - Select the **ML minimum confidence value**
  - Review **ML results** in the Spider Dashboard

SPIDER
5G CYBER RANGE

Conclusions and Future Work

Conclusions and Future Work

SPIDER
5G CYBER RANGE

- Conclusions:
  - **Mouseworld** is integrated into SPIDER by **traffic injection**
  - **ML** is integrated into SPIDER as **packet aggregator and classifier**
  - **Trainees** are able to run exercises using ML modules in SPIDER

- Future Work:
  - Move from VMs to **Containers**
  - Integration of **non-supervised** models for anomaly detection

References

ML toolbox:

Detection of Encrypted Cryptomining Malware Connections With Machine and Deep Learning (IEEE Access, JCR Q1)

SPIDER:

Christos Xenakis, Anna Angelogianni, Eleni Veroni, Eirini Karapistoli, Matthias Ghering, Neofytos Gerosavva, … Antonio Pastor. (2020). The SPIDER concept: A Cyber Range as a Service platform. Presented at the EUROPEAN CONFERENCE ON NETWORKS AND COMMUNICATIONS (EUCNC2020), VIRTUAL: Zenodo. http://doi.org/10.5281/zenodo.4030473

SPIDER
5G CYBER RANGE

Thanks!