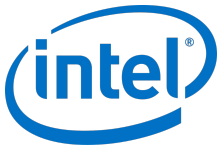
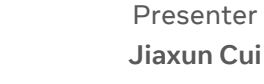
ICLR
International Conference On Learning Representations

MACTA: A Multi-agent Reinforcement Learning Approach for Cache Timing Attacks and Detection

Jiaxun Cui, Xiaomeng Yang*, Mulong Luo*, Geunbae Lee*, Peter Stone, Hsien-Hsin S. Lee, Benjamin Lee, G. Edward Suh, Wenjie Xiong^, Yuandong Tian^

* Equal second author contribution, ^ Equal supervising





Presenter
Jiaxun Cui
The University of Texas at Austin

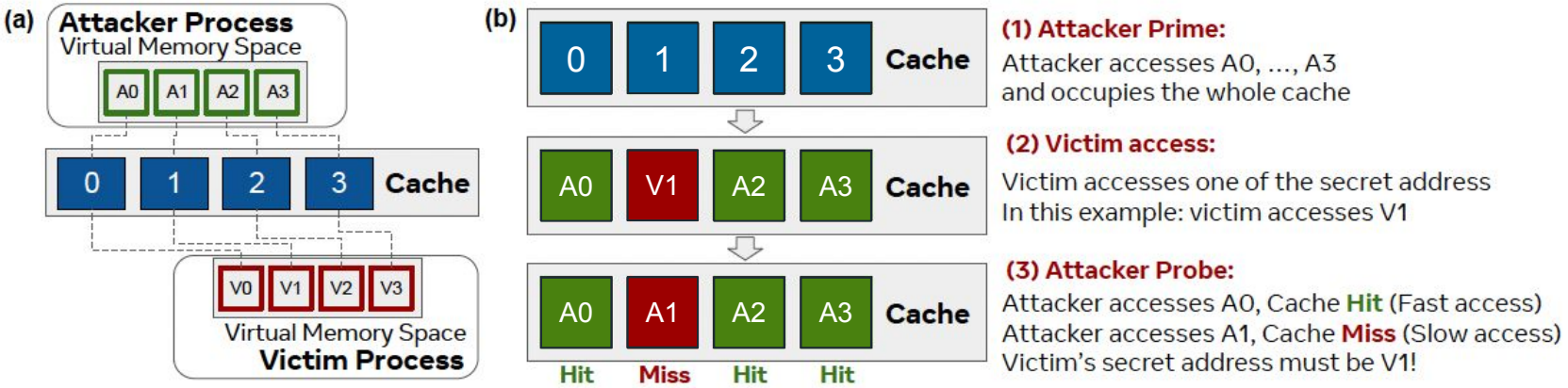
00 Overview

Overview

1. A multi-agent environment, **MA-AutoCAT**
2. A training framework, **MACTA**
3. A **generalizable** and **robust detector** that leverages **Transformer encoder** for **Cache Timing Attacks**

01 MOTIVATION

What is a Cache Timing Attack?



Reinforcement learning attackers can exploit the timing channel and steal information faster than human heuristic attacks[1].

[1] Luo, M., Xiong, W., Lee, G., Li, Y., Yang, X., Zhang, A., ... & Suh, G. E. (2023, February). Autocat: Reinforcement learning for automated exploration of cache-timing attacks. In 2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA) (pp. 317-332)..

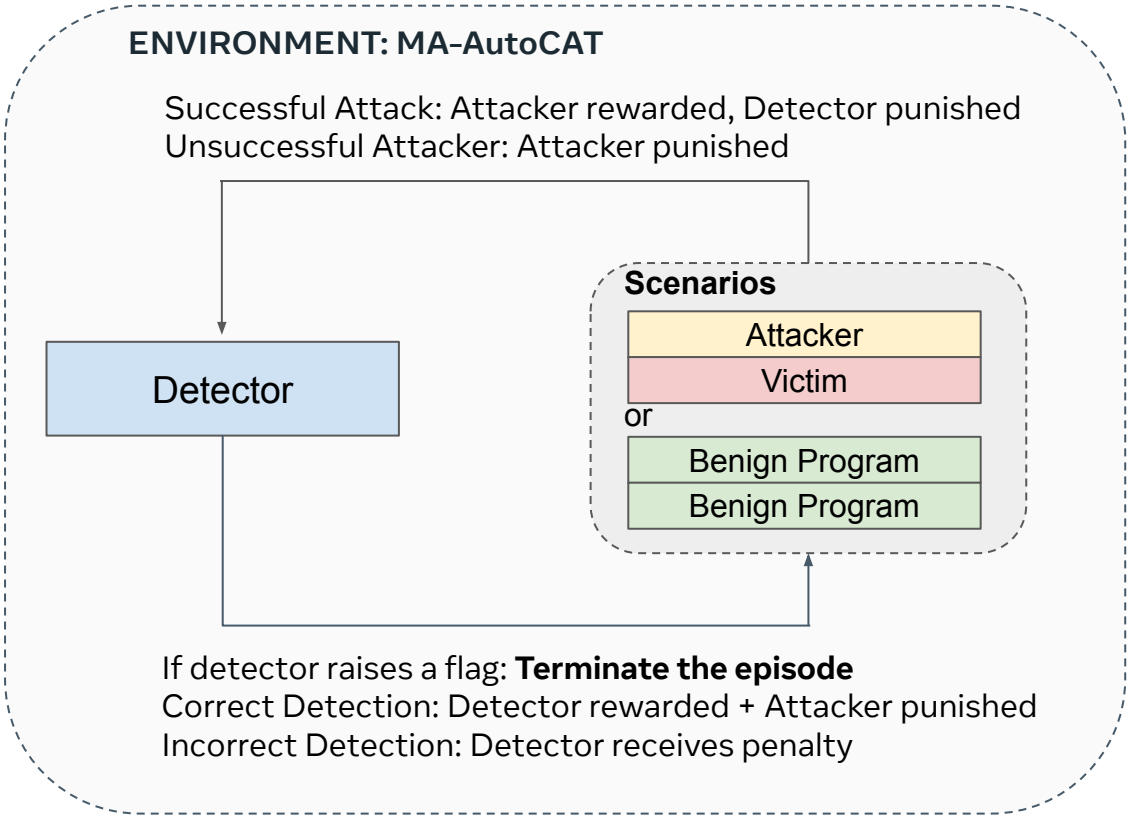
Meta

Why should we take a multi-agent perspective?

- **Existing Detectors are still potentially vulnerable to future attacks**
  - Detectors rely on strong domain knowledge and discovered attacker tactics
  - Attackers can evade detection by modifying behaviors

- **General-sum Markov Games**
  - Enable automatic discovery of both attacker and defender policies
  - Have a pool of diverse opponent strategies to develop robust policies

02 ENVIRONMENT



Fixed episode length: Max Step=64

Detector (D)
To find out whether there is an attacker as soon as possible

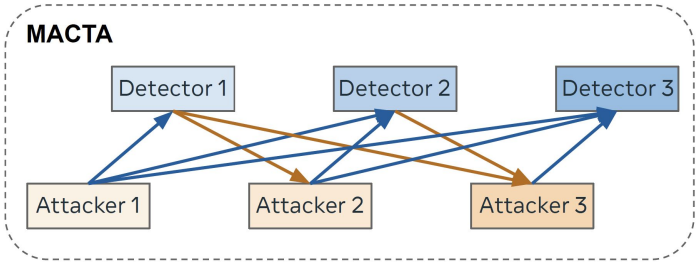
Attacker (A)
To attack (guess victim's secret) as many times as it can before the detector finds out

Victim (V)
It has a multi-bit secret address, which is the target of the attacker

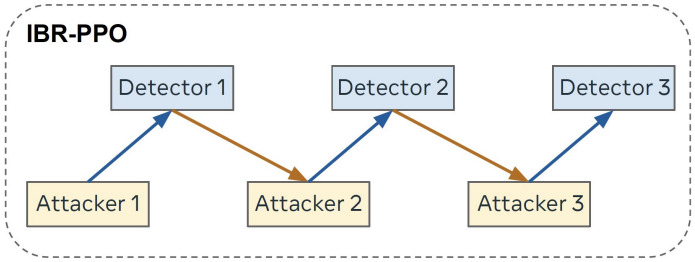
Benign Program (B)
It has action sequences sampled from normal programs in the SPEC-2017 dataset, and the actions are projected to the valid attacker action space of a specific cache configuration

03 METHOD

MACTA Key Concepts



P → Q Q learns policy against P using PPO



Iterated Best Response training could result in Cyclic Policies

$$\Pi^i_{\tau+1} \leftarrow \Pi^i_\tau \cup \{\pi^i_*(\mathbb{U}(\Pi^{-i}_\tau))\}$$

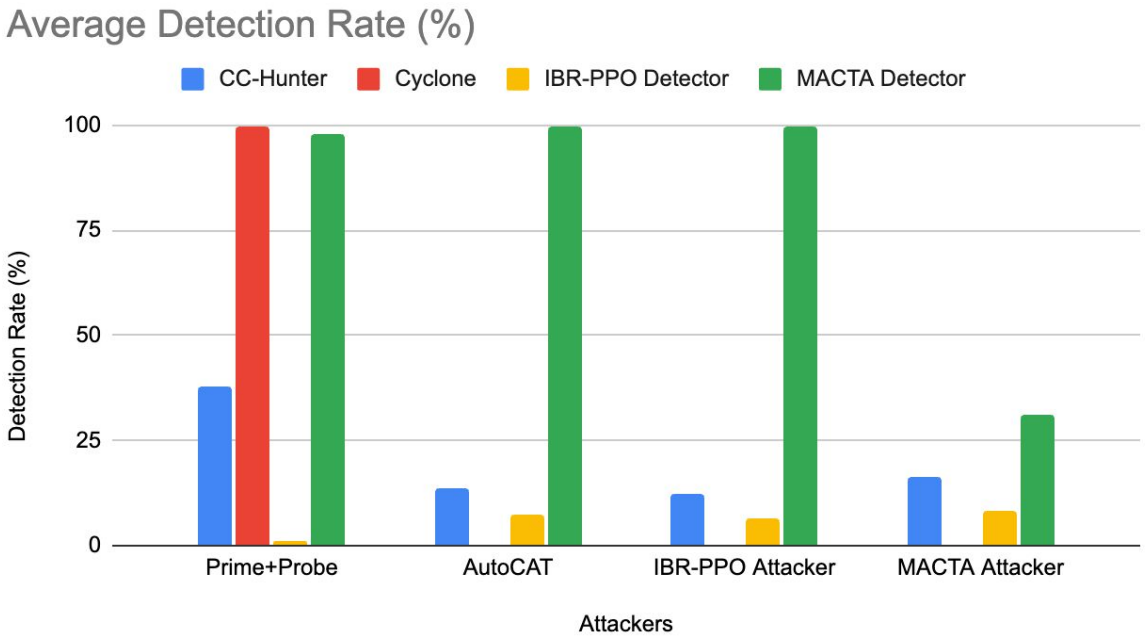
1. **Transformer** observation encoder
2. Maintain a policy **pool** for each agent and increase the pool size with policy checkpoints during training
3. Approximate Best Responses to a **uniform mixture** of opponents using (Dual-Clip) Proximal Policy Optimization (PPO) [2] [3]

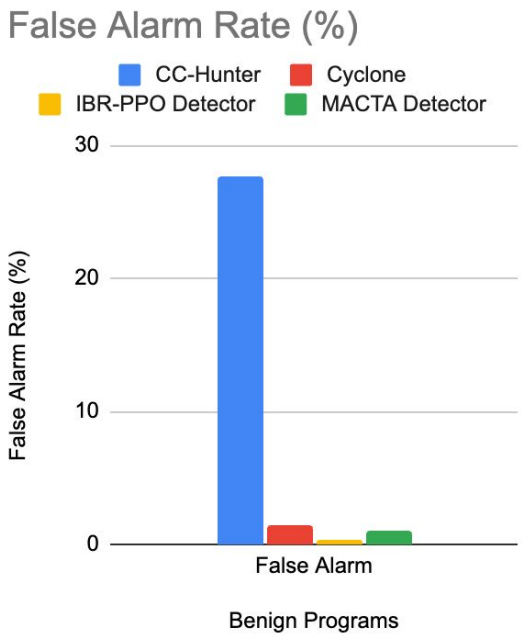
[2] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
[3] Ye, D., Liu, Z., Sun, M., Shi, B., Zhao, P., Wu, H., ... & Huang, L. (2020, April). Mastering complex control in moba games with deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 6672-6679).

04 Results

Detector Evaluation

Average over 10000 episodes and 10 test datasets



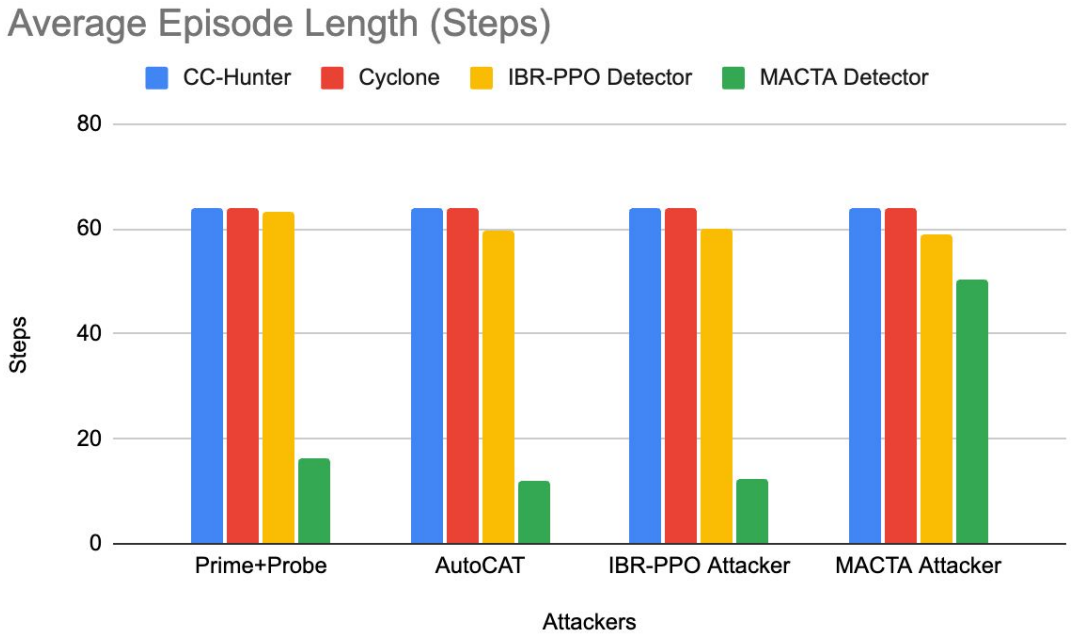


MACTA detector is able to outperform baseline detectors and **generalize** to **unseen attackers** while maintaining **low false alarm rate** for benign programs.

Meta

Detector Evaluation

Average over 10000 episodes and 10 test datasets

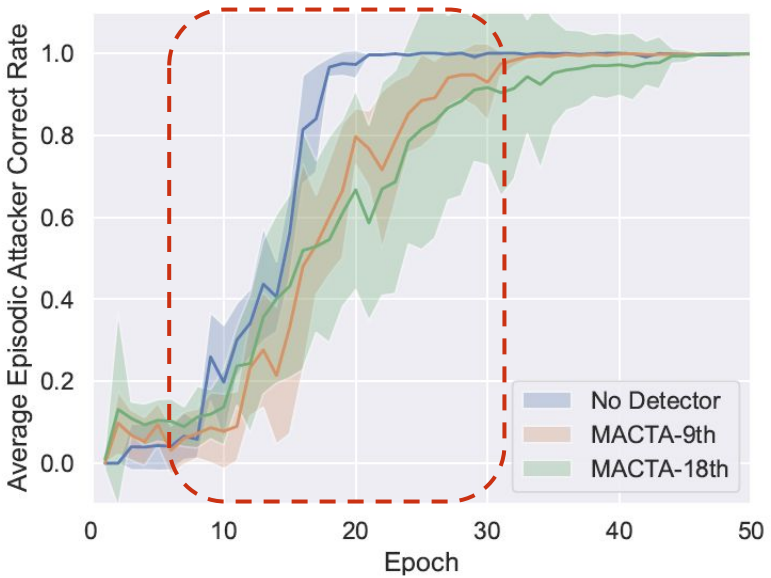


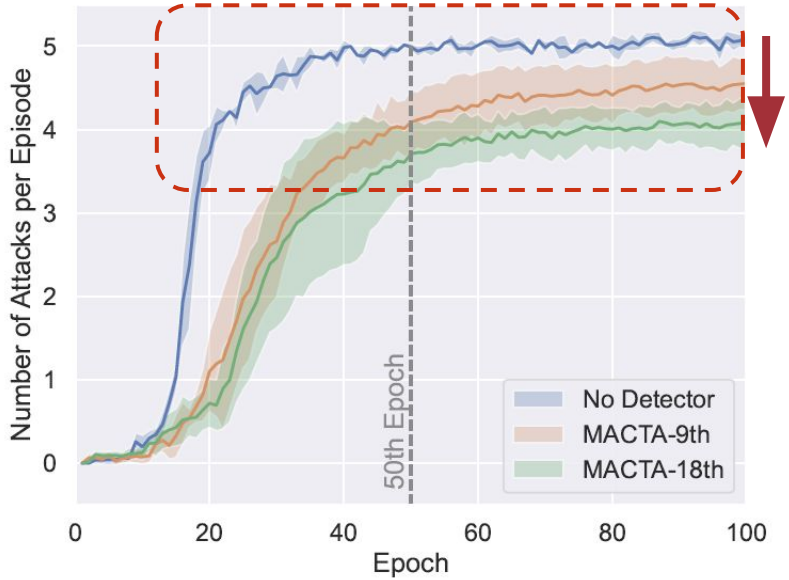
MACTA terminates attackers early to prevent further information leakage

Meta

Detector “Exploitability”

Load a pre-trained detector, train an “exploiter” attacker of it from scratch





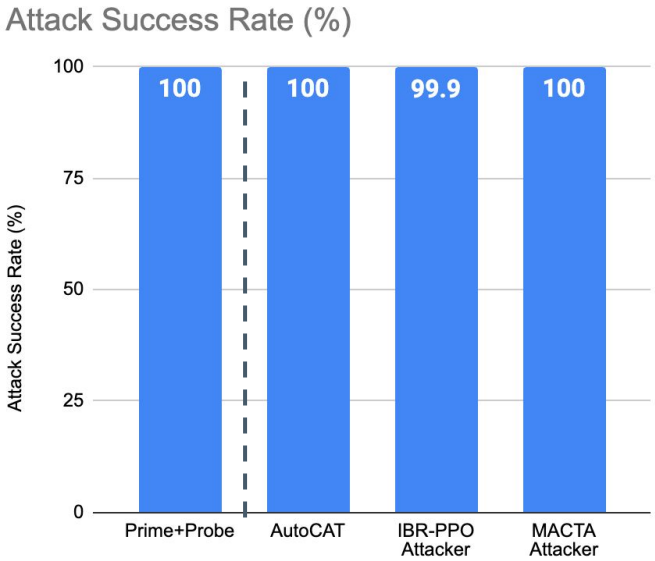
1. Slow down future adaptive attackers’ learning speed
2. Reduces the information leakage against adaptive attackers

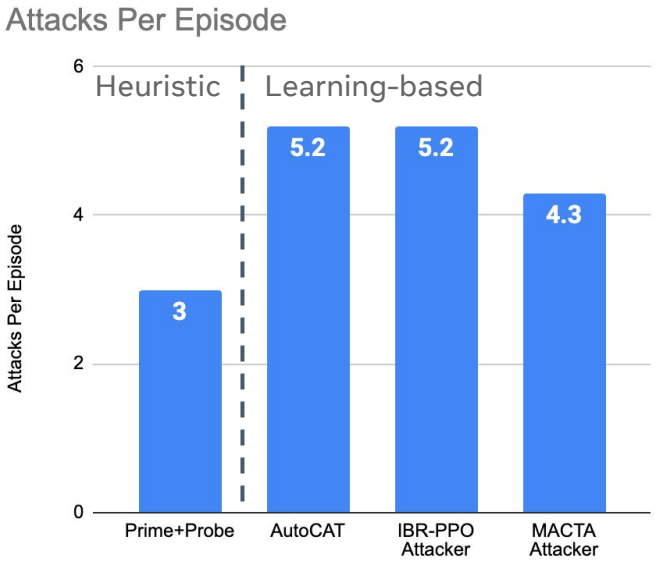
Jiaxun Cui, The University of Texas at Austin

Meta

Attacker Evaluation

Average over 10000 episodes

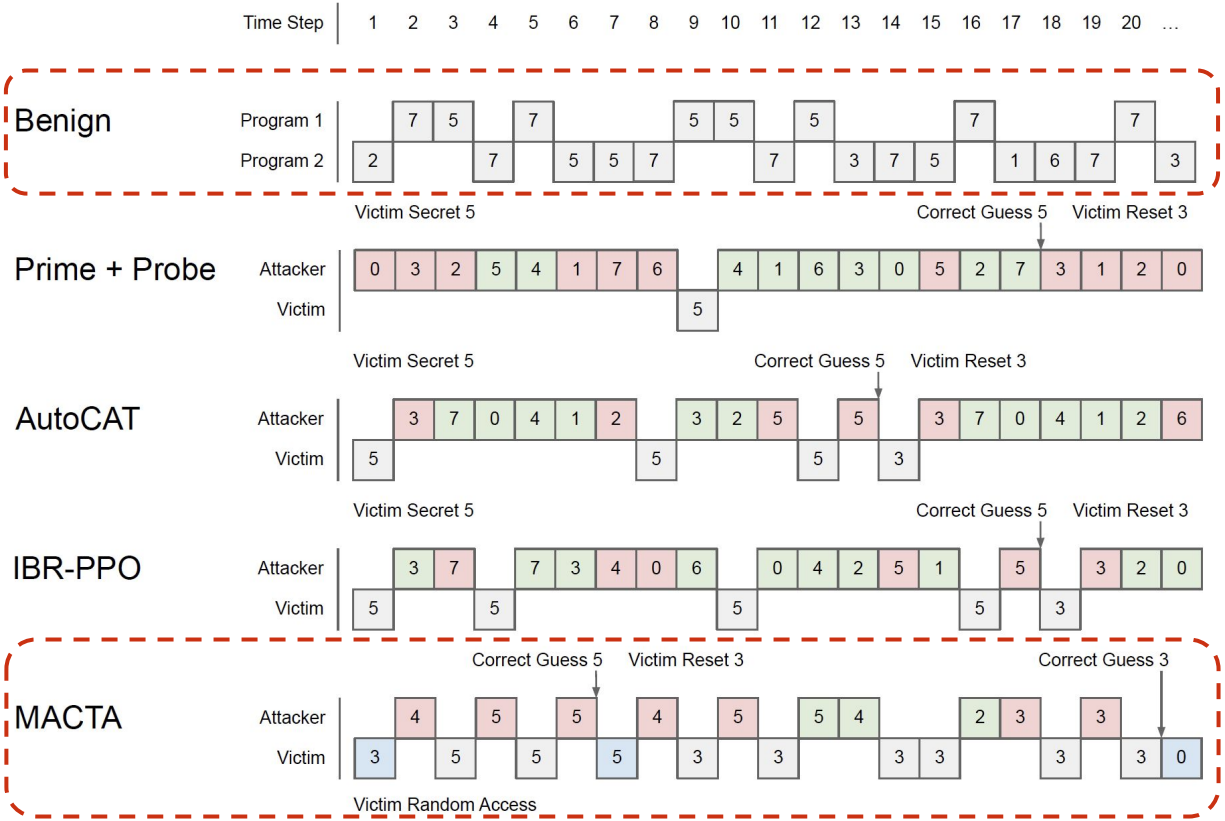




1. All of the attackers are conducting effective attacks that are transferable to real hardwares.
2. MACTA attacker has the fewest attacks per episode among learning-based attackers.

Meta

Qualitative Evaluation



3 Cache Miss Observed by the Attacker

3 Cache Hit Observed by the Attacker

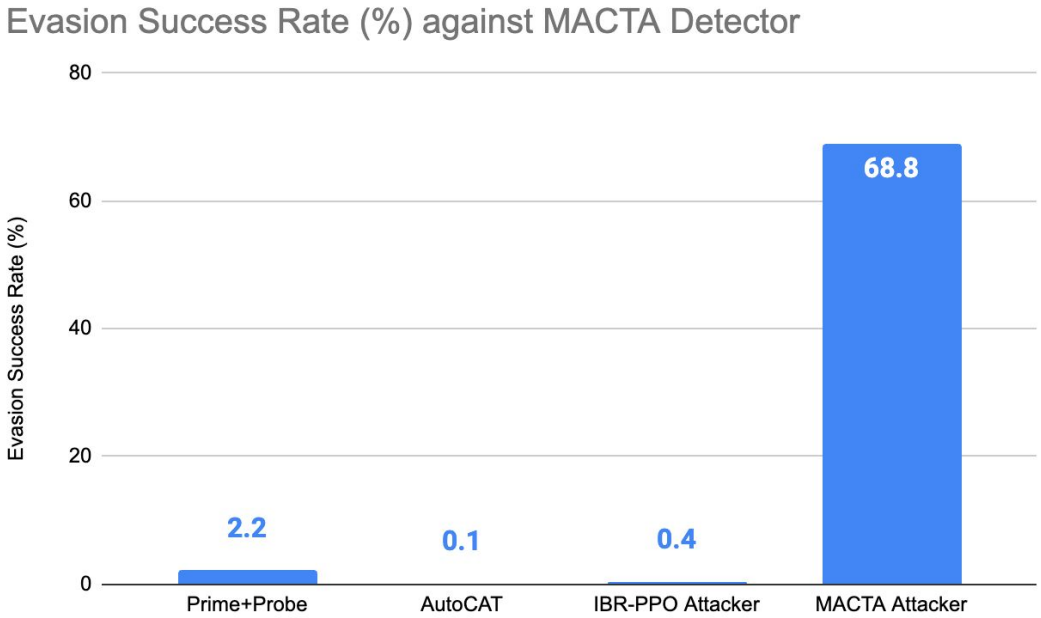
3 Cache Set Access Not Observable to the Attacker

3 Victim Random Access Not Observable to the Attacker

Meta

Attacker Evaluation

Average over 10000 episodes and 10 test datasets



MACTA attacker can evade the strongest detector with highest success rates

Jiaxun Cui, The University of Texas at Austin

Meta

ICLR
International Conference On Learning Representations

MACTA: A Multi-agent Reinforcement Learning Approach for Cache Timing Attacks and Detection

Jiaxun Cui, Xiaomeng Yang*, Mulong Luo*, Geunbae Lee*, Peter Stone, Hsien-Hsin S. Lee, Benjamin Lee, G. Edward Suh, Wenjie Xiong^, Yuandong Tian^

Paper

Code

Meta