Shicheng ZHOU, Jingju LIU, Yuliang LU, Jiahai YANG, Yue ZHANG, Jie CHEN, 2025. Mind the Gap: towards generalizable autonomous penetration testing via domain randomization and meta-reinforcement learning. *Front Inform Technol Electron Eng*, 26(12):2511-2528. https://doi.org/10.1631/FITEE.2500100

Mind the Gap: towards generalizable autonomous penetration testing via domain randomization and meta-reinforcement learning

Key words: Cybersecurity; Penetration testing; Reinforcement learning; Domain randomization; Meta-reinforcement learning; Large language model

Corresponding author: Jingju LIU, Yuliang LU, Yue Zhang
E-mail: liujingju17@nudt.edu.cn; luyuliang@nudt.edu.cn; zhangyue@nudt.edu.cn
ORCID: Jingju LIU, https://orcid.org/0009-0005-9506-6903; Yuliang LU, https://orcid.org/0000-0002-8502-9907; Yue ZHANG, https://orcid.org/0009-0007-3570-2132

Research challenges

- **Training environment dilemma:** the conflict between the realism of the RL training environment and the efficiency of the training process.
- **Generalization gap and reality gap:** (1) Training environments have limited diversity, whereas real-world environments are unknown for agents and unpredictably diverse. The differences between them are known as the reality gap. (2) Agents' learned policies often exhibit poor performance when transferred to unseen testing (real-world) scenarios.

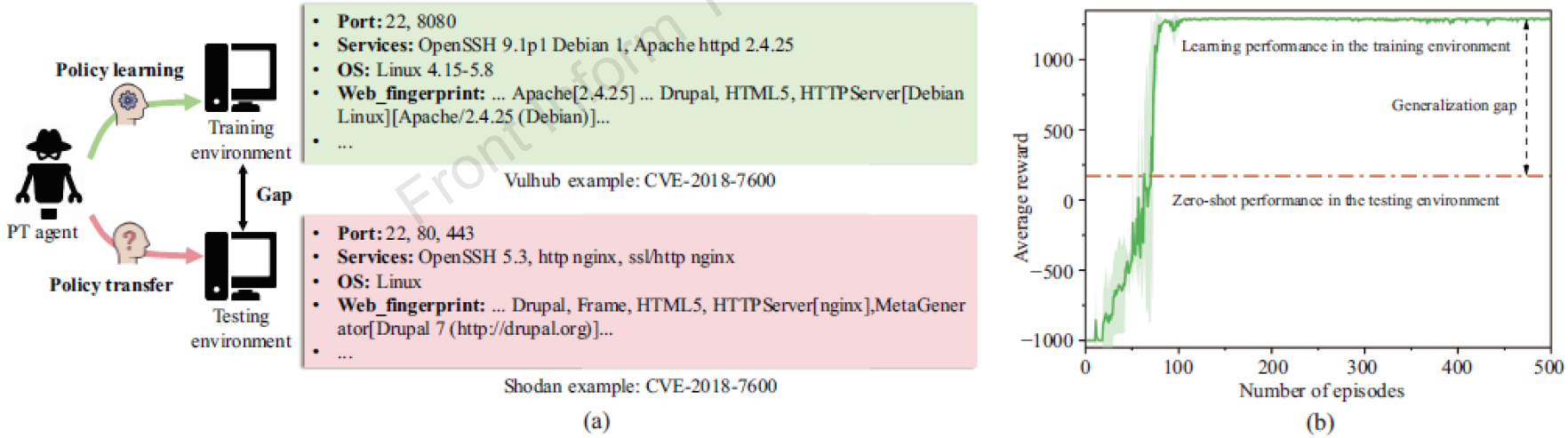

Fig. 1 Gaps between the training and testing environments: (a) reality gap where hosts with the same vulnerability have different configurations; (b) generalization gap that shows the learning curve of APRIL in the training environment and its zero-shot performance in the testing environment

Main contributions

- We propose GAP, a generalizable autonomous pentesting framework that works on a real-to-sim-to-real pipeline. This framework enables end-to-end policy learning in unknown real environments as well as the construction of realistic simulations, while improving agents' generalization ability.
- To bridge the generalization gap and achieve fast policy adaptation, we are among the first to apply domain randomization in the autonomous pentesting domain and propose an LLM-powered domain randomization method for environment augmentation. We further apply meta-RL to improve the agents' generalization ability to unseen environments by leveraging the generated simulations.
- We conduct simulations on various vulnerable virtual machines, with results showing that GAP can enable policy learning in various realistic environments, achieve zero-shot policy transfer in similar environments, and achieve rapid policy adaptation in dissimilar environments.

Overview of GAP

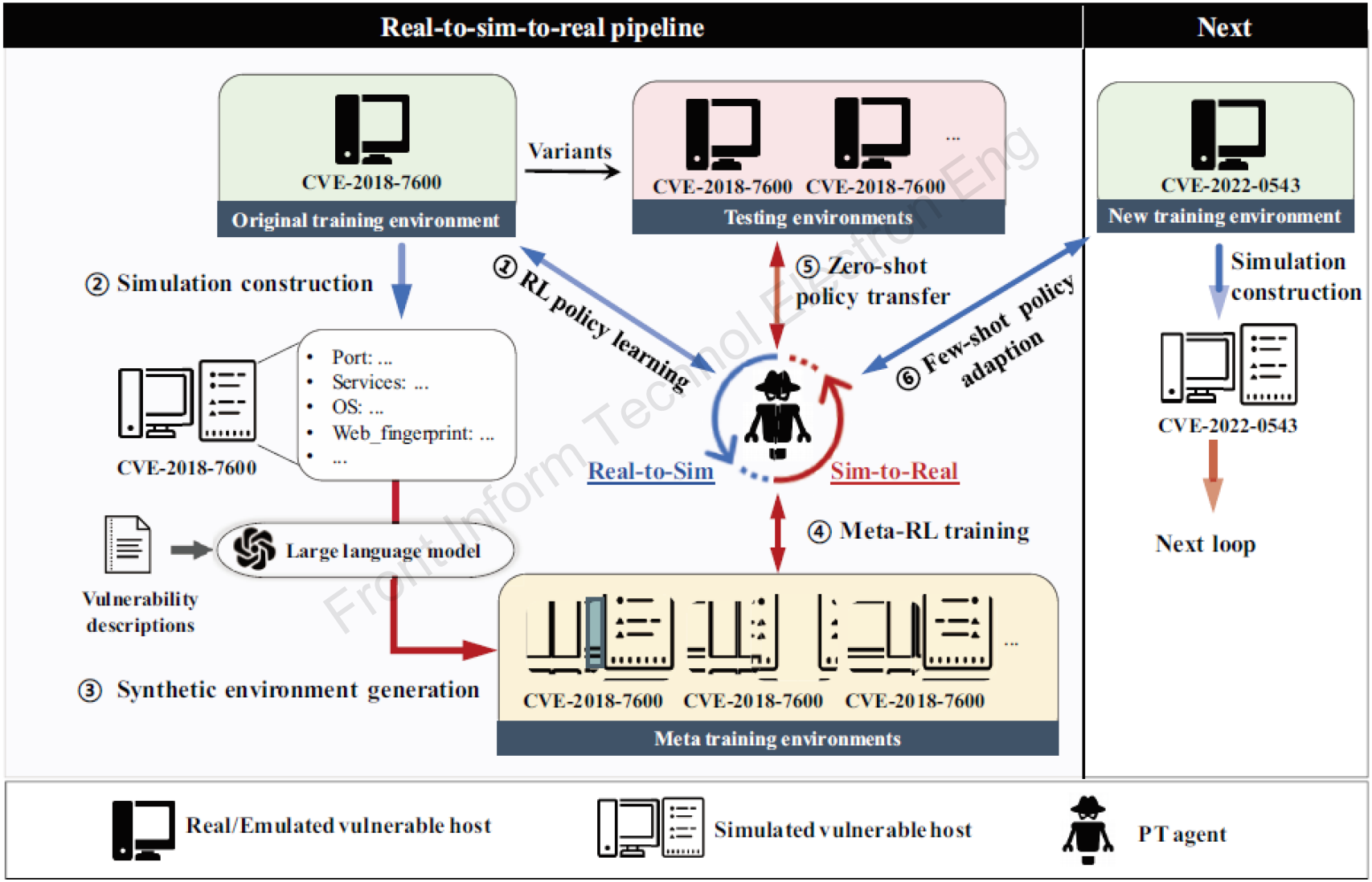

Fig. 2 Overview of GAP. References to color refer to the online version of this figure

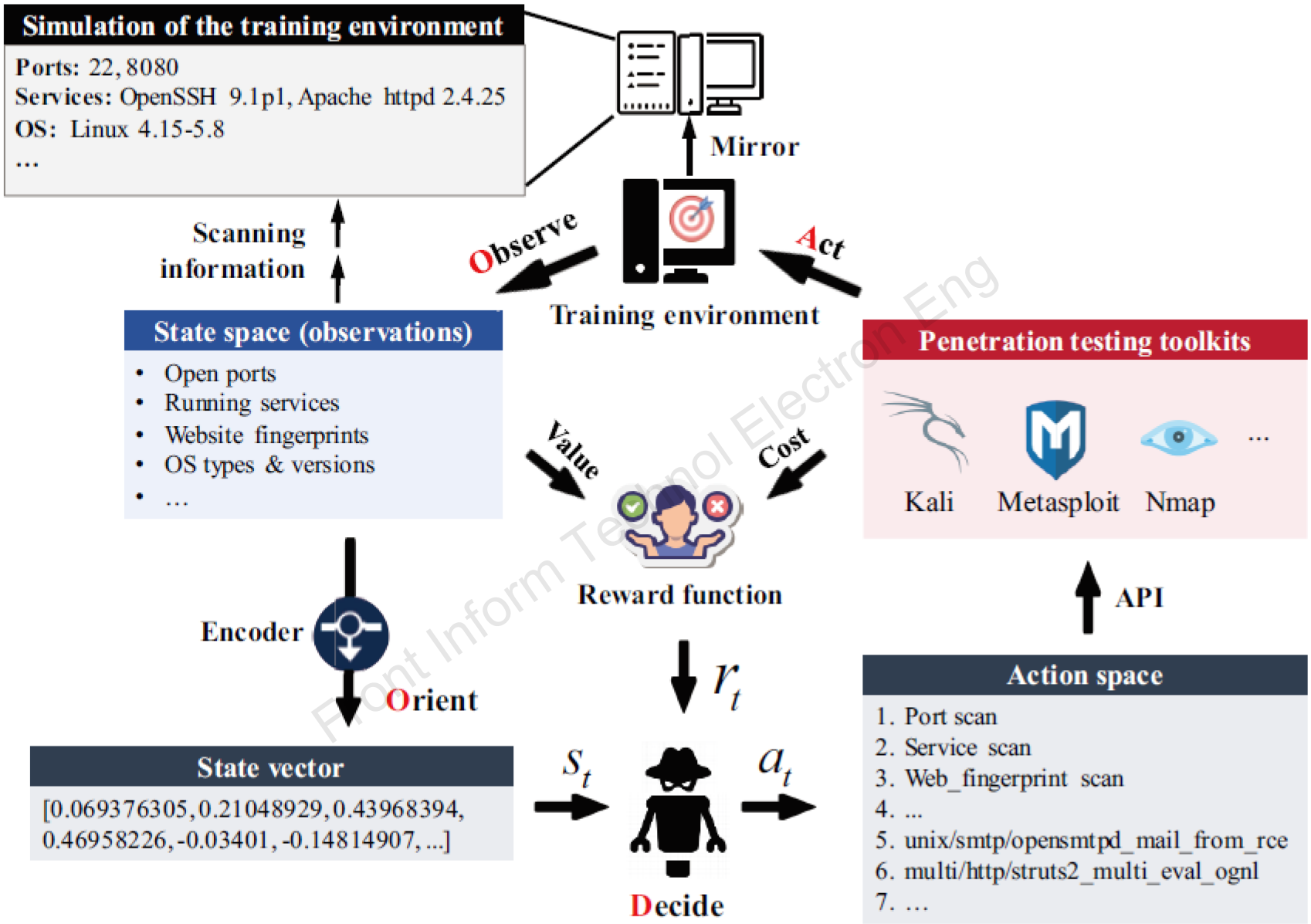

Fig. 3 The process of policy learning and construction of simulated environments

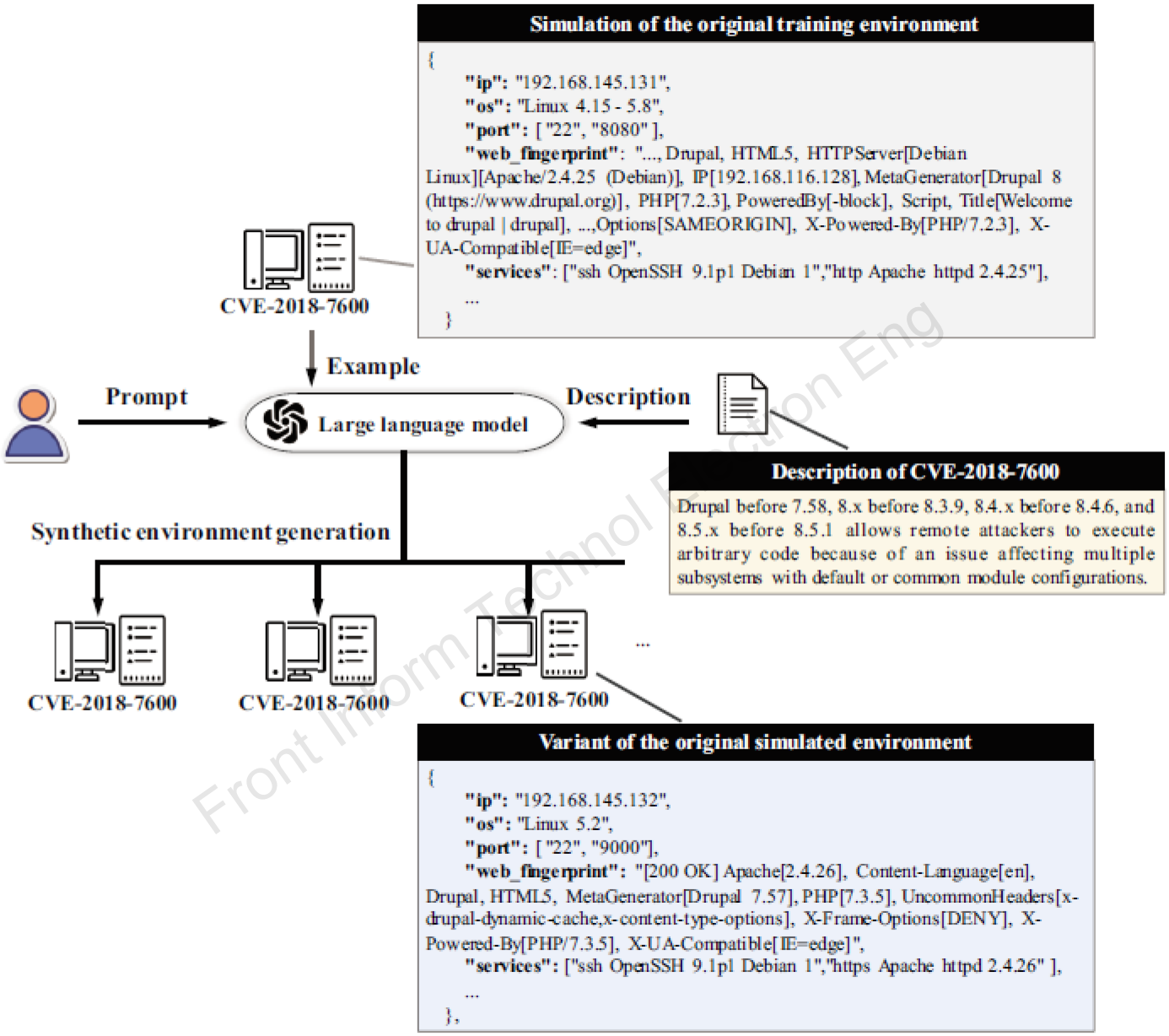

Fig. 5 The workflow and example of synthetic environment generation using a large language model. In this example, we construct the simulation by using the CVE-2018-7600 vulnerable host from Vulhub as the original training environment. Then, we use GLM-4 (Team GLM, 2024) for domain randomization

Results

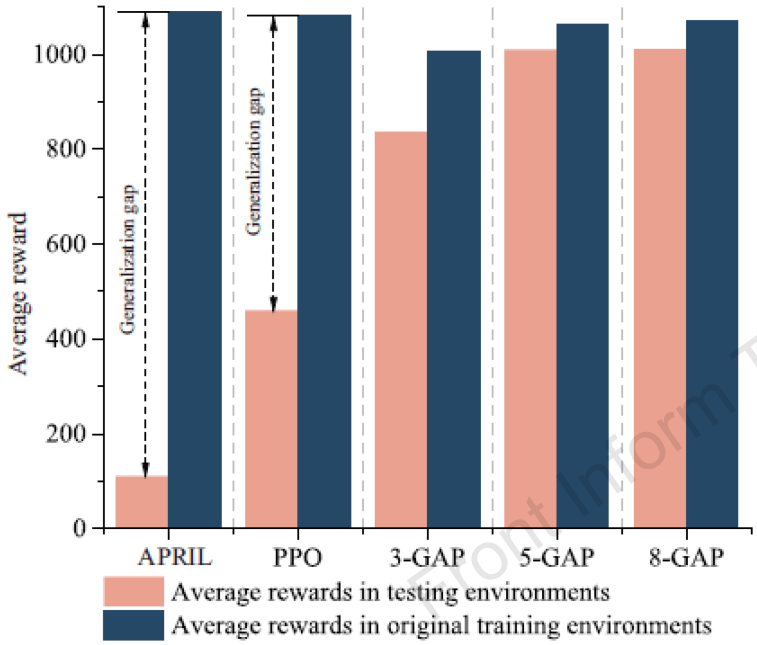

Fig. 8 Zero-shot generalization performance in the original training environments and testing environments

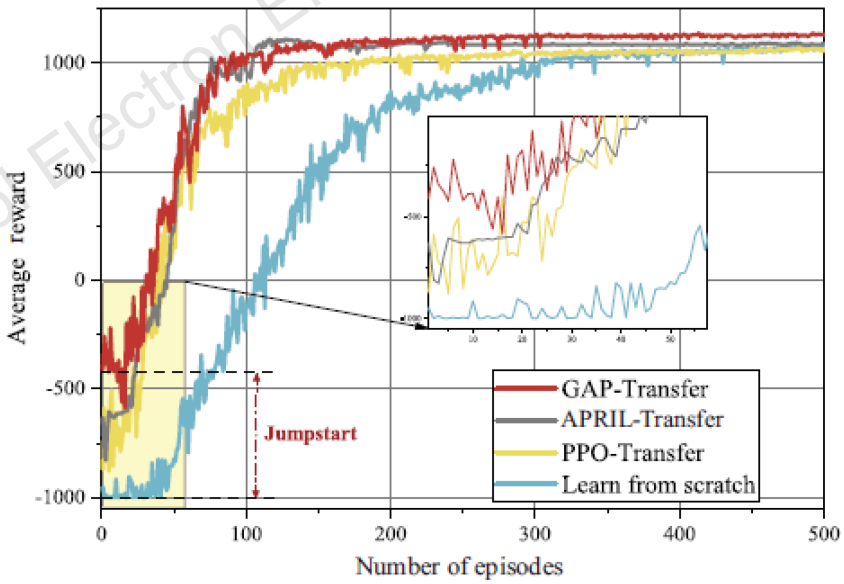

Fig. 9 Learning curves of different methods in testing environments. References to color refer to the online version of this figure

Conclusions

In this paper, we present GAP, an autonomous pentesting framework for efficient policy training in realistic environments and for training generalizable agents capable of drawing inferences about other cases from one instance—a key to the broad application of autonomous pentesting agents. To achieve this, GAP introduces a real-to-sim-to-real pipeline that (1) enables end-to-end policy learning in unknown realistic environments while constructing realistic simulation analogs and (2) improves agents' generalization ability by leveraging domain randomization and meta-RL learning. The preliminary evaluations demonstrate that GAP allows pentesting agents for end-to-end policy learning in realistic environments, bridging the generalization gap for zero-shot policy transfer in similar environments, and facilitating rapid policy adaptation in dissimilar environments.