Pythia: Remote Oracles for the Masses

Shin-Yeh Tsai, Mathias Payer, Yiying Zhang

WukLab

PURDUE UNIVERSITY

EPFL

UC San Diego

Datacenter Security Is Important

TECHNOLOGY NEWS JULY 30, 2019 / 1:10 PM / 14 DAYS AGO

Akamai raises forecast on cloud security, content delivery demand

OCTOBER 28, 2015

Researchers show how side-channel attacks can be used to steal encryption keys on Amazon's cloud servers

by Worcester Polytechnic Institute

Cybersecurity

Microsoft Wants More Security Researchers to Hack Into Its Cloud

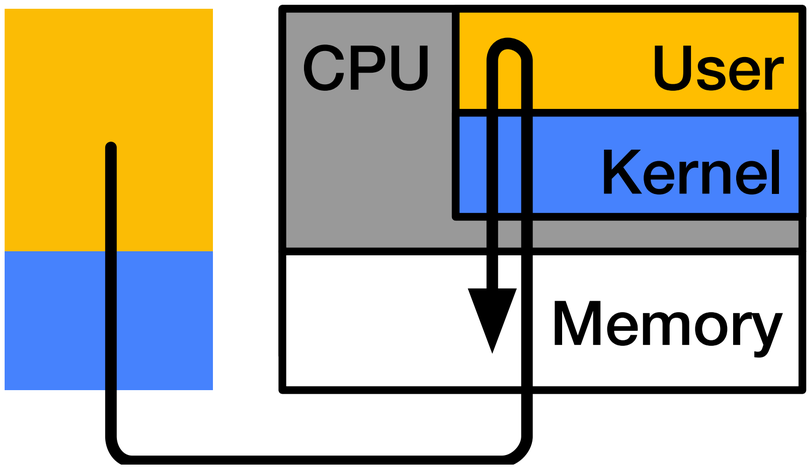
NEW YORK TIMES CYBER-ATTACK REMINDS DATA CENTERS OF SECURITY IMPORTANCE

Datacenter Needs and Trends

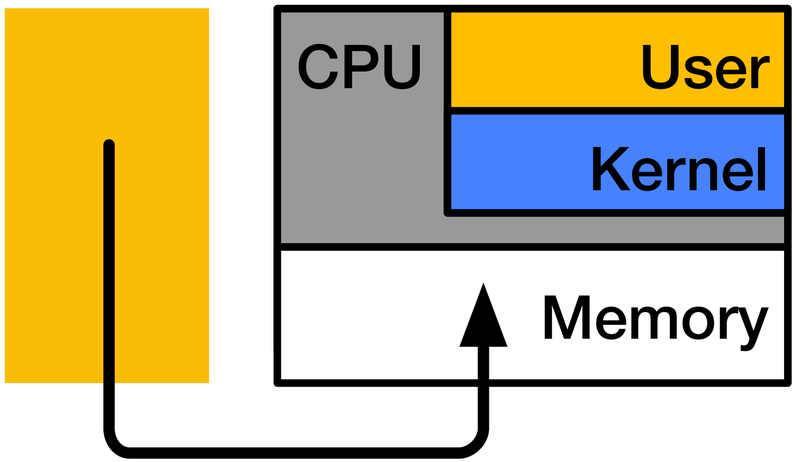
- **Faster network communication to support distributed applications**
- **Large memory to store application data**
- **Low CPU utilization for lower energy cost**

RDMA Fits Datacenter Needs

(Remote Direct Memory Access)

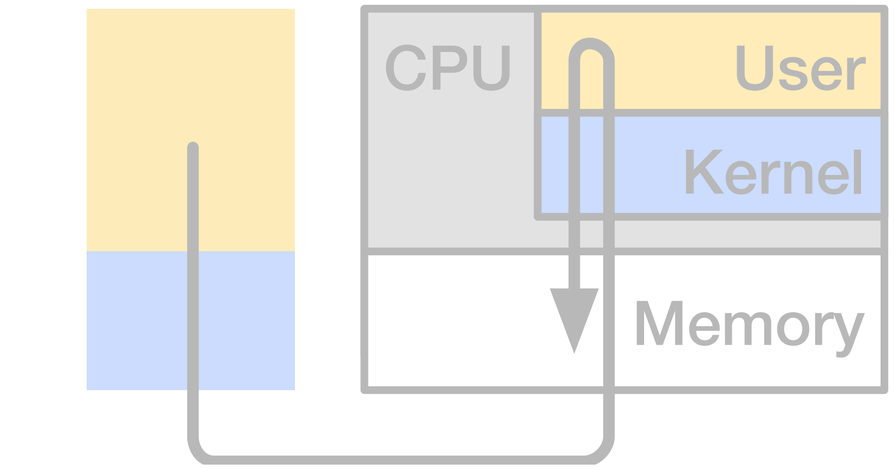


Classic Communication

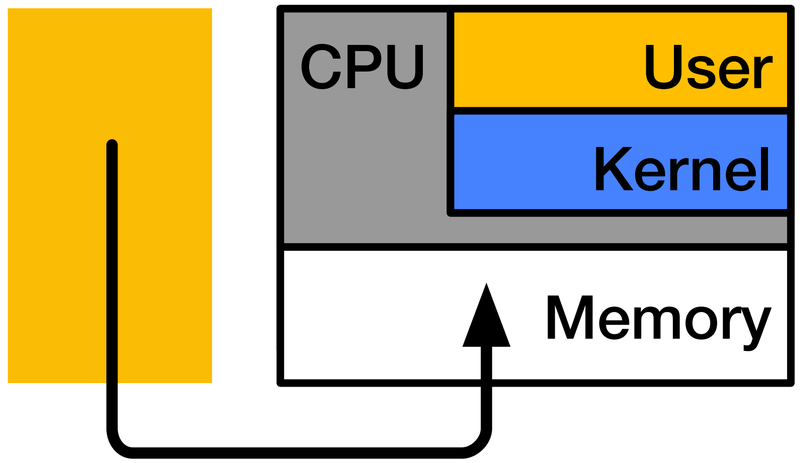


RDMA

RDMA Fits Datacenter Needs
(Remote Direct Memory Access)



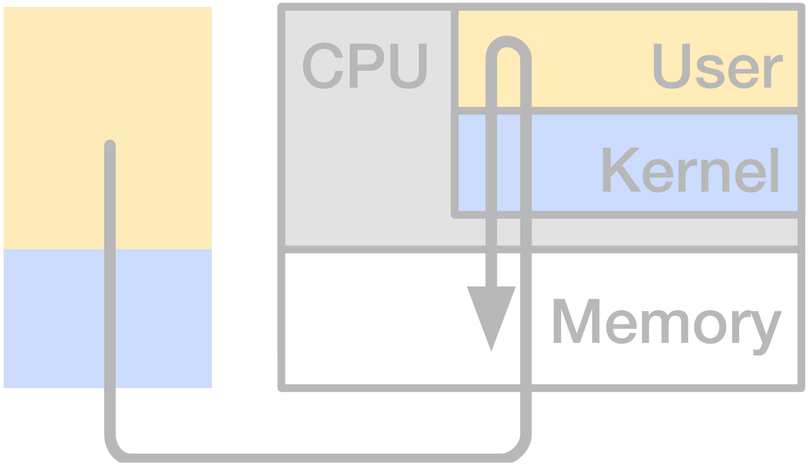
Classic Communication



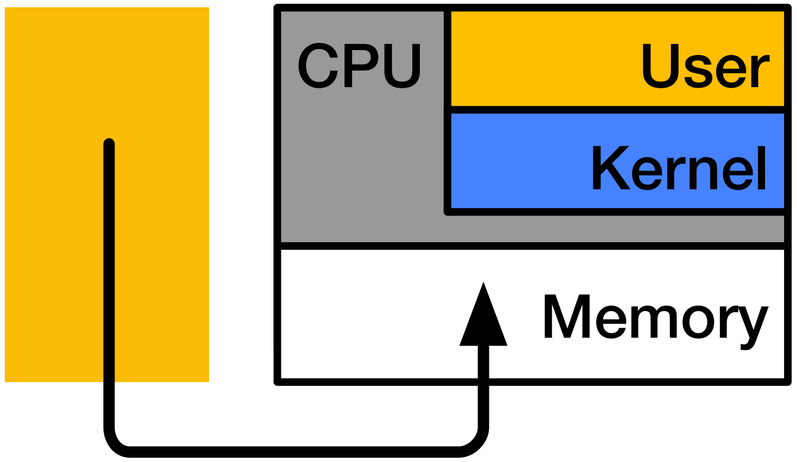
RDMA

- Read/write remote memory
- Bypassing kernel
- Memory zero copy

RDMA Fits Datacenter Needs
(Remote Direct Memory Access)



Classic Communication



RDMA

- Read/write remote memory
- Bypassing kernel
- Memory zero copy

– Low latency
– High throughput
– Low CPU utilization

Availability of Linux RDMA on Microsoft Azure

Posted on July 9, 2015

Tejas Karmarkar, Princ

We are excited to announce
in our cloud journey and in
set of users. Linux RDMA m

Microsoft Azure

Availability of Linux RDMA on Microsoft Azure

Posted on July 9, 2015

Tejas Karmarkar, Princ

We are excited to announce
in our cloud journey and in
set of users. Linux RDMA m

Microsoft Azure

Alibaba Cloud's Support for RDMA

Alibaba Cloud supports Super Com
dedicated to RDMA communication
intelligence, machine learning, scie
processing, and other scenarios.

Alibaba Cloud

RDMA-Based Datacenter Applications

RDMA-Based Datacenter Applications

Pilaf
[ATC ’13]

APUS
[SoCC ’17]

Hotpot
[SoCC ’17]

FaRM
[NSDI ’14]

Octopus
[ATC ’17]

Orion
[FAST ’19]

Wukong
[OSDI ’16]

HERD
[SIGCOMM ’14]

Storm
[SYSTOR '19]

NAM-DB
[VLDB ’17]

DRTM+H
[OSDI ’18]

Cell
[ATC ’16]

DrTM
[SOSP ’15]

LITE
[SOSP ’17]

KV-Direct
[SOSP ’17]

FaSST
[OSDI ’16]

FaRM + Xact
[SOSP ’15]

Mojim
[ASPLOS ’15]

RSI
[VLDB ’16]

DrTM+R
[EuroSys ’16]

RDMA-Based Datacenter Applications

Performance

FaRM
[NSDI '14]

Octopus
[ATC '17]

Orion
[FAST '19]

Wukong
[OSDI '16]

HERD
[SIGCOMM '14]

Storm
[SYSTOR '19]

NAM-DB
[VLDB '17]

DRTM+H
[OSDI '18]

Cell
[ATC '16]

DrTM
[SOSP '15]

LITE
[SOSP '17]

KV-Direct
[SOSP '17]

FaSST
[OSDI '16]

FaRM + Xact
[SOSP '15]

RSI
[VLDB '16]

Mojim
[ASPLOS '15]

DrTM+R
[EuroSys '16]

RDMA-Based Datacenter Applications

Performance

Scalability

FaRM
[NSDI ’14]

Octopus
[ATC ’17]

HERD
[SIGCOMM ’14]

NAM-DB
[VLDB ’17]

Storm
[SYSTOR '19]

DrTM
[SOSP ’15]

DRTM+H
[OSDI ’18]

Cell
[ATC ’16]

LITE
[SOSP ’17]

KV-Direct
[SOSP ’17]

FaRM + Xact
[SOSP ’15]

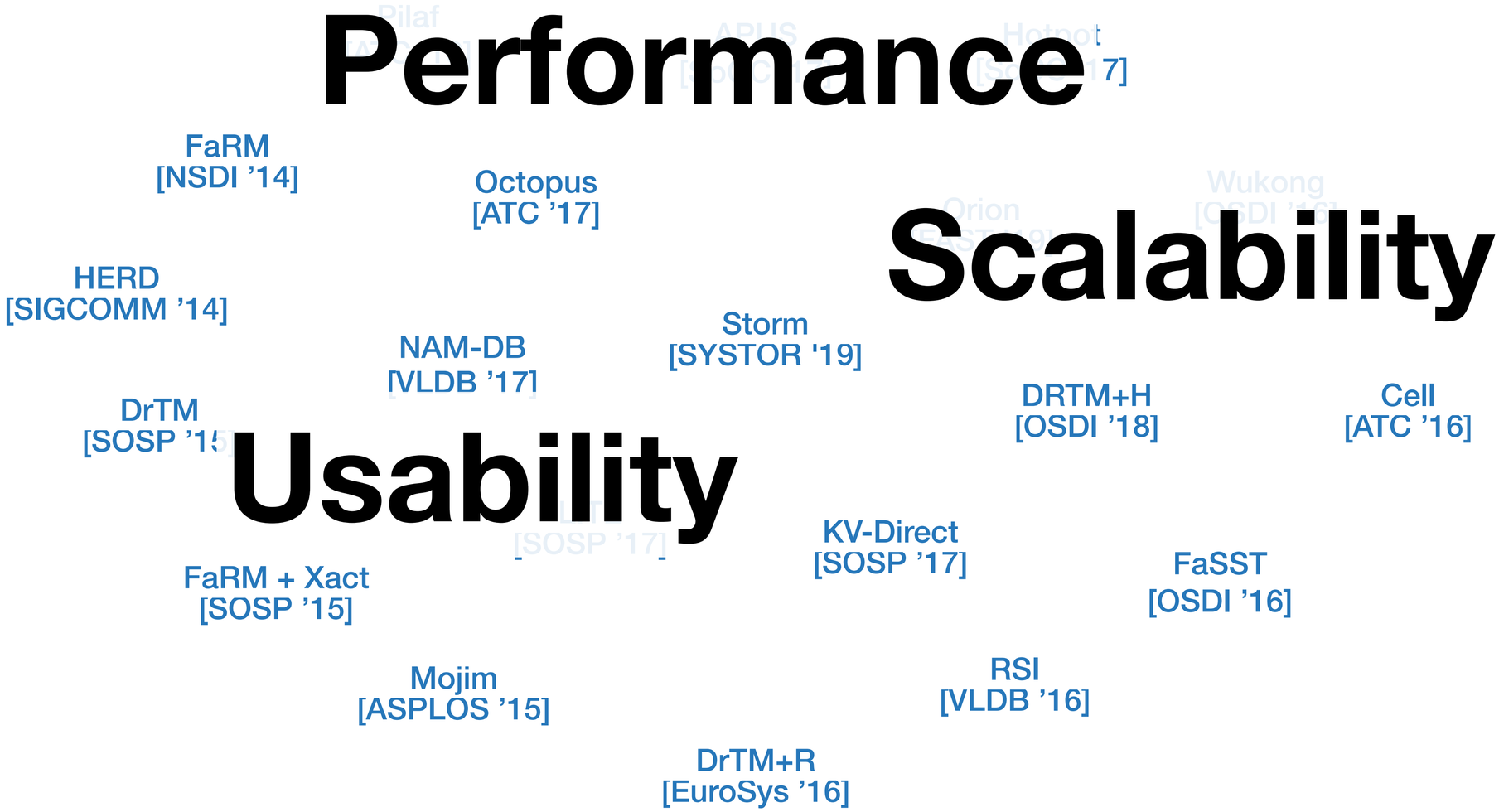
FaSST
[OSDI ’16]

Mojim
[ASPLOS ’15]

RSI
[VLDB ’16]

DrTM+R
[EuroSys ’16]

RDMA-Based Datacenter Applications



RDMA-Based Datacenter Applications

Performance

FaRM
[NSDI '14]

Octopus
[ATC '17]

Scalability

HERD
[SIGCOMM '14]

Storm
[SYSTOR '19]

NAM-DB
[VLDB '17]

DRTM+H
[OSDI '18]

Cell
[ATC '16]

DrTM
[SOSP '15]

Usability

KV-Direct
[SOSP '17]

FaRM + Xact
[SOSP '15]

FaSST

What about Security?

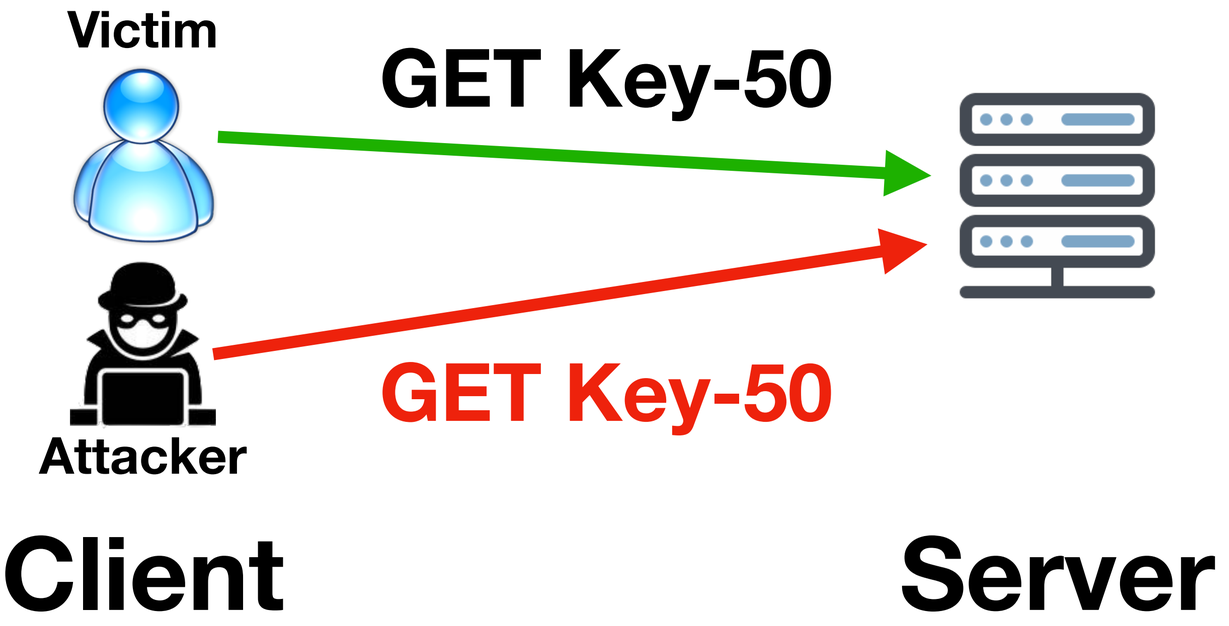
DrTM+R
[EuroSys '16]

Pythia

The First Side-Channel Attack on RDMA

Threat Model

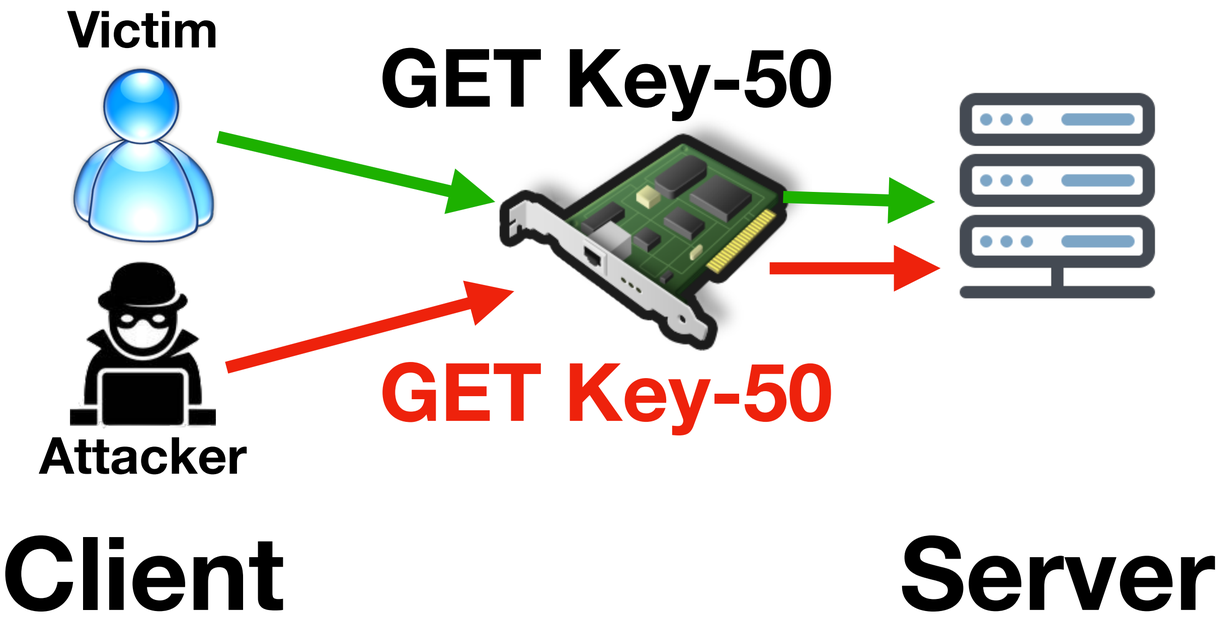
- Server hosts data in memory exported via RDMA
- The attacker and the victim can be on different machines
- The attacker wants to learn the access pattern of the victim
- The attacker cannot observe the network traffic



Did the victim access key-50? (Y/N)

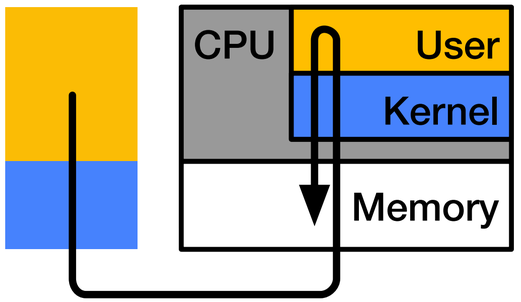
Threat Model

- Server hosts data in memory exported via RDMA
- The attacker and the victim can be on different machines
- The attacker wants to learn the access pattern of the victim
- The attacker cannot observe the network traffic

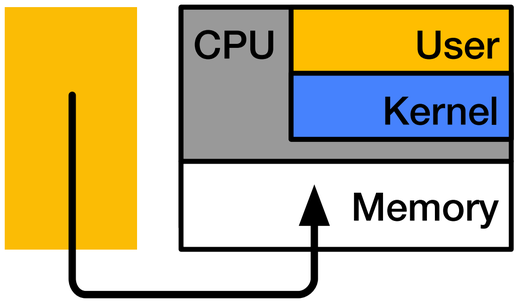


Did the victim access key-50? (Y/N)

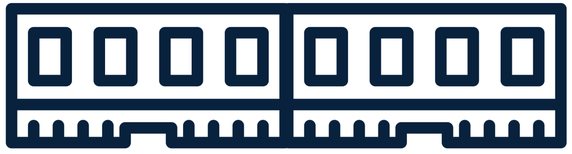
Traditional NIC and RDMA NIC



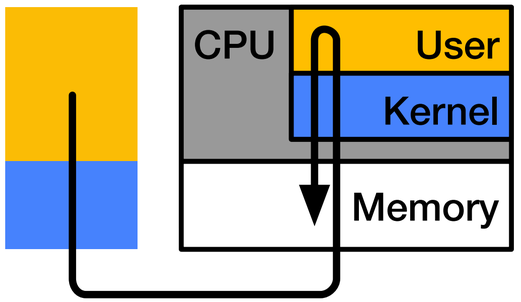
Classic Communication



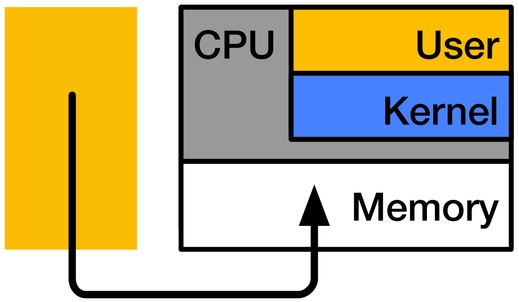
RDMA



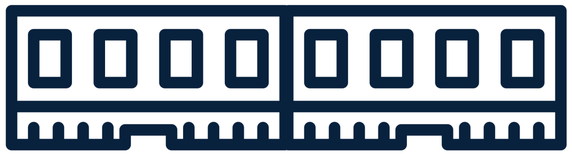
Traditional NIC and RDMA NIC



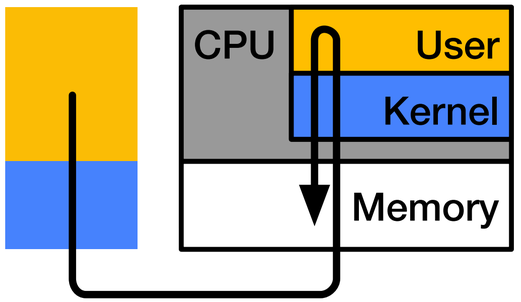
Classic Communication



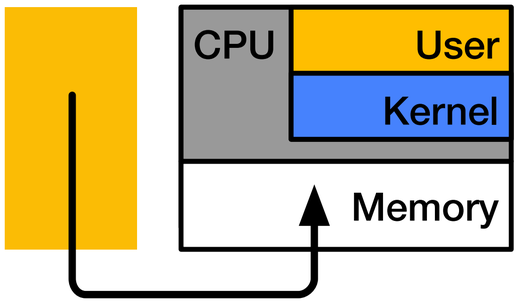
RDMA



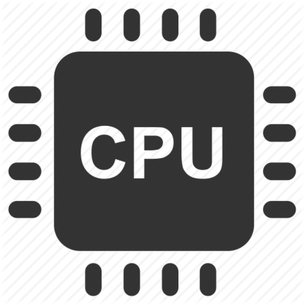
Traditional NIC and RDMA NIC

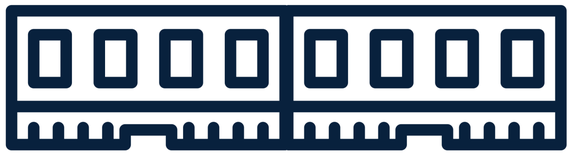


Classic Communication

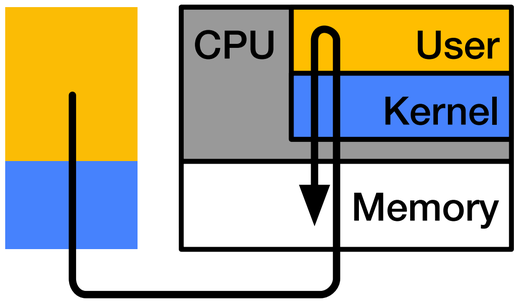


RDMA

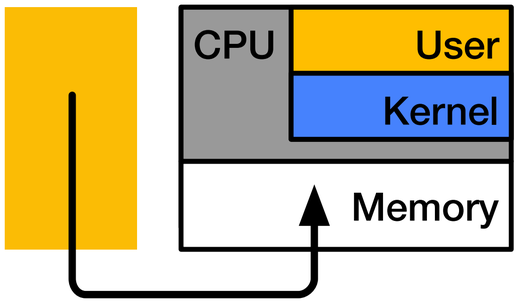




Traditional NIC and RDMA NIC

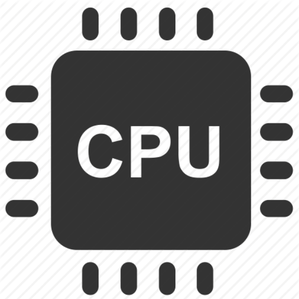


Classic Communication

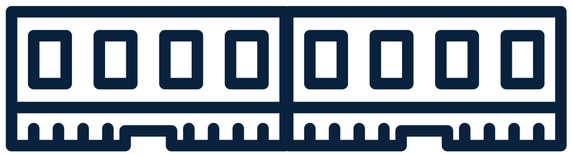


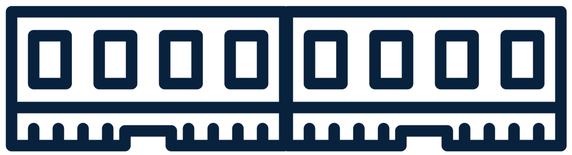
RDMA

Massive Overhead

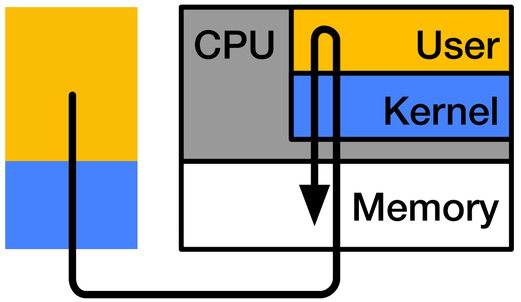


1. Address mapping
2. Permission checking
3. Resource isolation

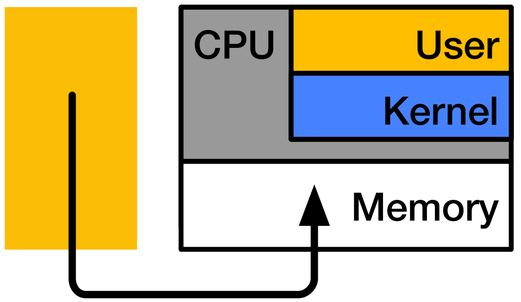




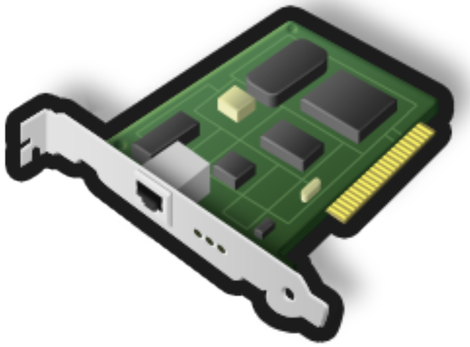
Traditional NIC and RDMA NIC



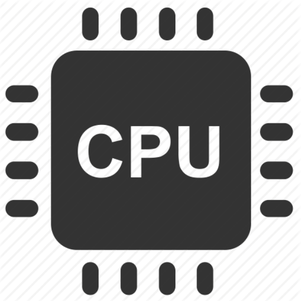
Classic Communication



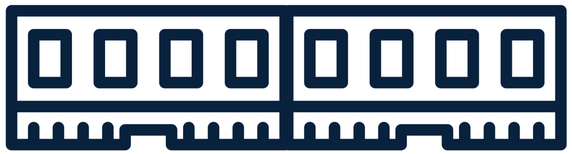
RDMA

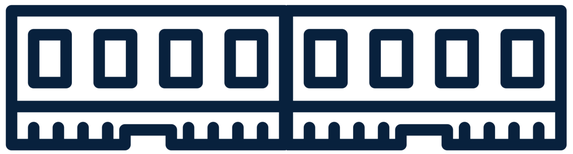


Massive Overhead

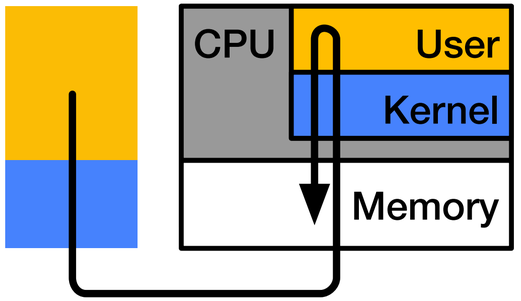


1. Address mapping
2. Permission checking
3. Resource isolation

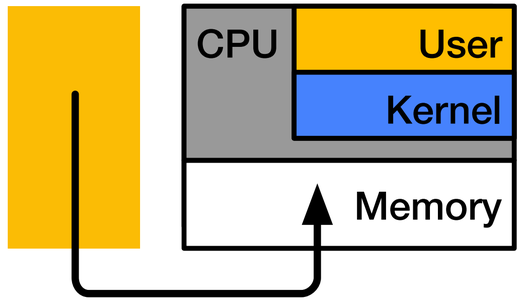




Traditional NIC and RDMA NIC

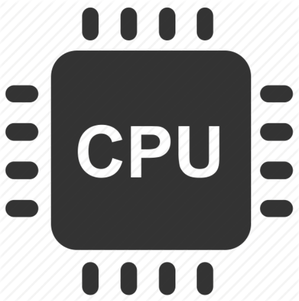


Classic Communication

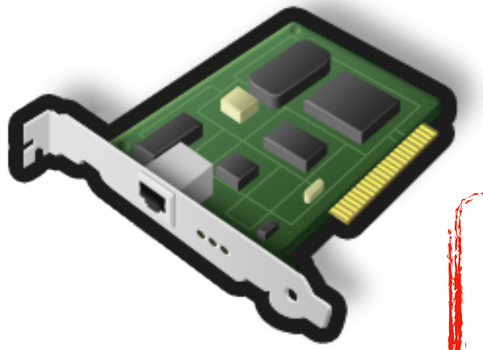


RDMA

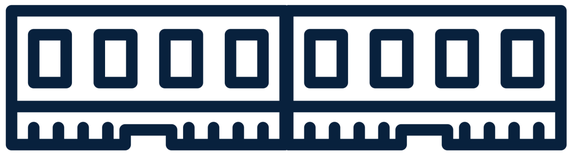
Massive Overhead

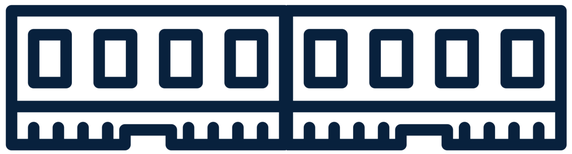


1. Address mapping
2. Permission checking
3. Resource isolation



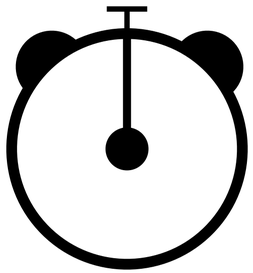
Memory Region
1. rkey/lkey
2. Address

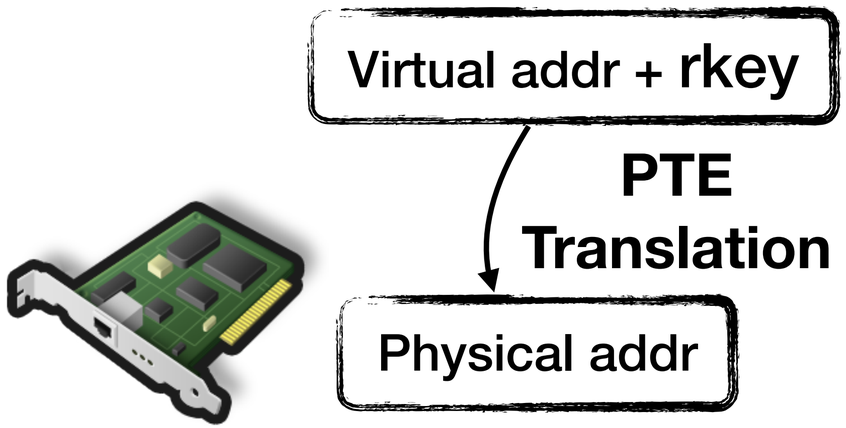




RDMA NIC - Metadata in SRAM

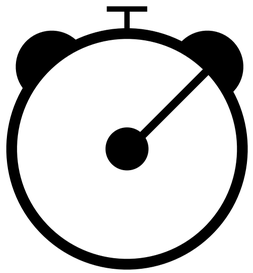
- Queue Pairs
- Memory Regions
- Page Table Entries

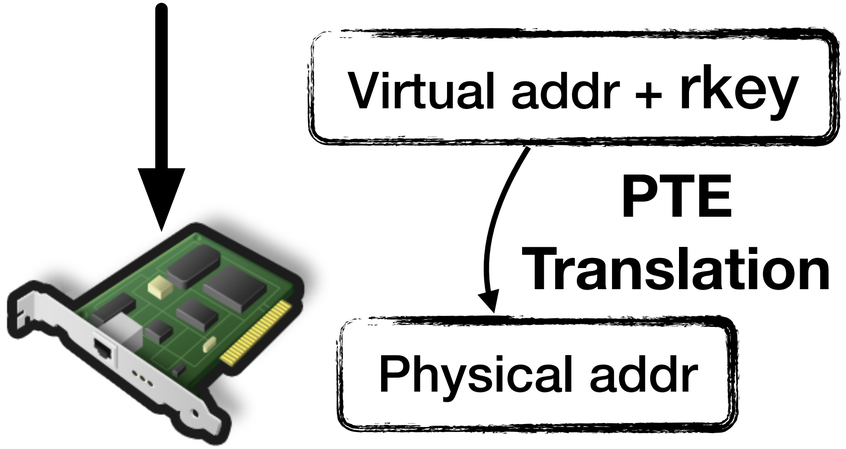




RDMA NIC - Metadata in SRAM

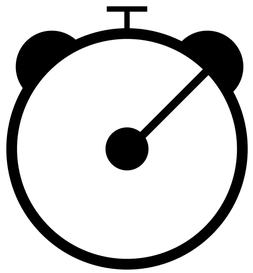
- Queue Pairs
- Memory Regions
- Page Table Entries

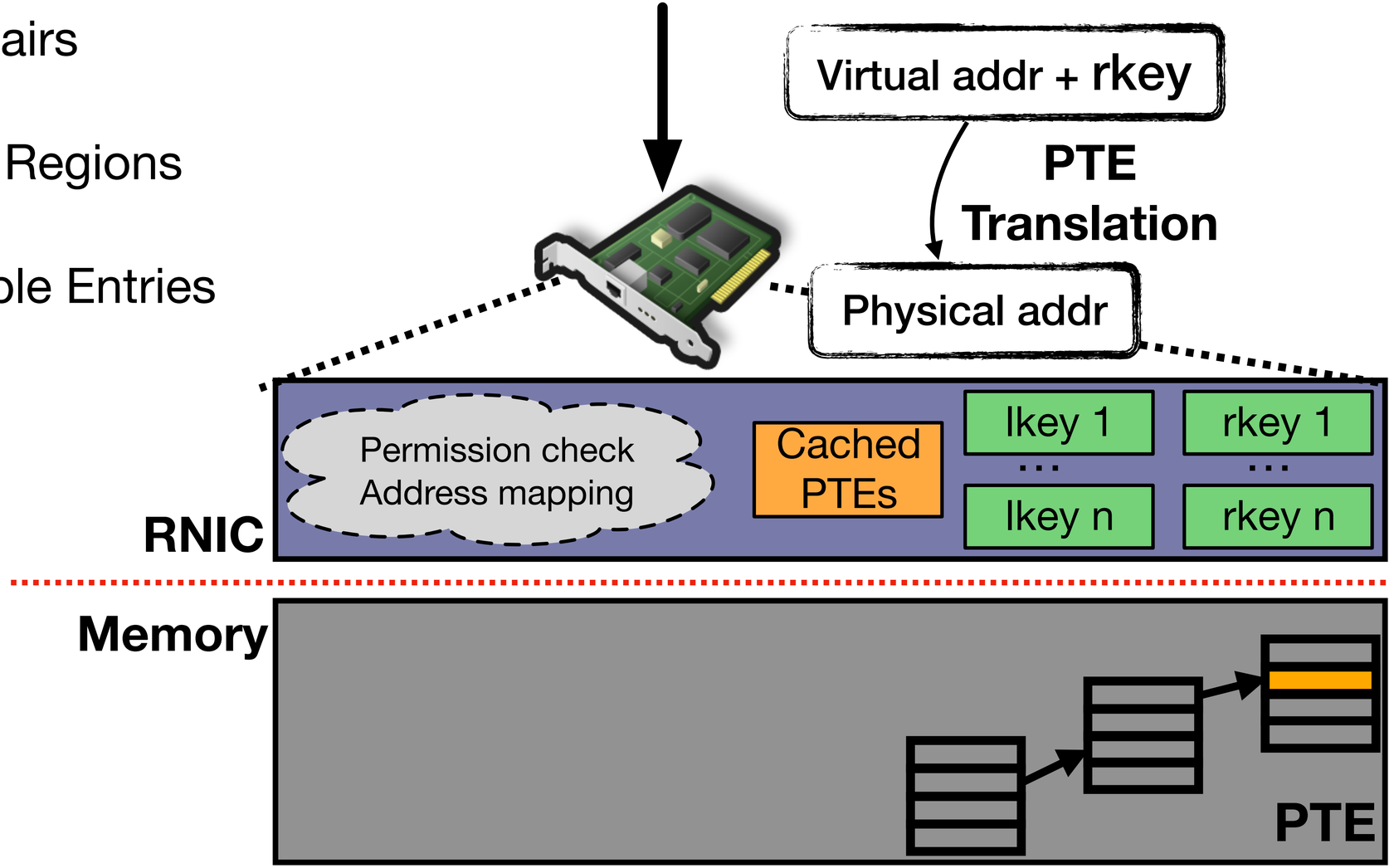




RDMA NIC - Metadata in SRAM

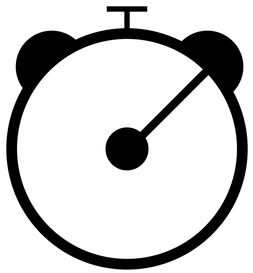
- Queue Pairs
- Memory Regions
- Page Table Entries

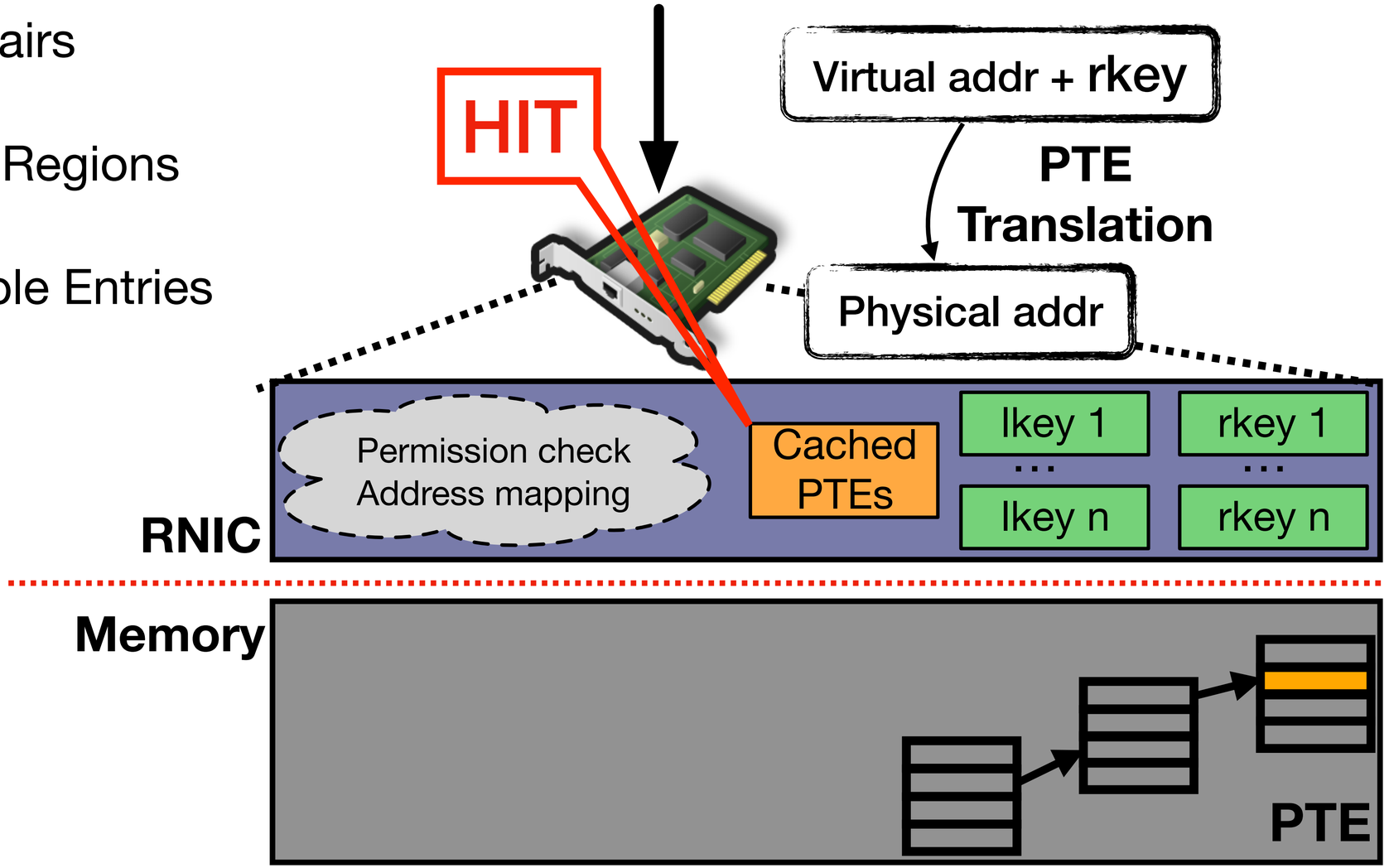




RDMA NIC - Metadata in SRAM

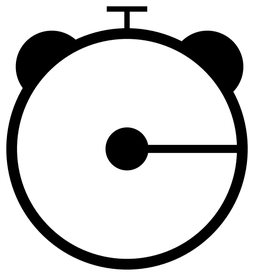
- Queue Pairs
- Memory Regions
- Page Table Entries

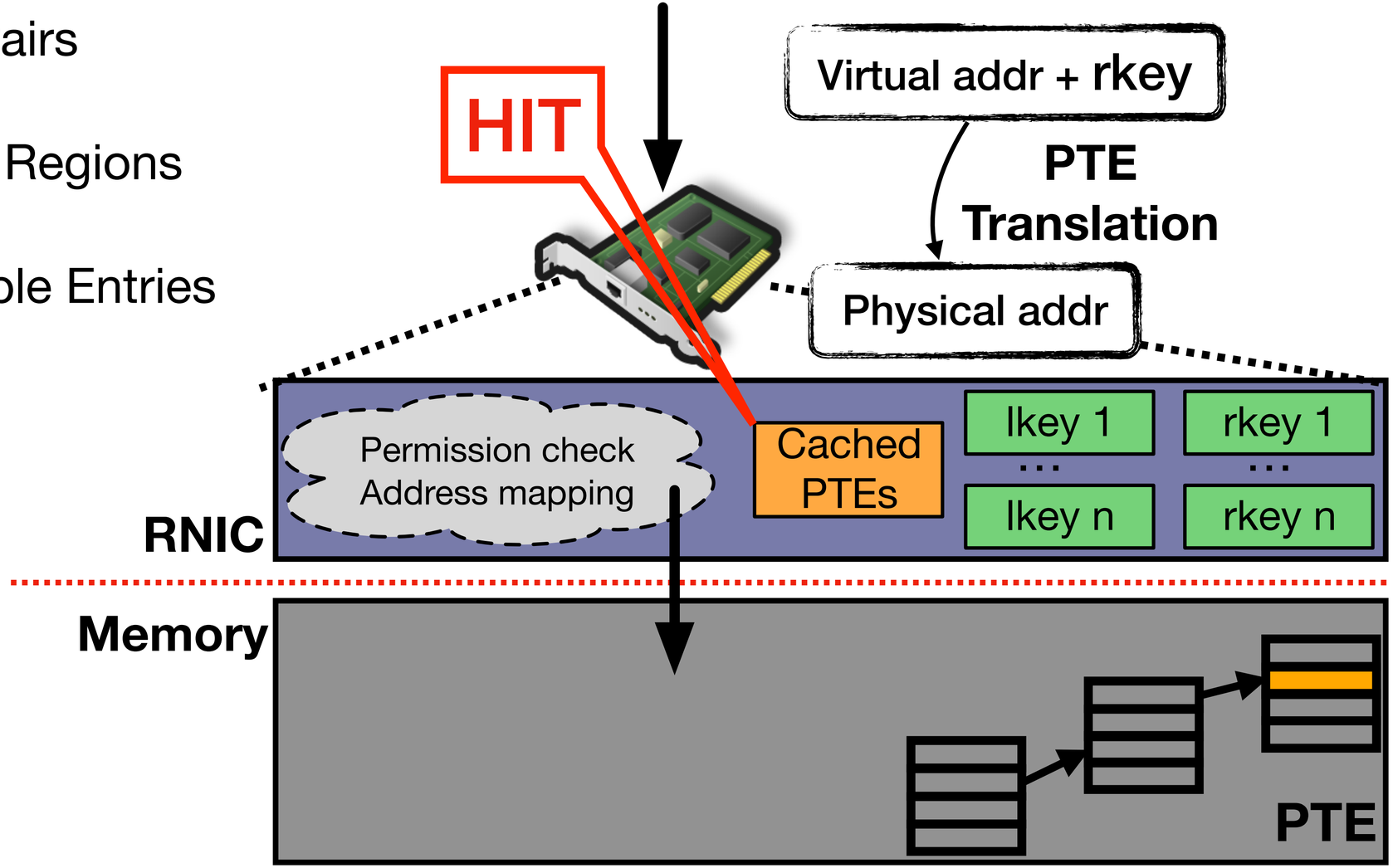




RDMA NIC - Metadata in SRAM

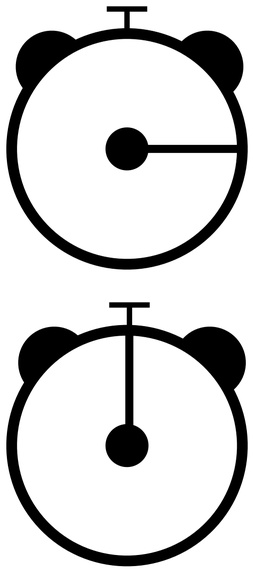
- Queue Pairs
- Memory Regions
- Page Table Entries

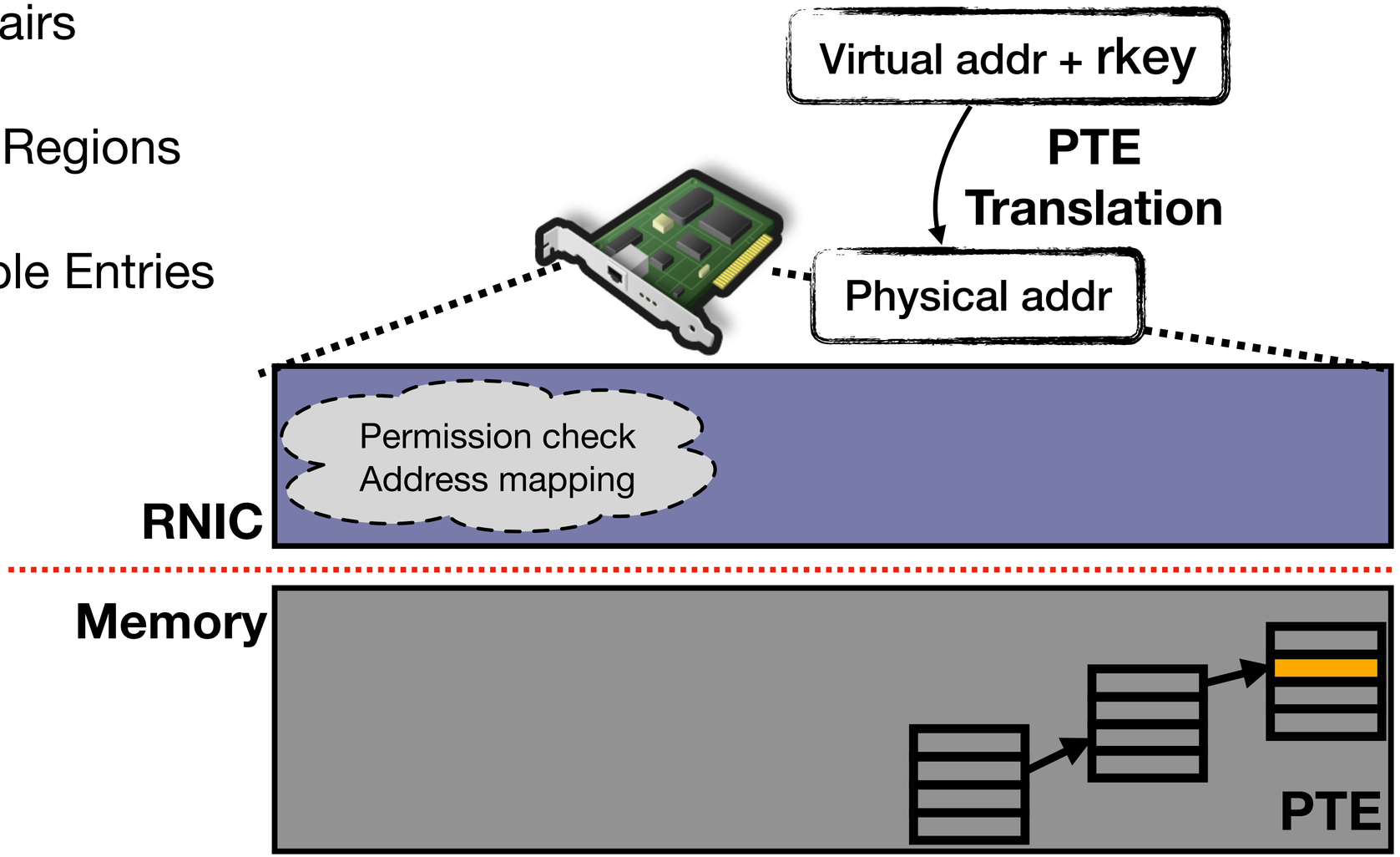




RDMA NIC - Metadata in SRAM

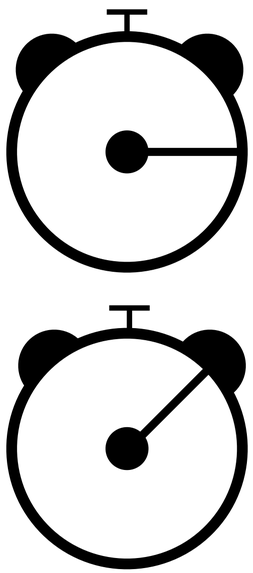
- Queue Pairs
- Memory Regions
- Page Table Entries

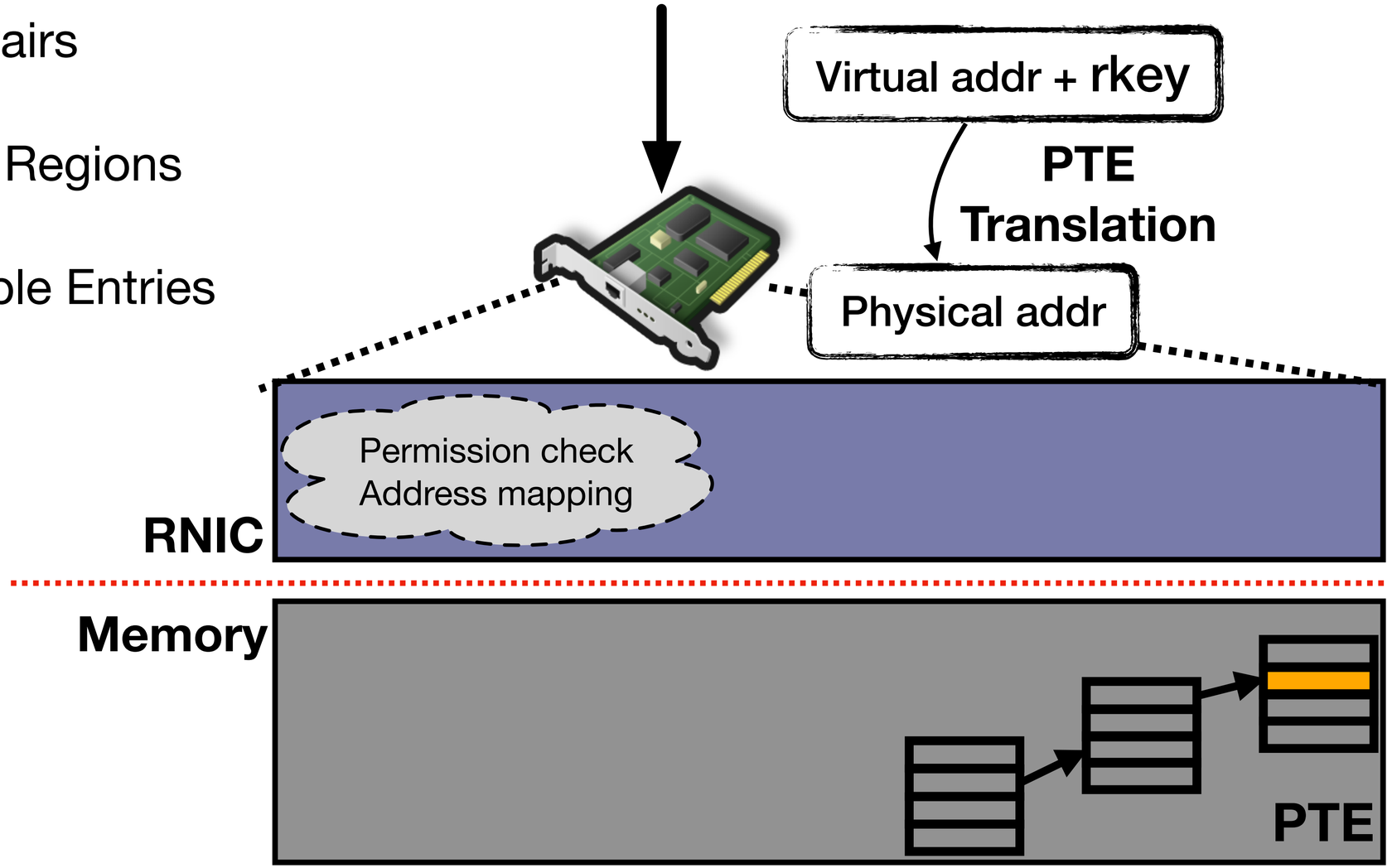




RDMA NIC - Metadata in SRAM

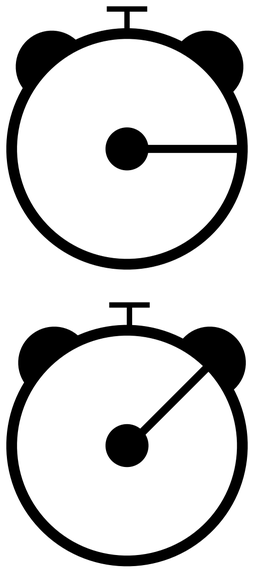
- Queue Pairs
- Memory Regions
- Page Table Entries

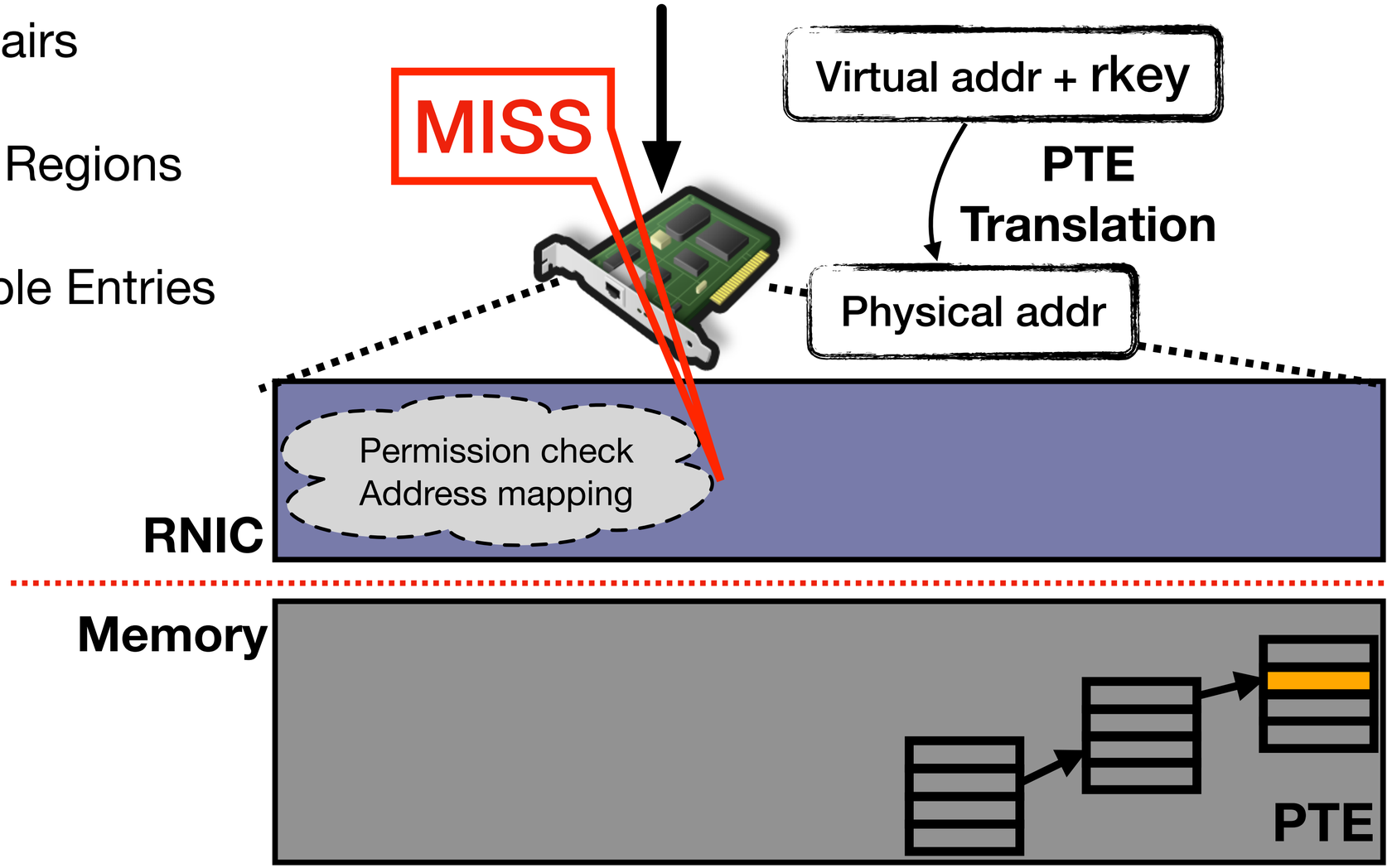




RDMA NIC - Metadata in SRAM

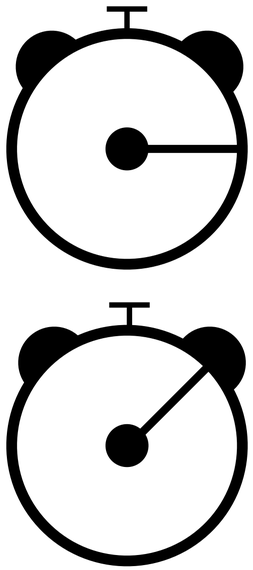
- Queue Pairs
- Memory Regions
- Page Table Entries

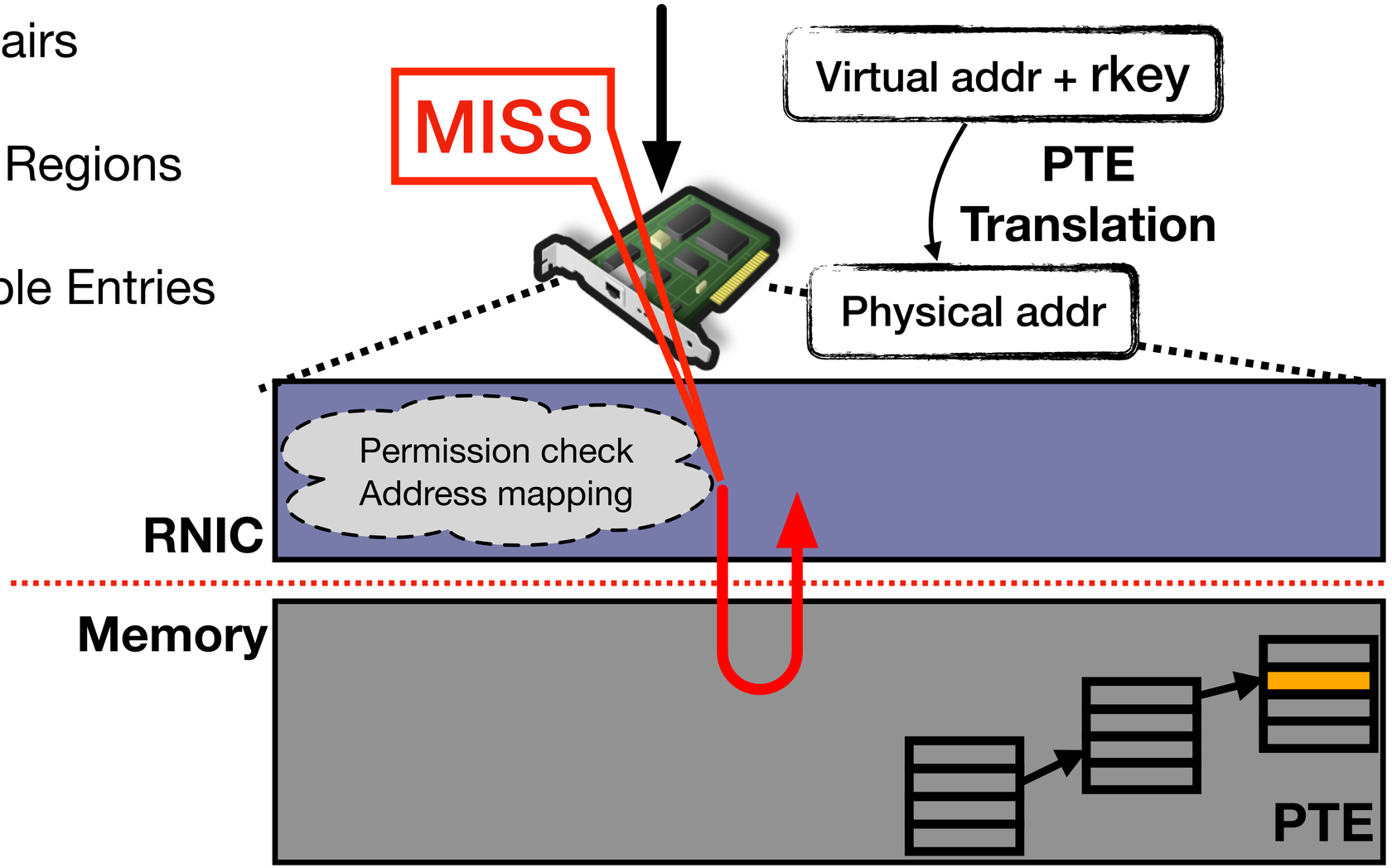




RDMA NIC - Metadata in SRAM

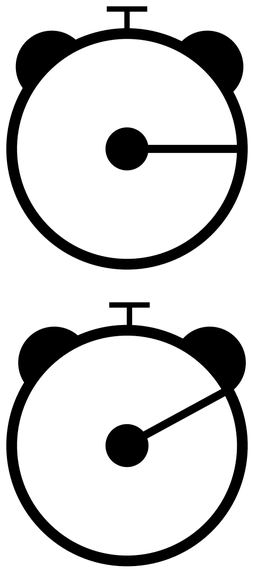
- Queue Pairs
- Memory Regions
- Page Table Entries

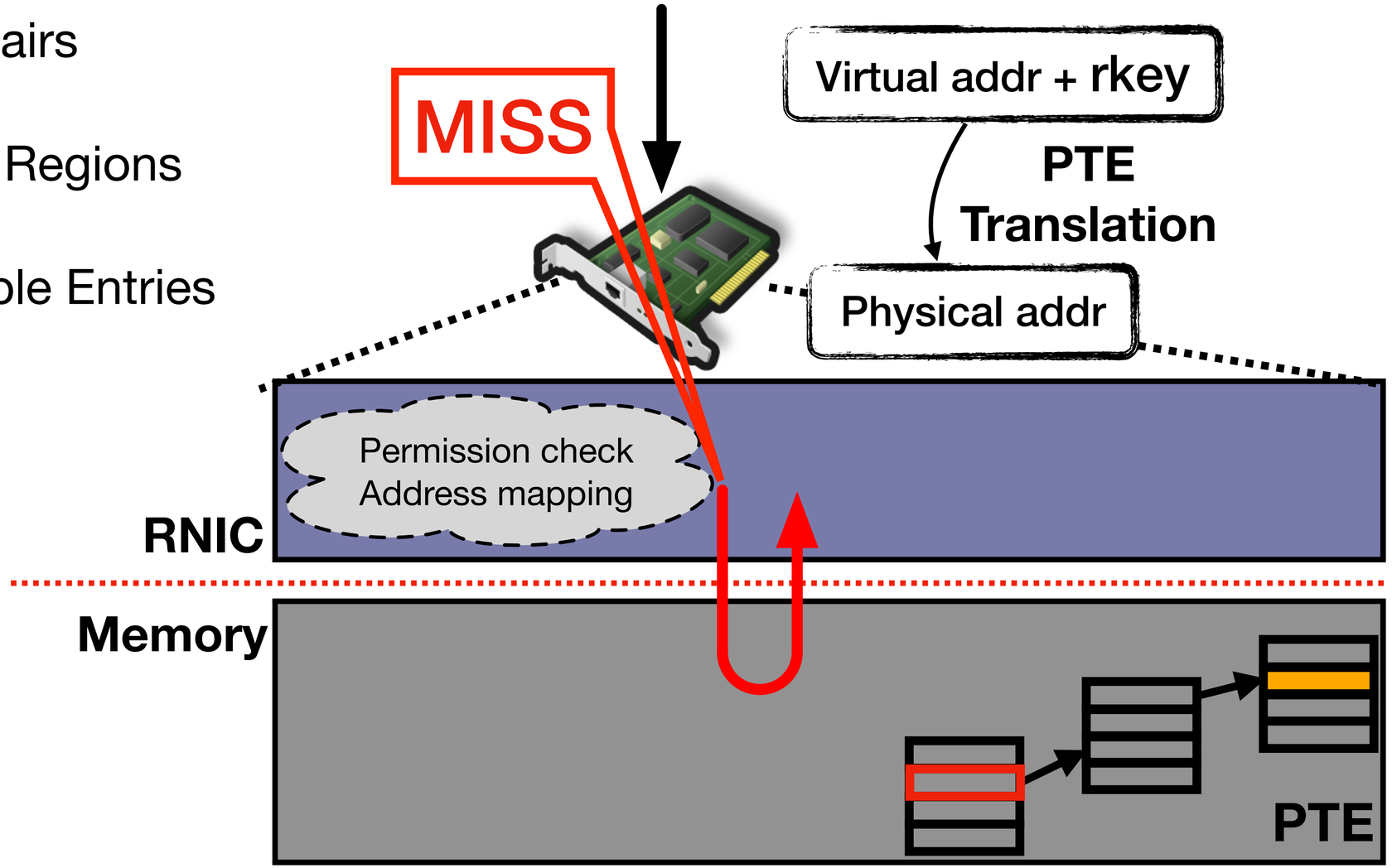




RDMA NIC - Metadata in SRAM

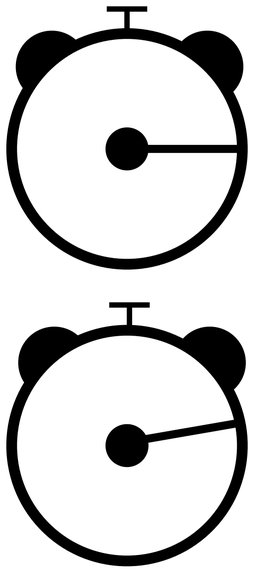
- Queue Pairs
- Memory Regions
- Page Table Entries

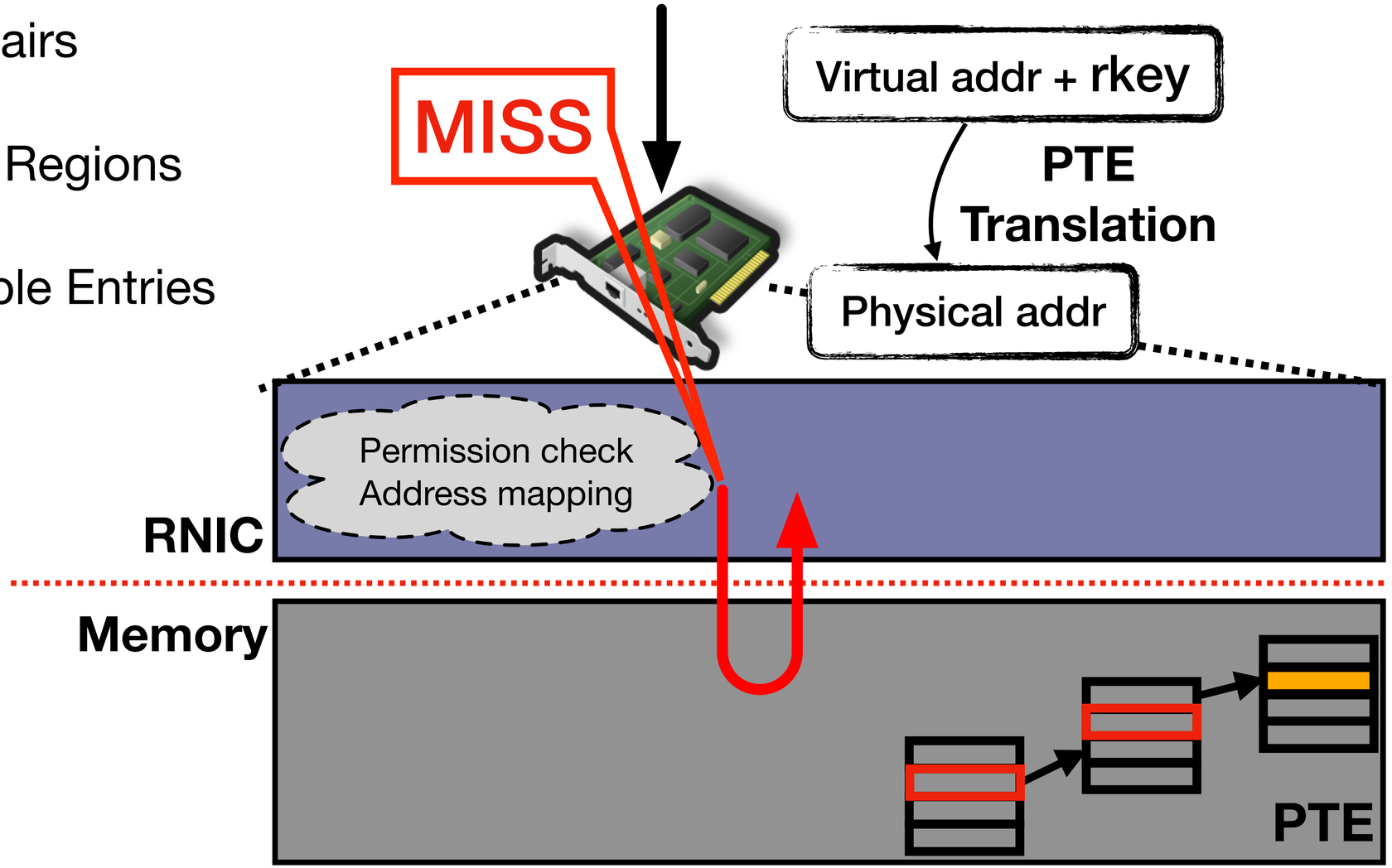




RDMA NIC - Metadata in SRAM

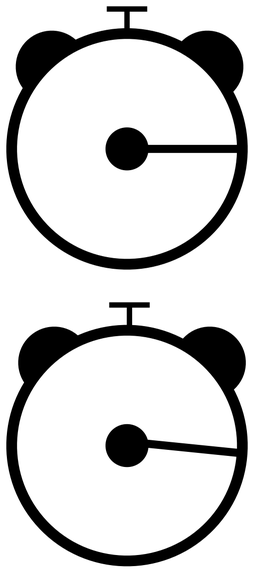
- Queue Pairs
- Memory Regions
- Page Table Entries

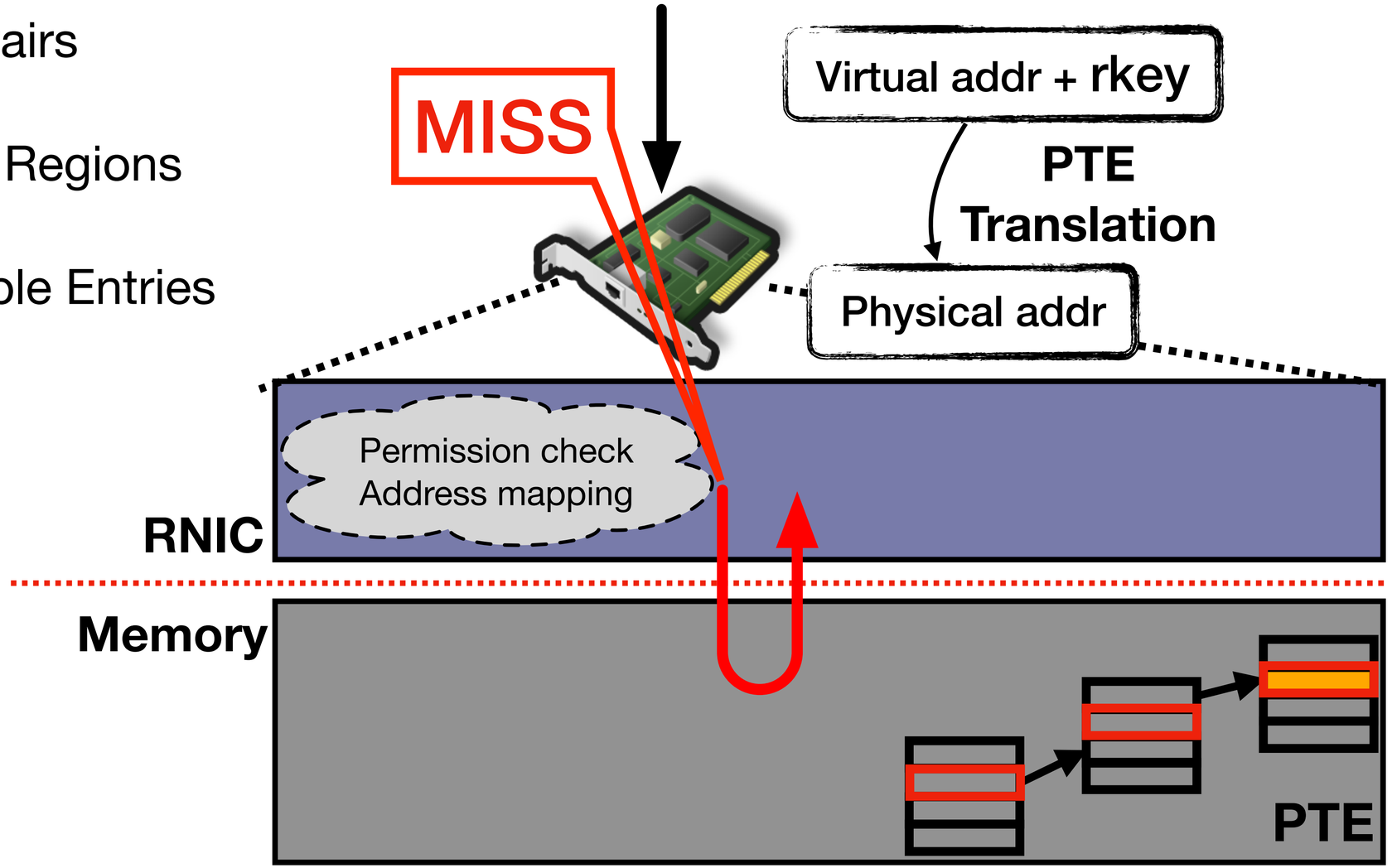




RDMA NIC - Metadata in SRAM

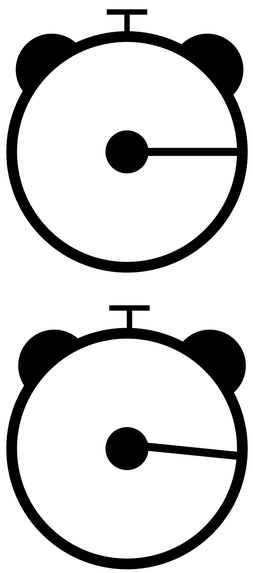
- Queue Pairs
- Memory Regions
- Page Table Entries

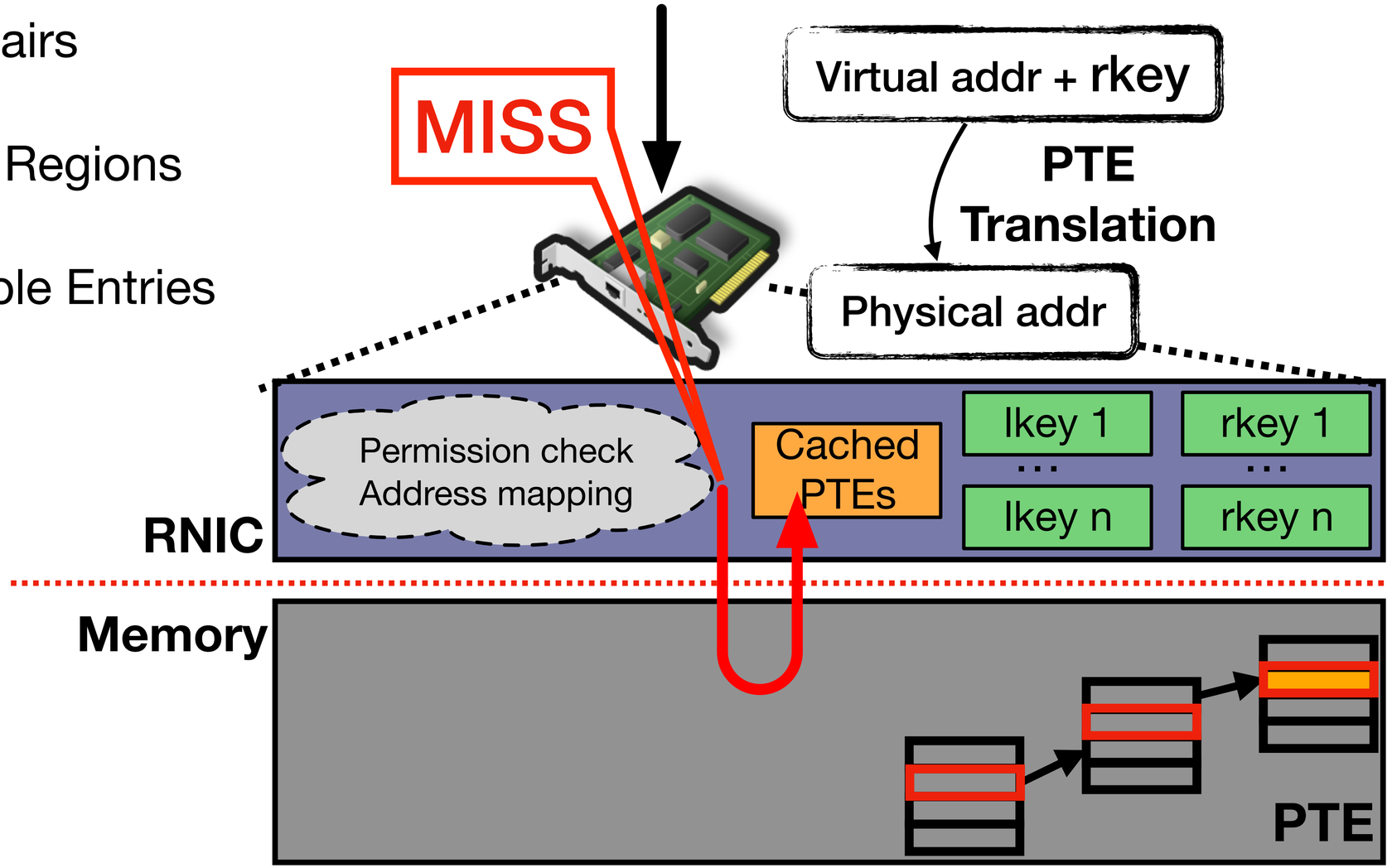




RDMA NIC - Metadata in SRAM

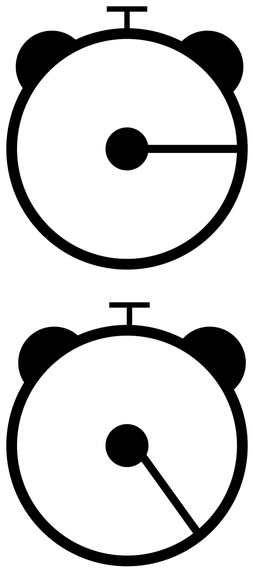
- Queue Pairs
- Memory Regions
- Page Table Entries

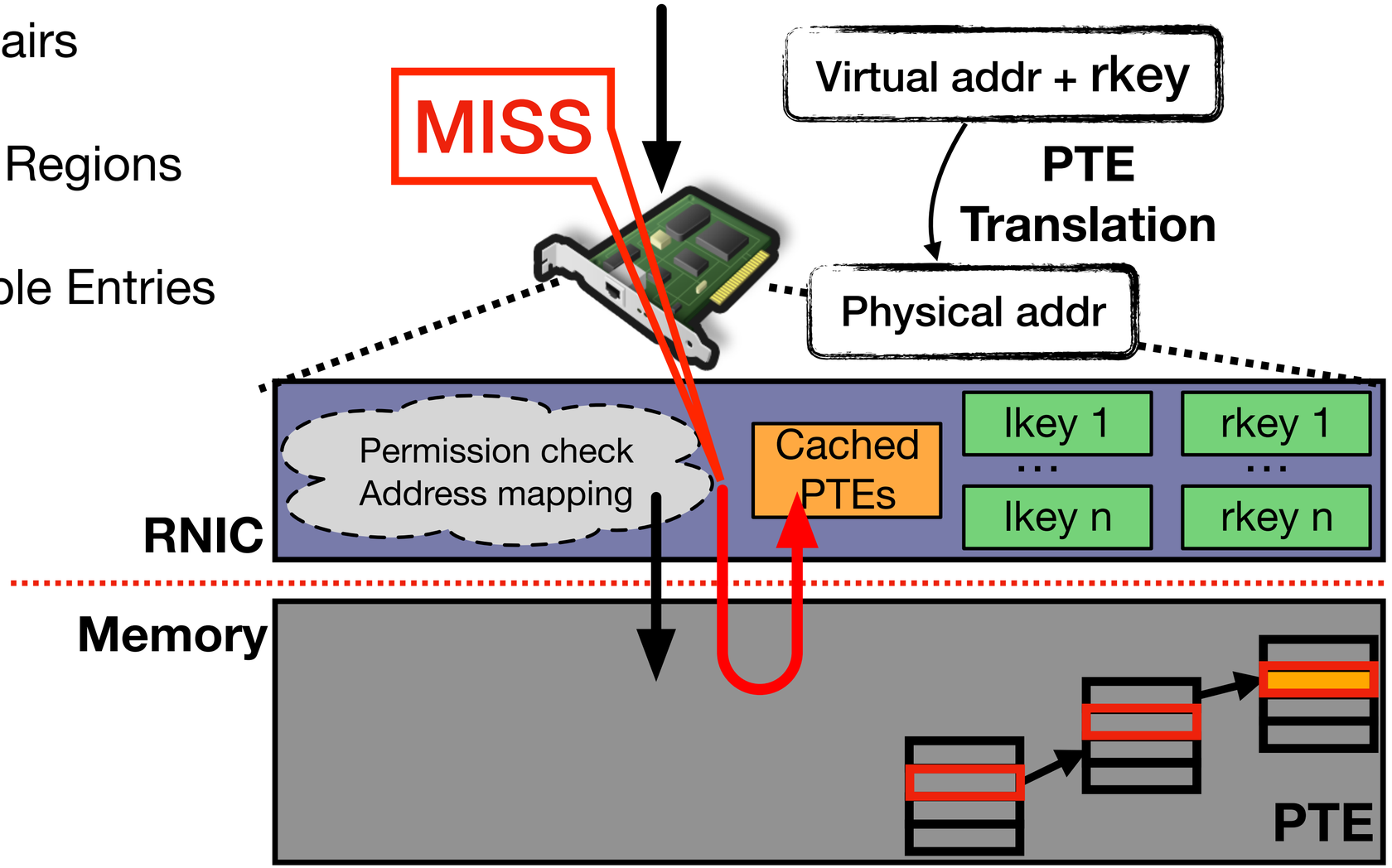




RDMA NIC - Metadata in SRAM

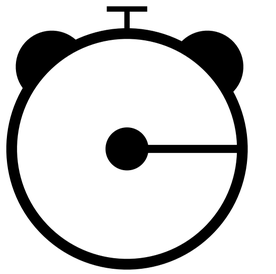
- Queue Pairs
- Memory Regions
- Page Table Entries



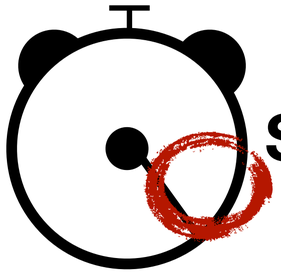


RDMA NIC - Metadata in SRAM

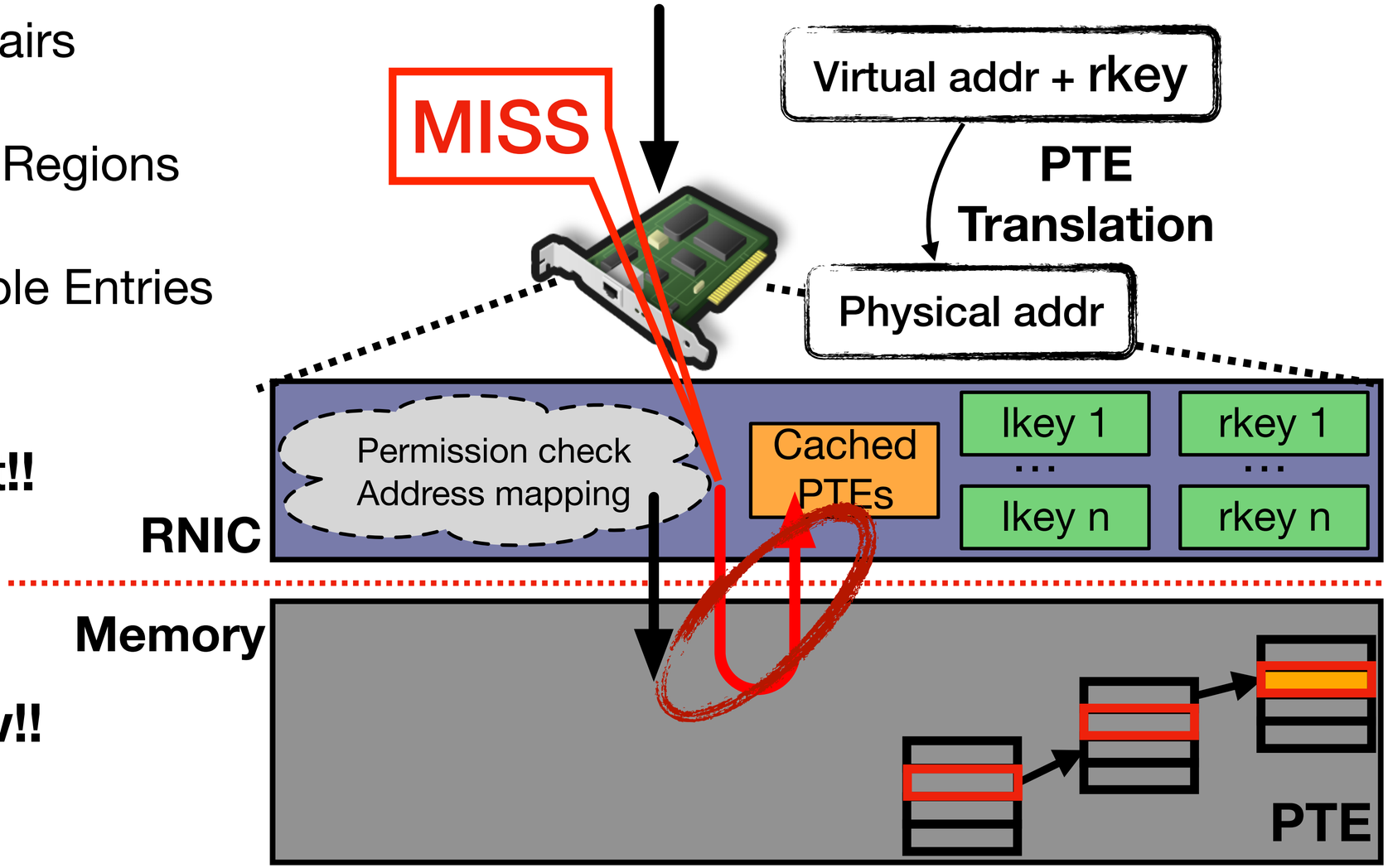
- Queue Pairs
- Memory Regions
- Page Table Entries



Fast!!



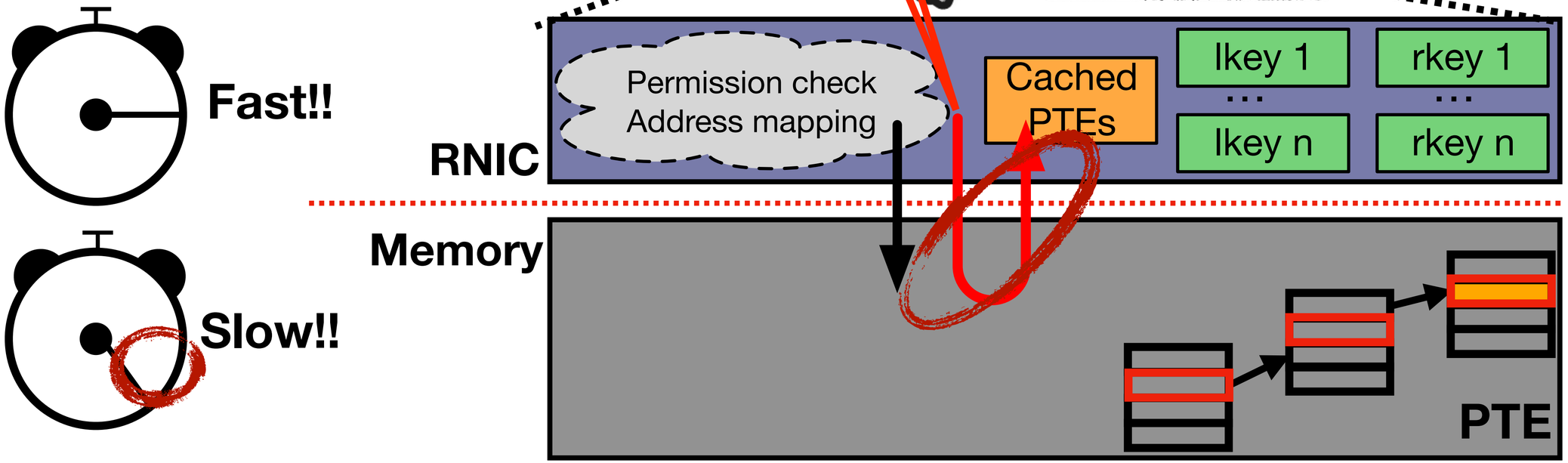
Slow!!



RDMA NIC - Metadata in SRAM

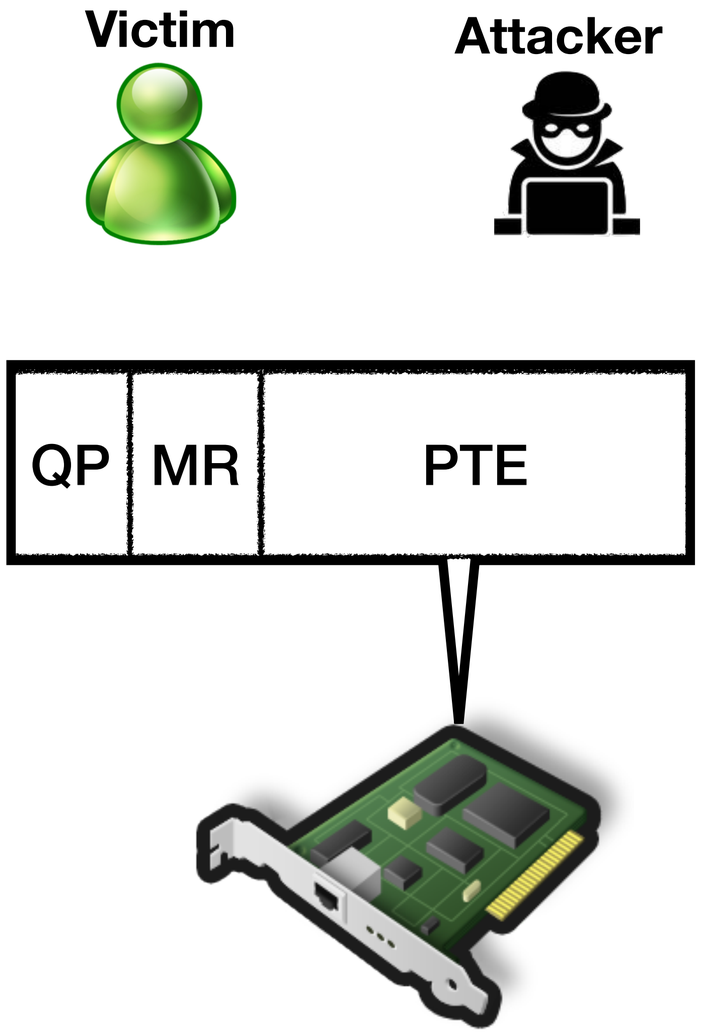
- Queue Pairs
- Memory Regions
- Page Table Entries

Side channel!!



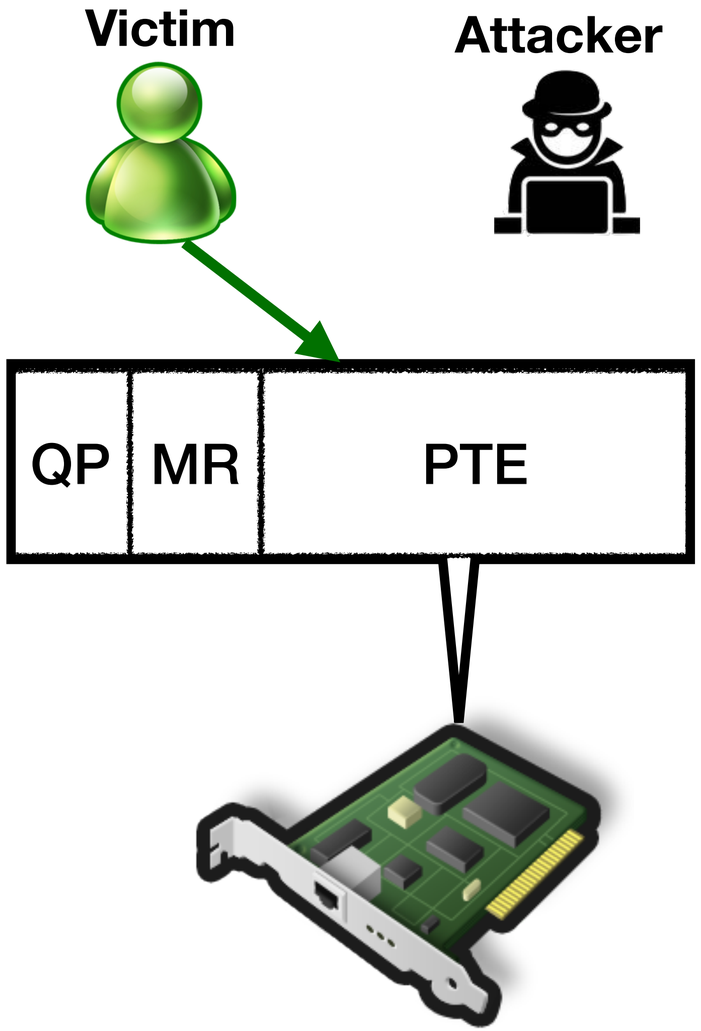
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



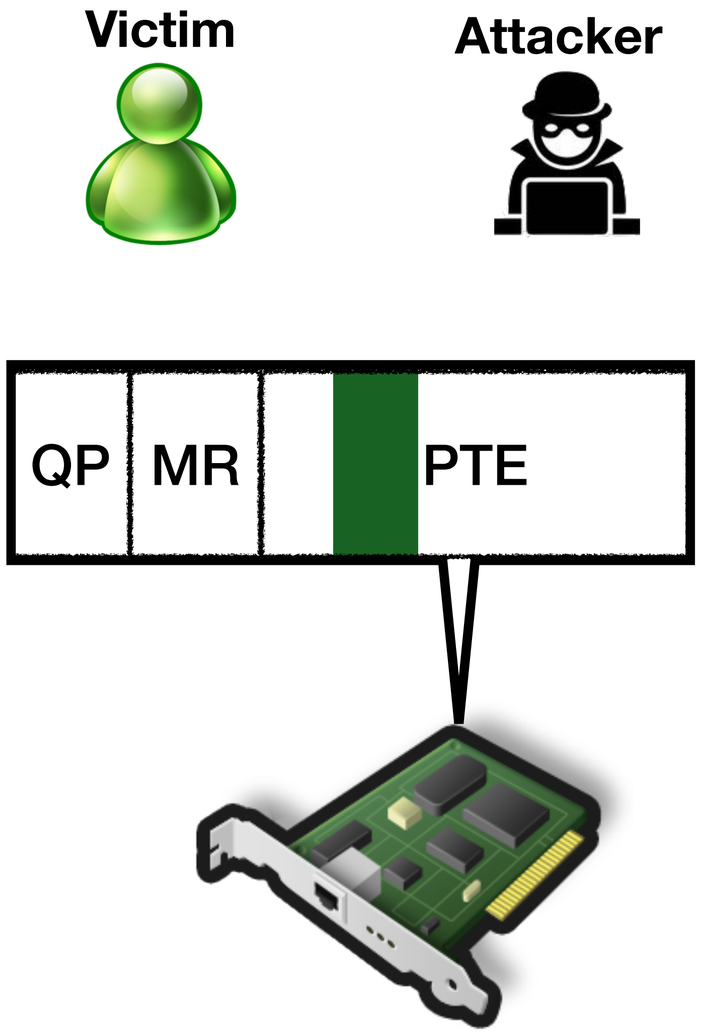
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



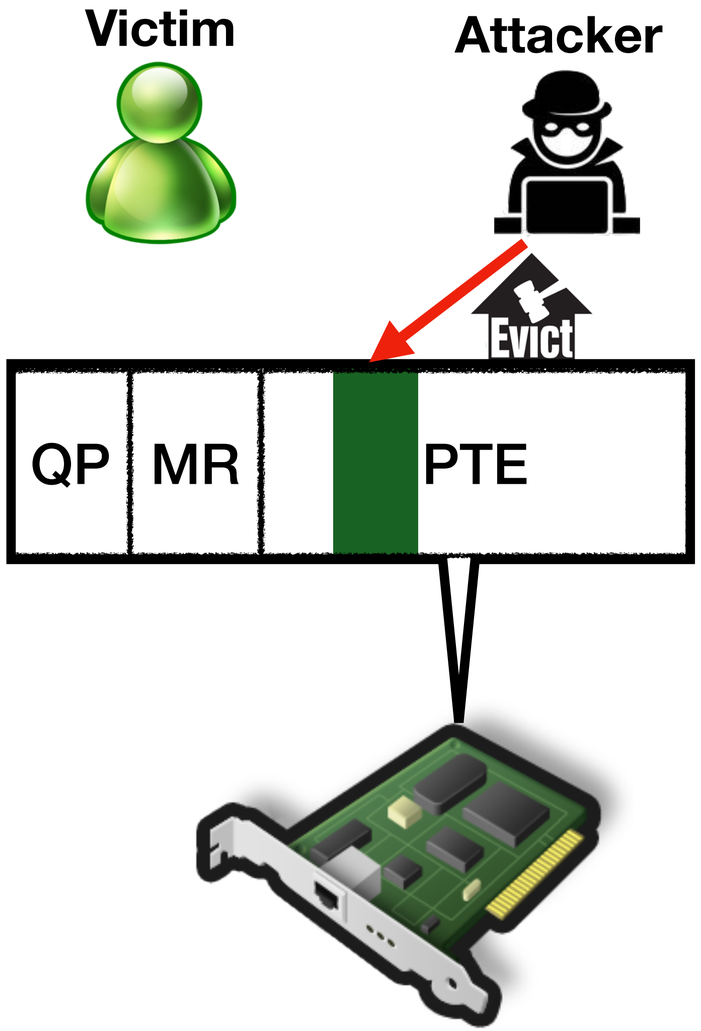
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



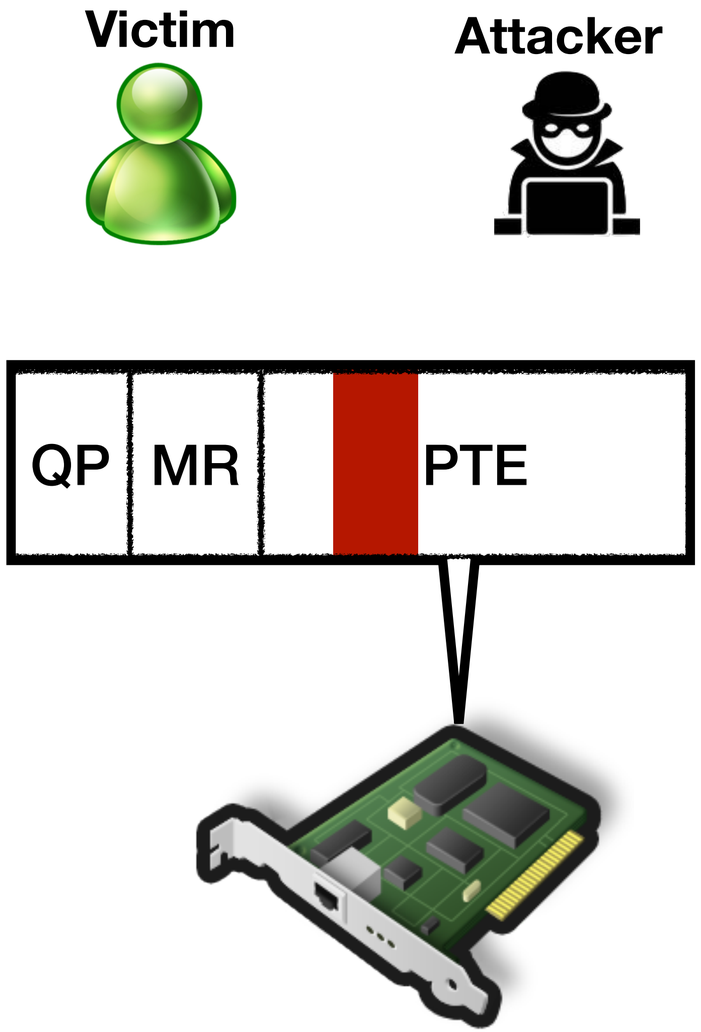
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



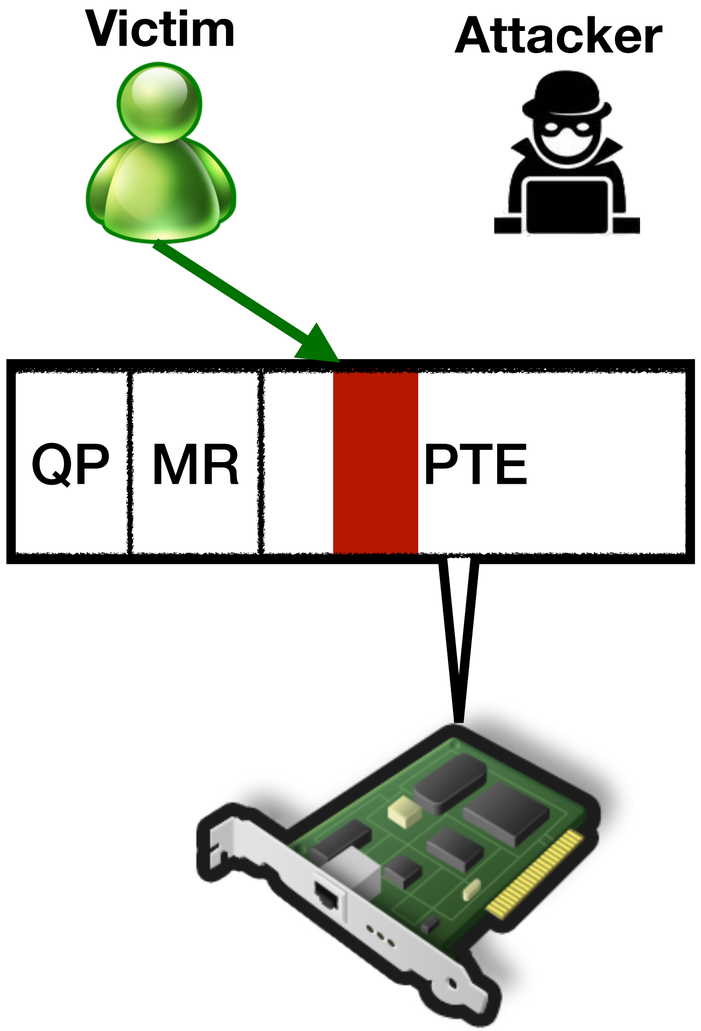
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



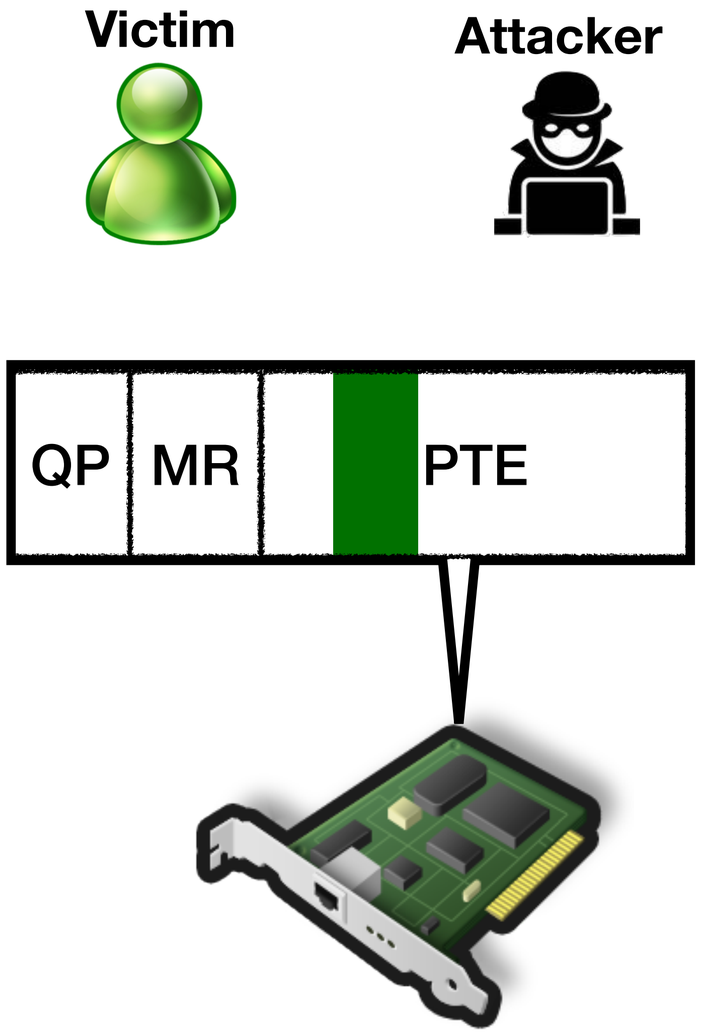
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



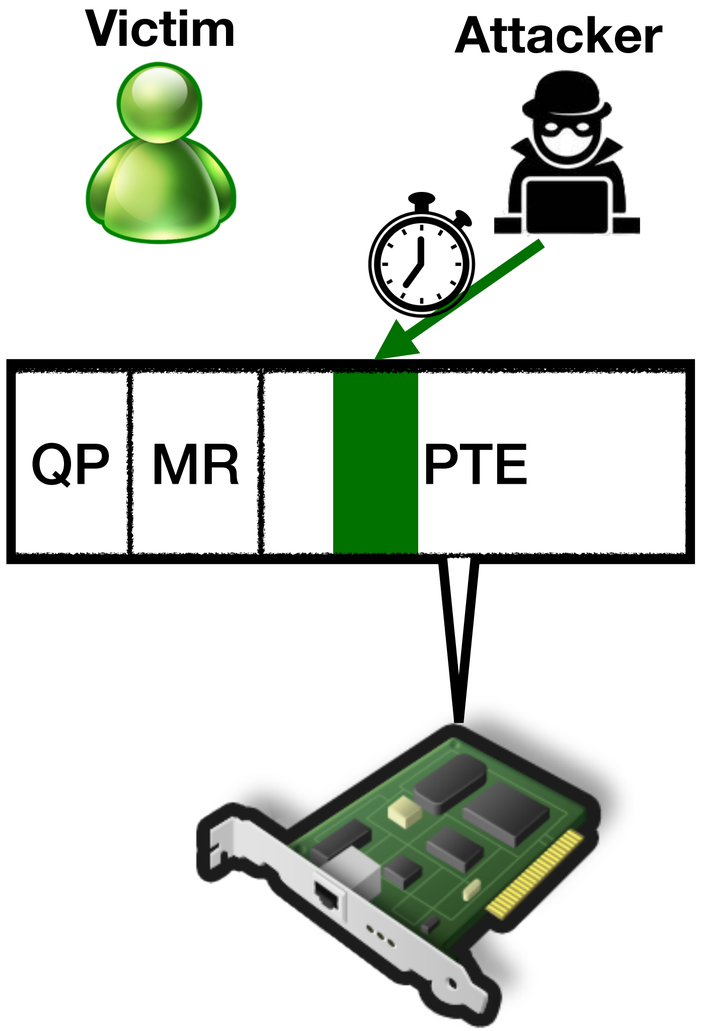
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



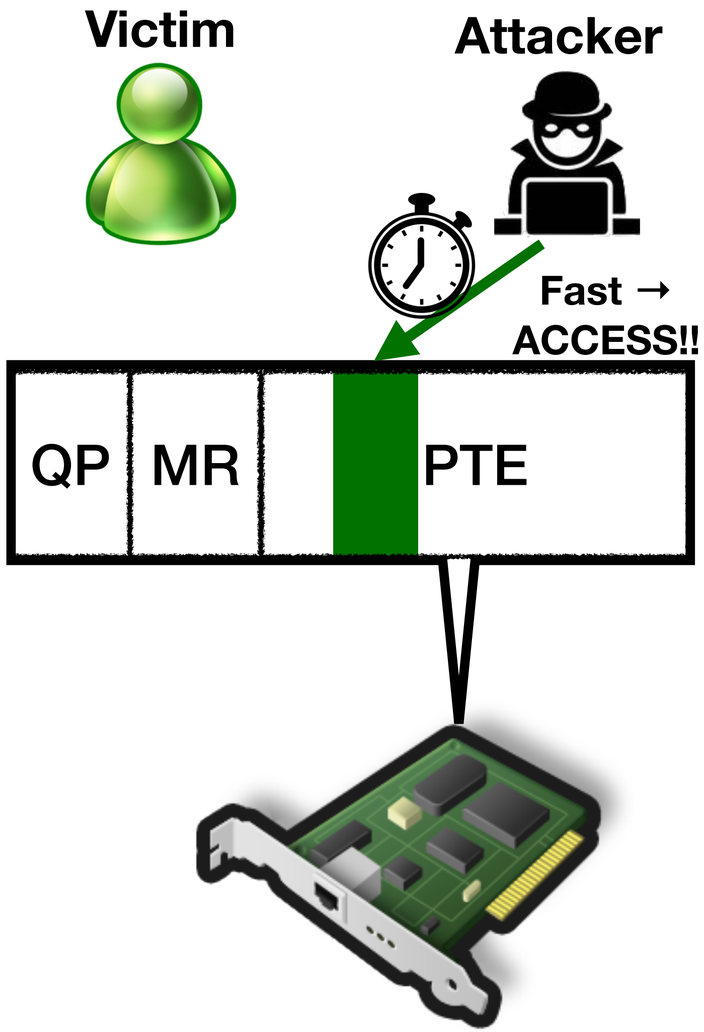
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



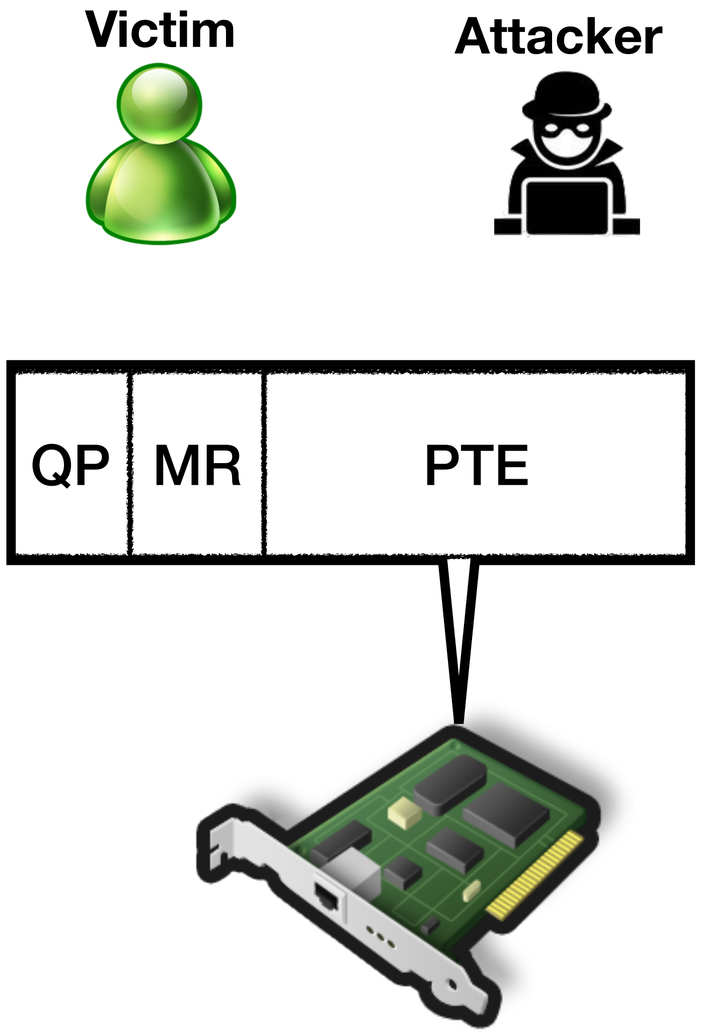
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



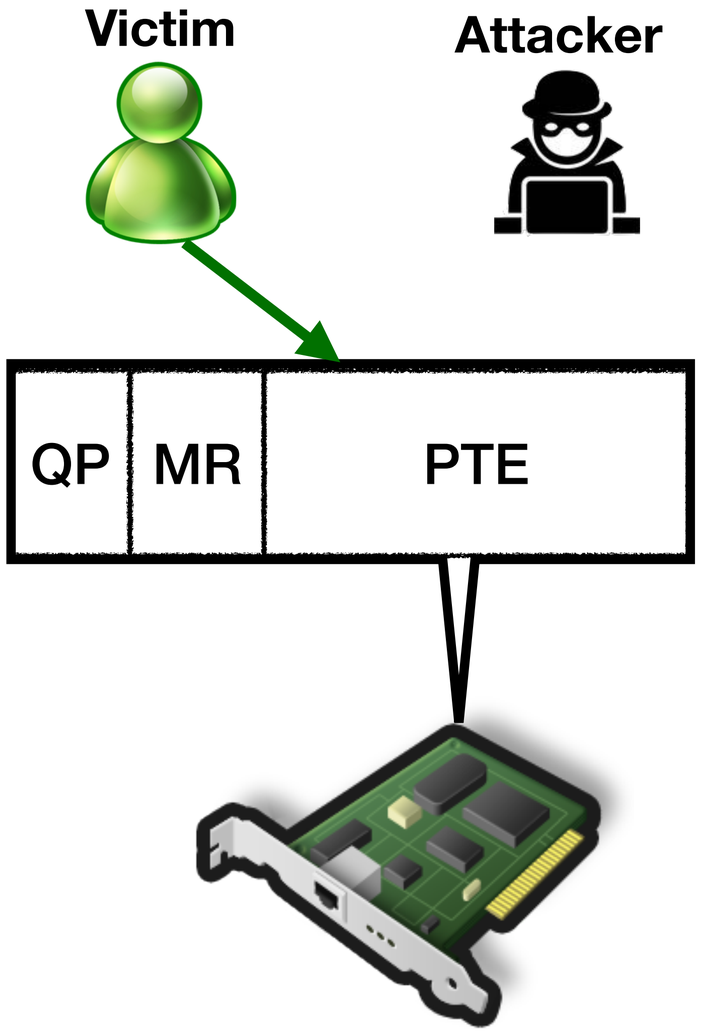
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



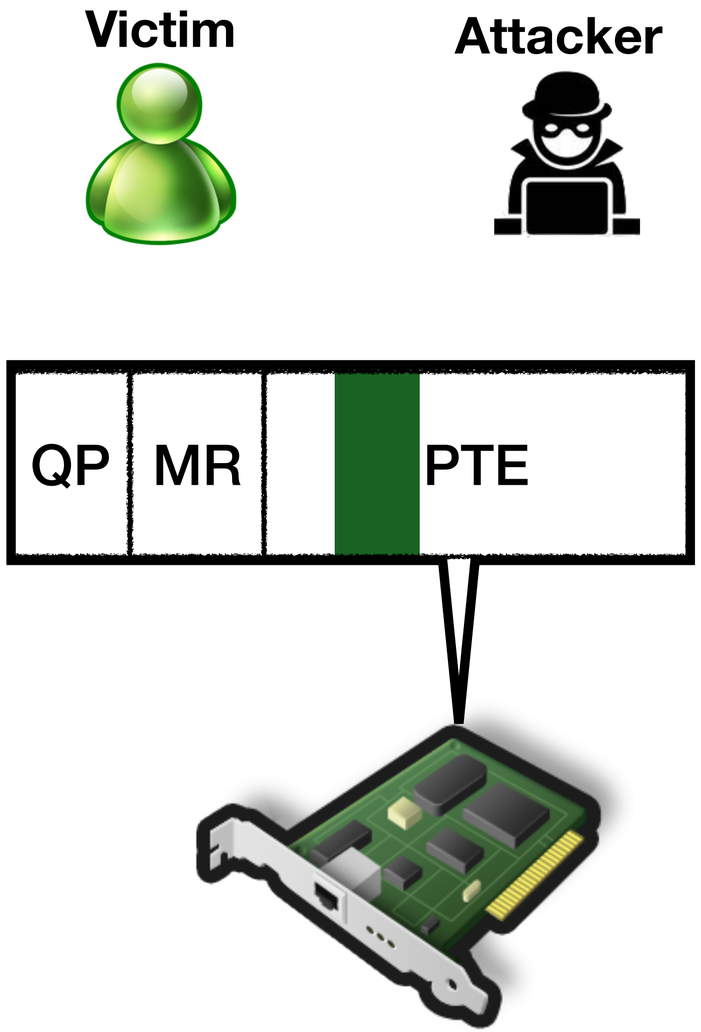
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



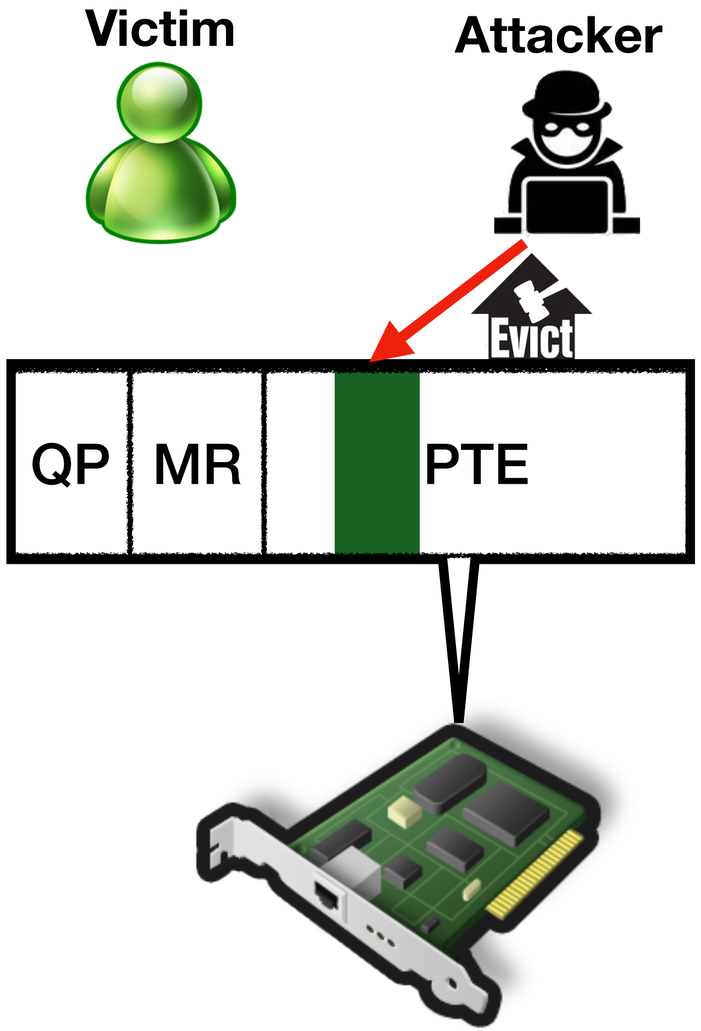
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



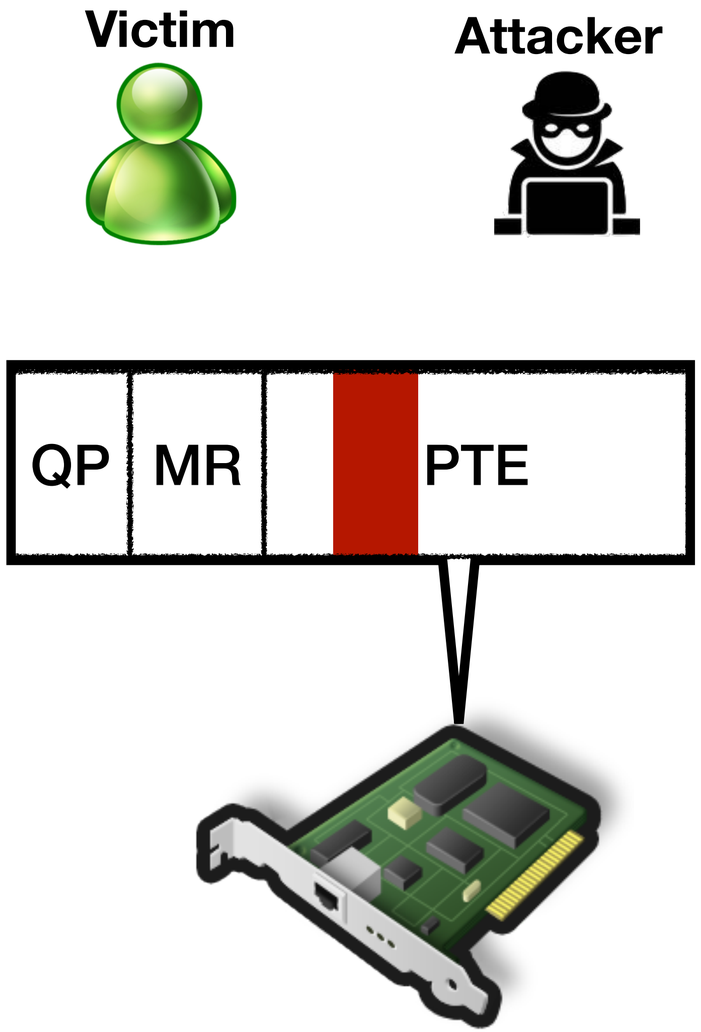
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



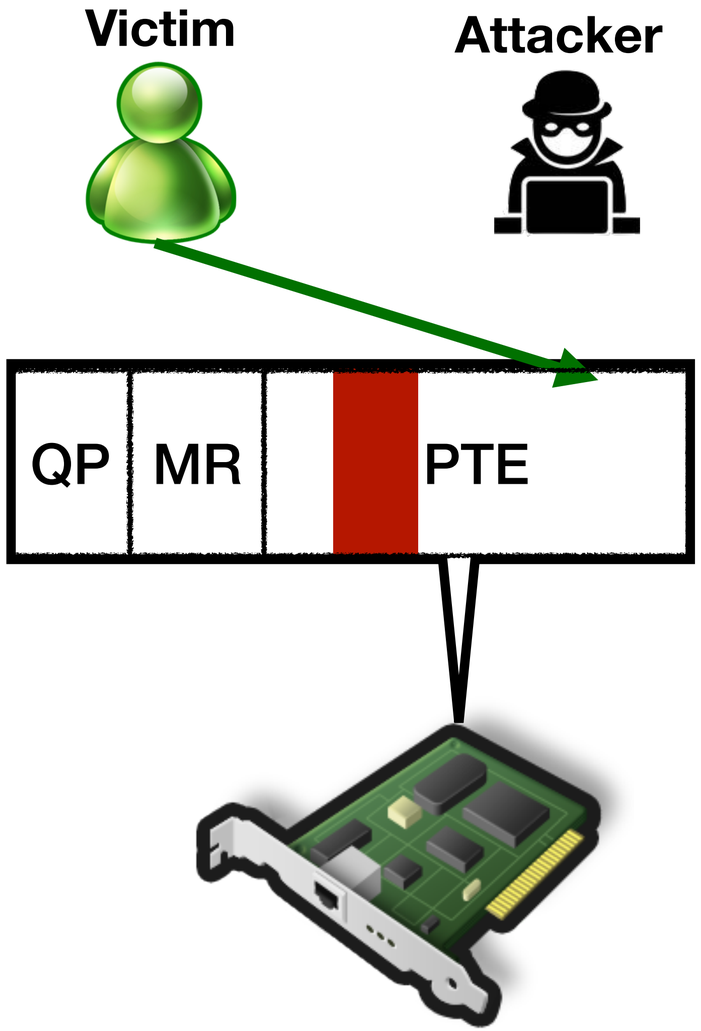
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



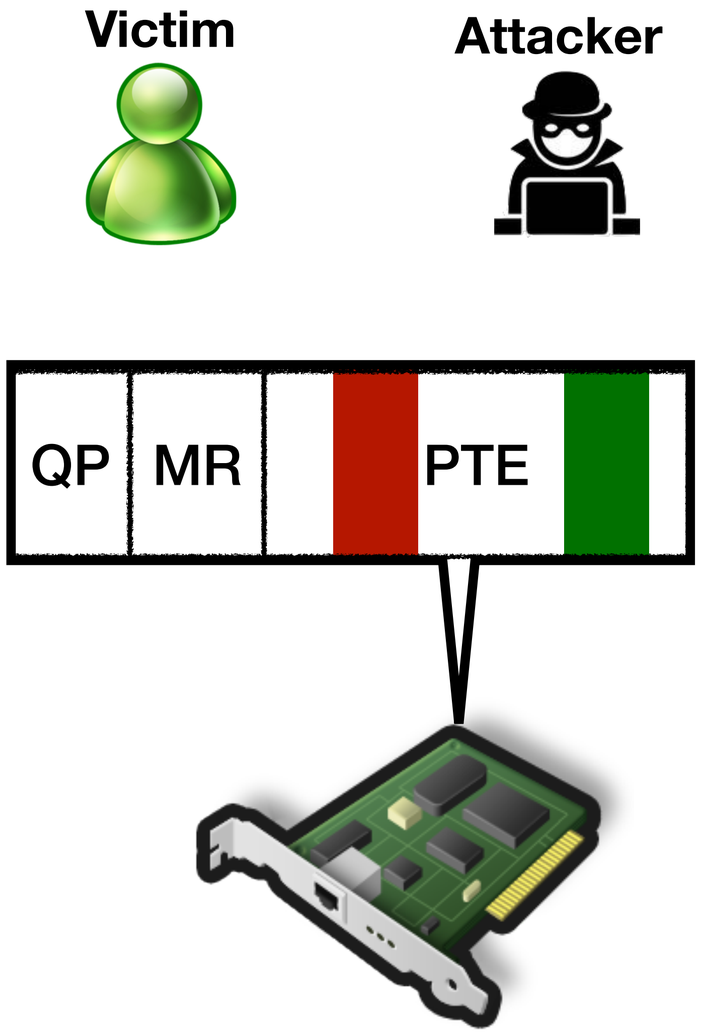
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



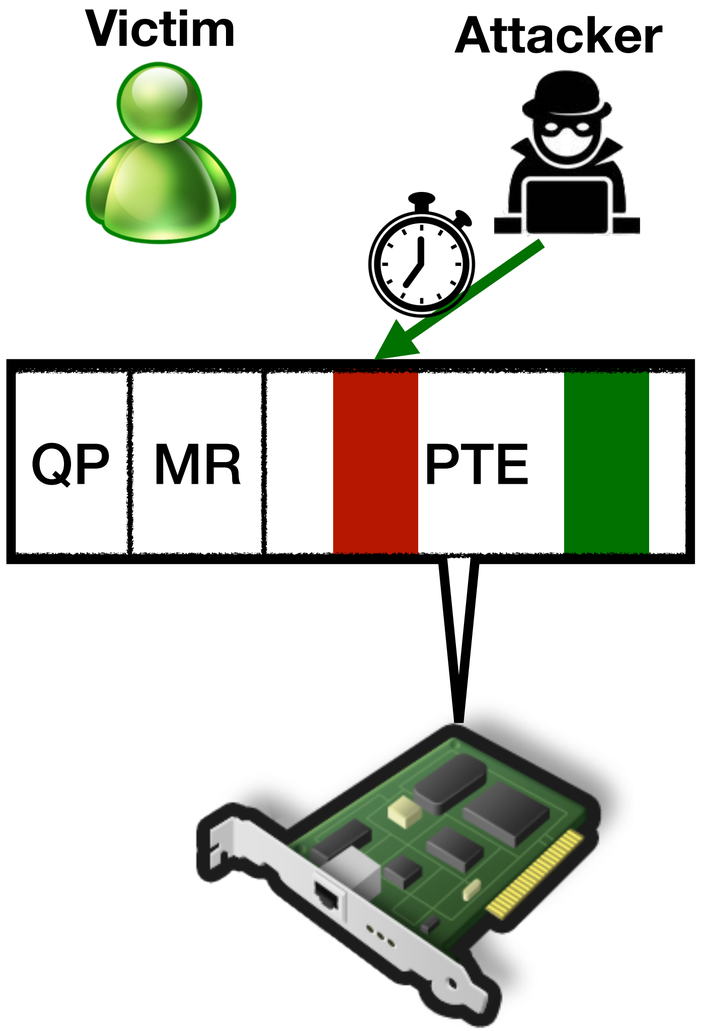
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



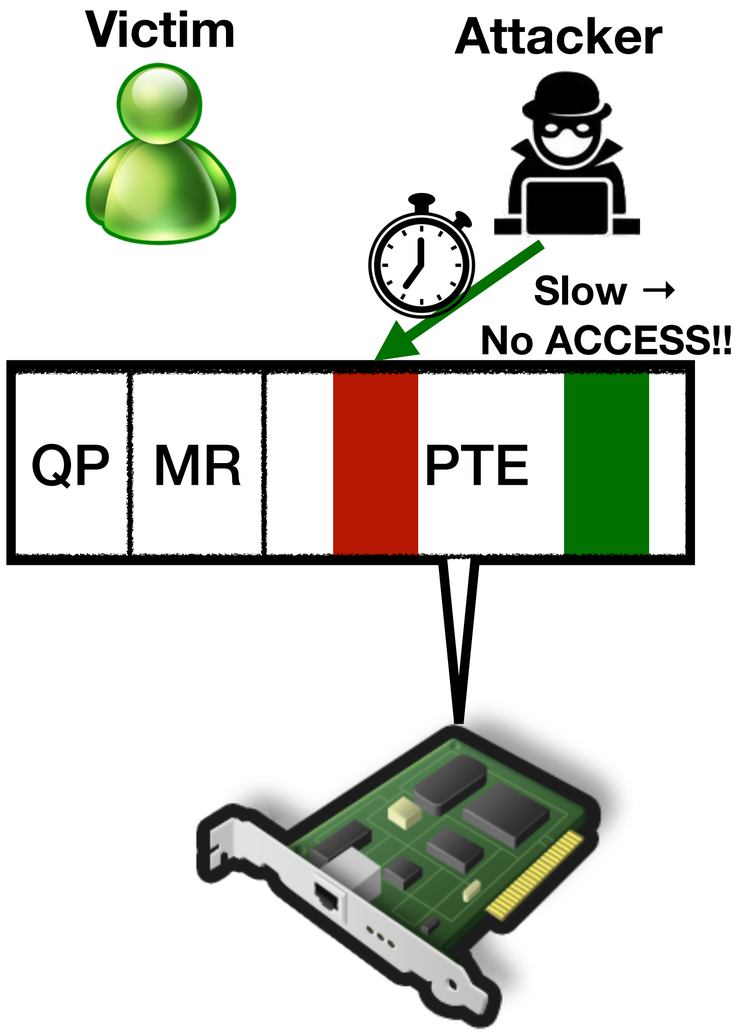
Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



Pythia Basic Idea

- **EVICT** SRAM
- Wait a bit
- Measure time to **RELOAD** a page
  - Fast → access
  - Slow → no access
- Repeat



Why Reverse-Engineering?

- Basic Evict+Reload is too slow

25 ms

Why Reverse-Engineering?

- Basic Evict+Reload is too slow

25 ms

Why Reverse-Engineering?

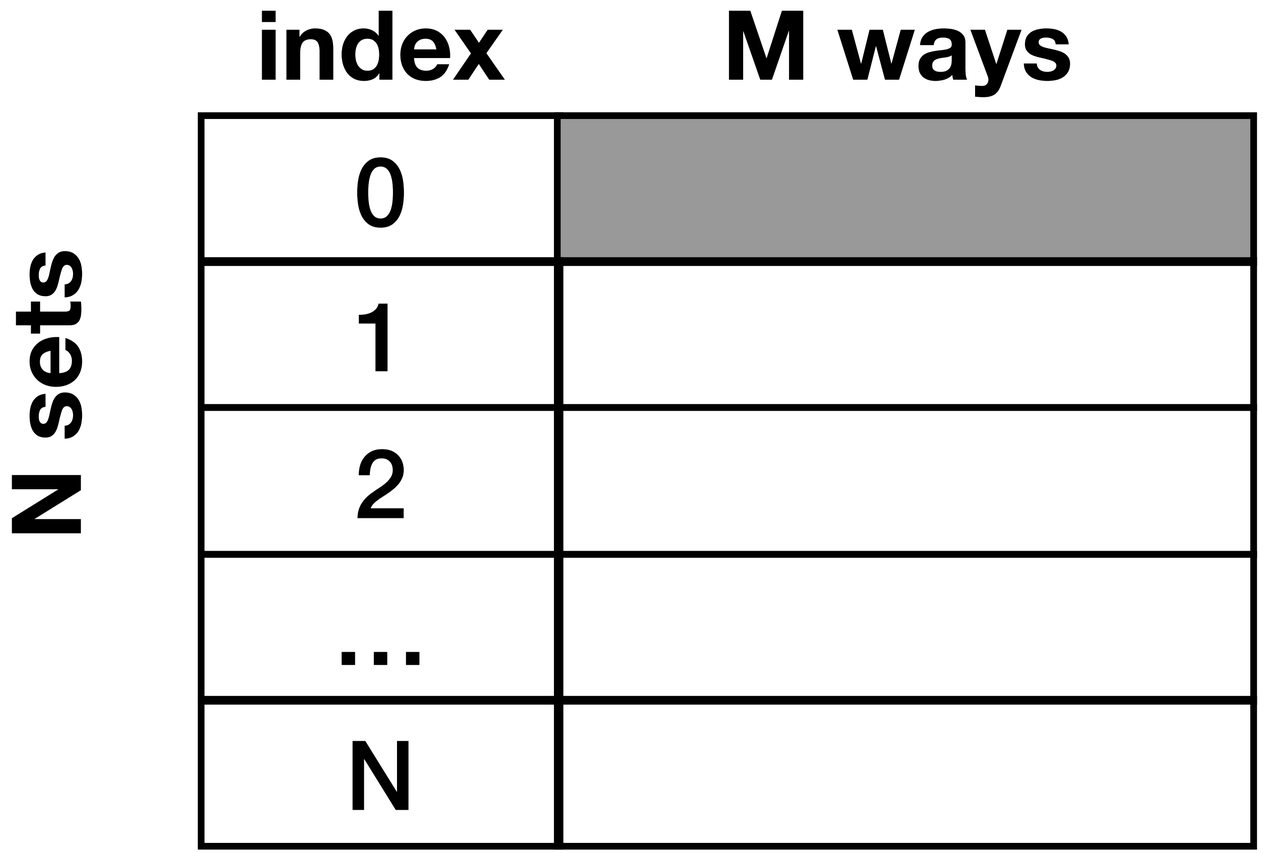
- Basic Evict+Reload is too slow

25 ms

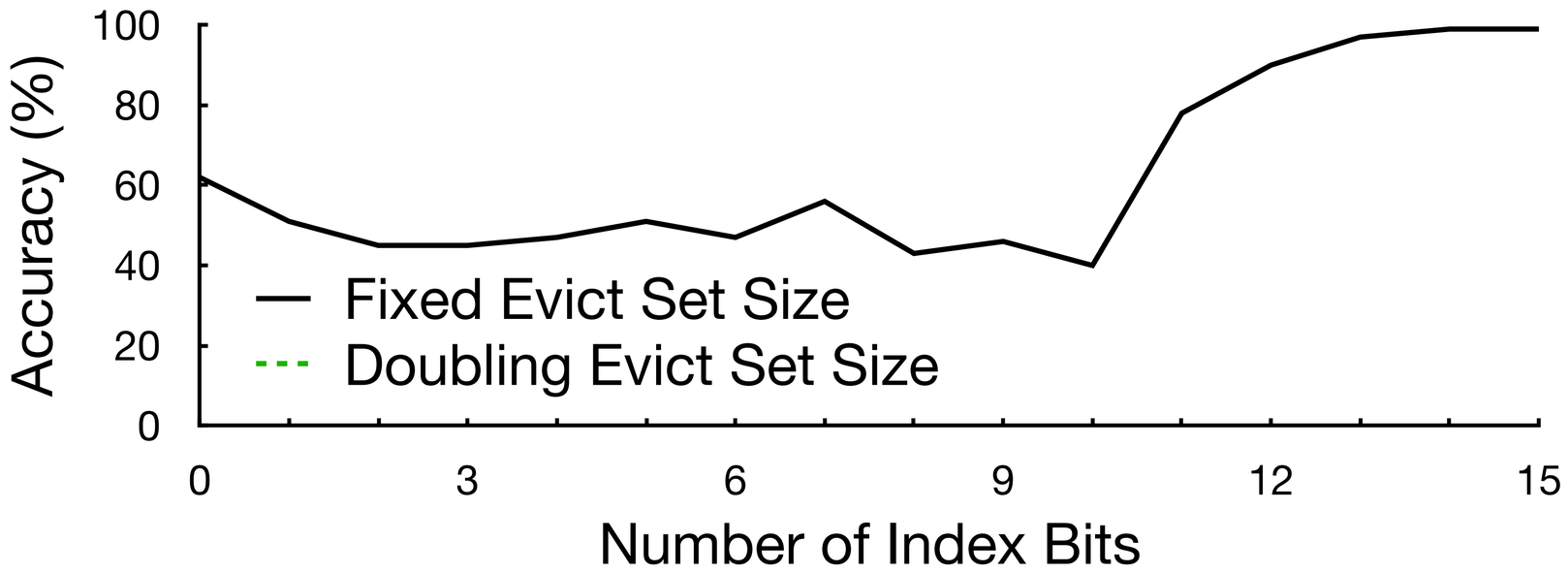
1/500

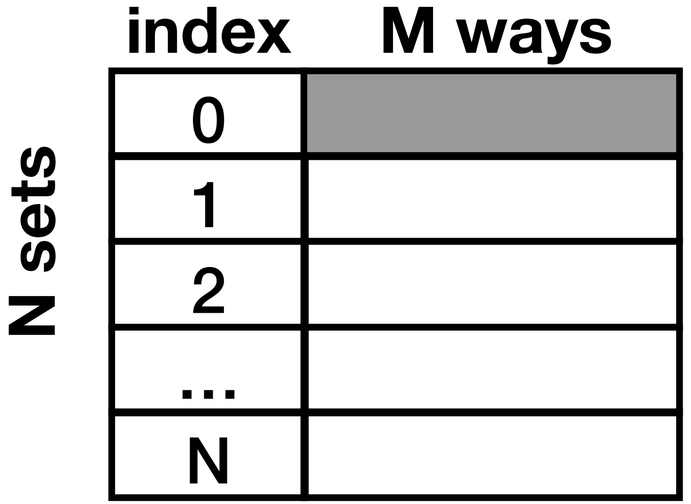
50 us

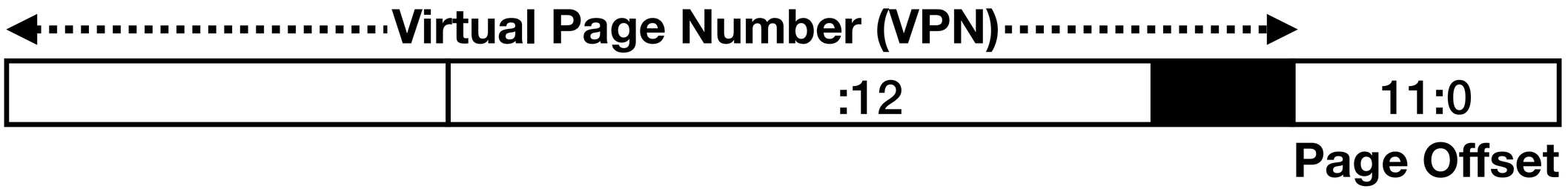
Effect of Number of Index Bits



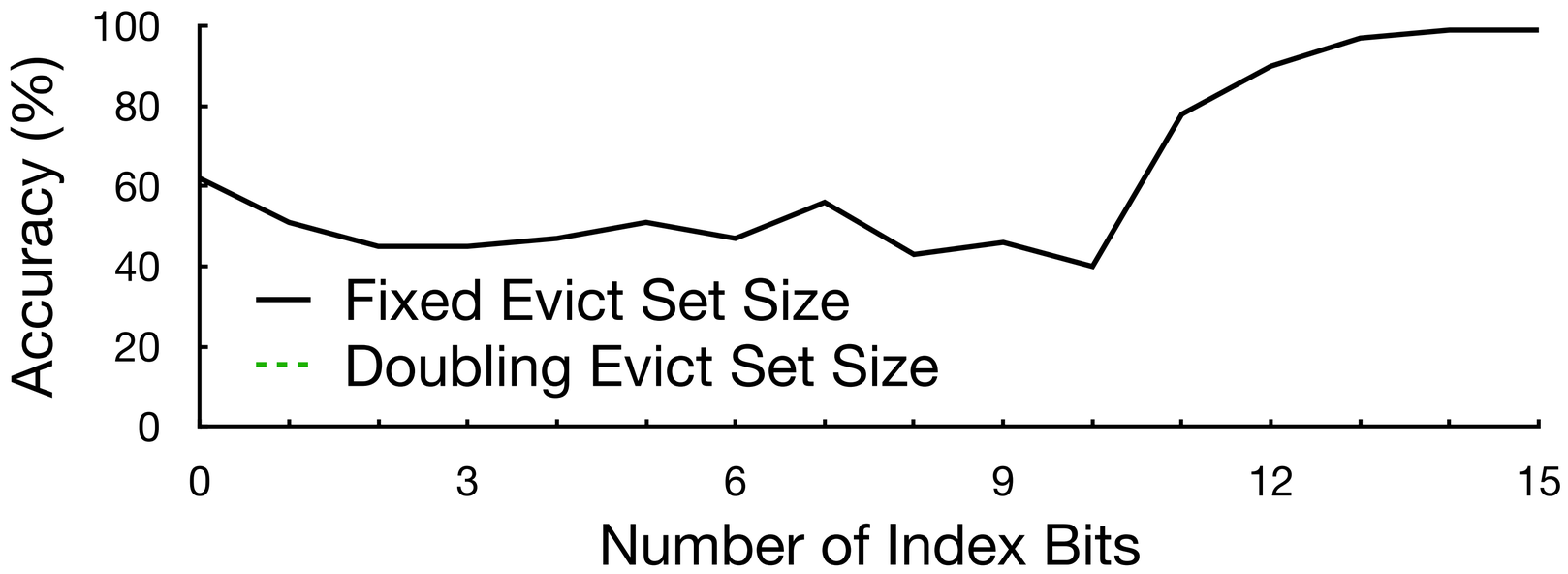
Effect of Number of Index Bits

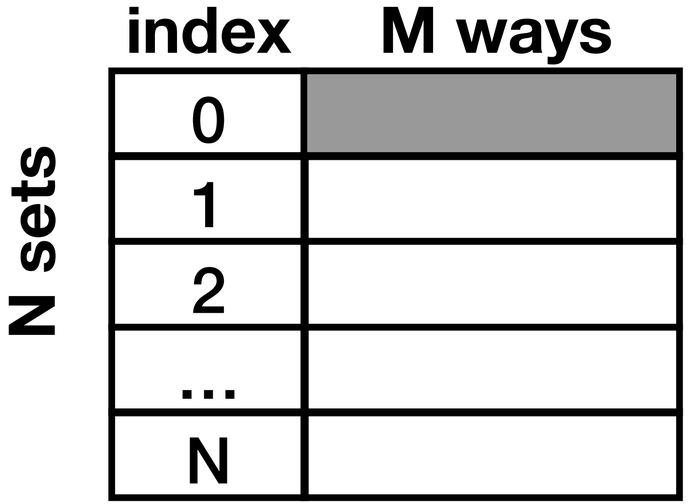


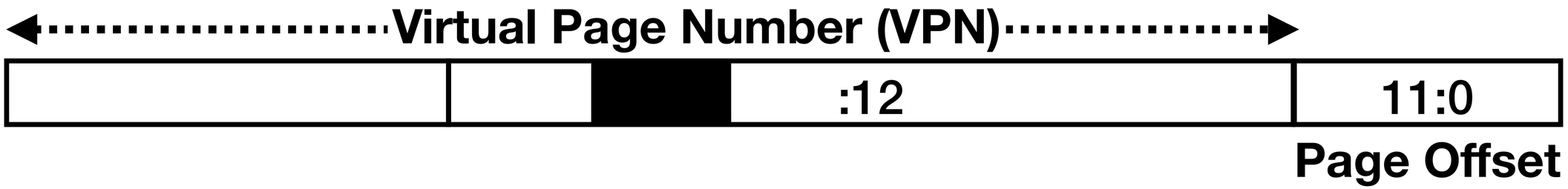




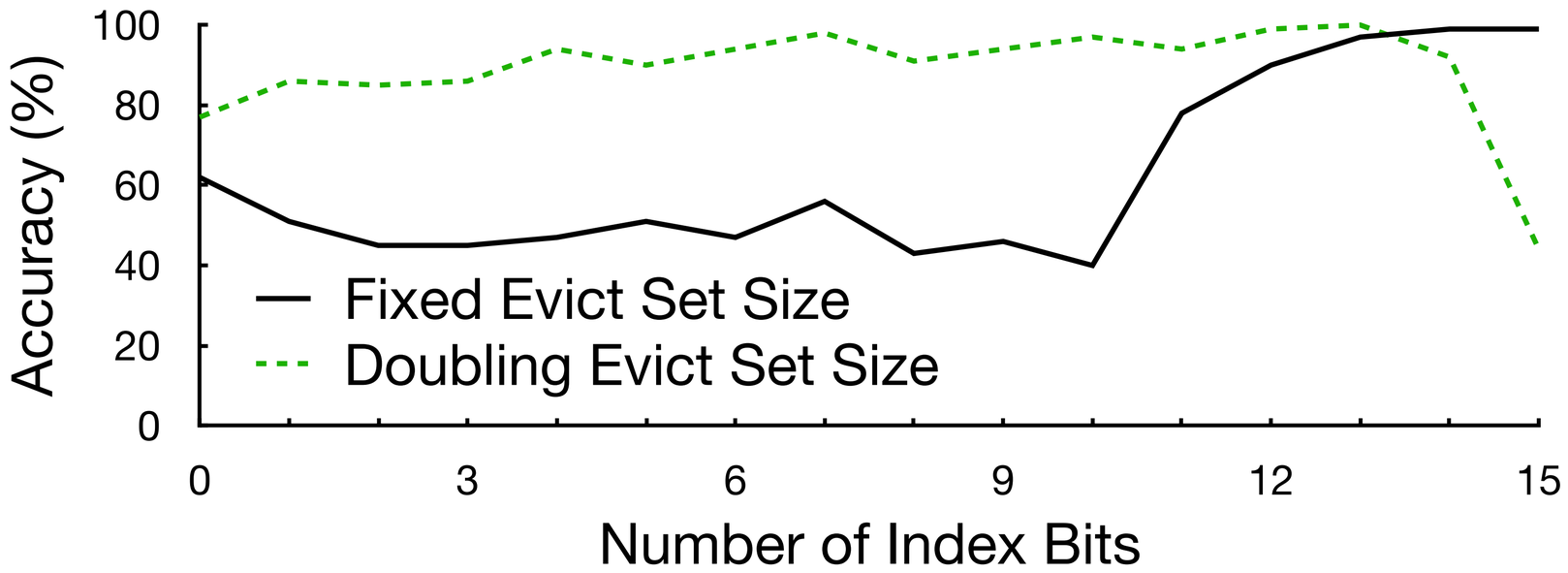
Effect of Number of Index Bits

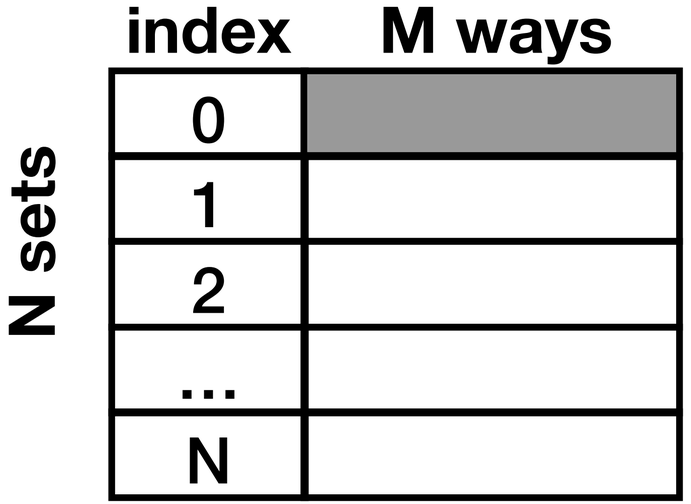


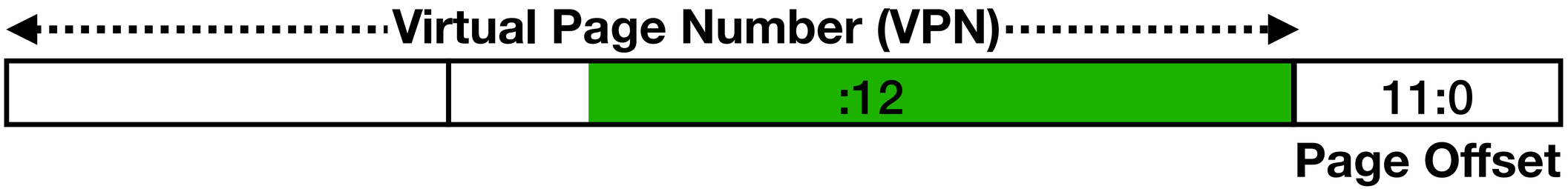




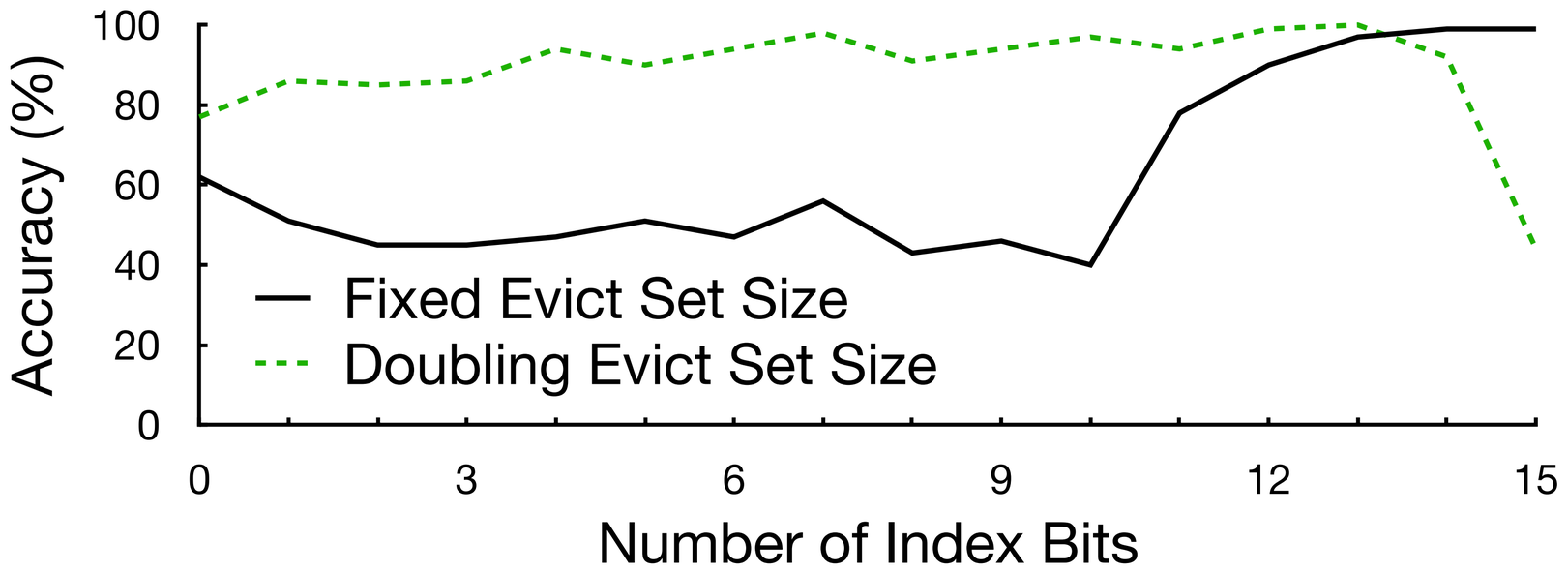
Effect of Number of Index Bits

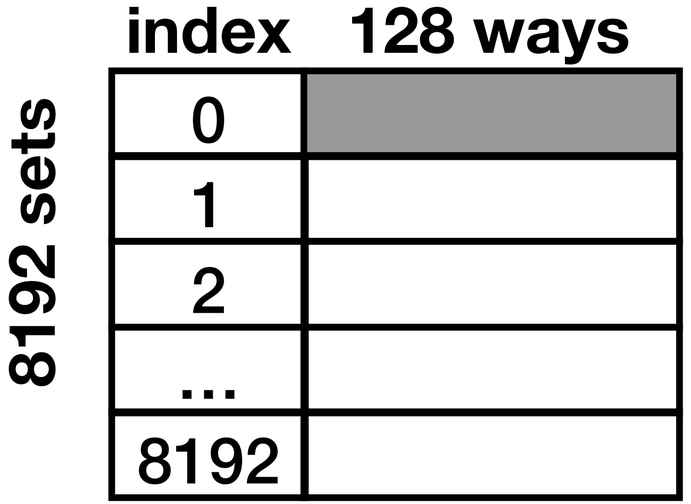


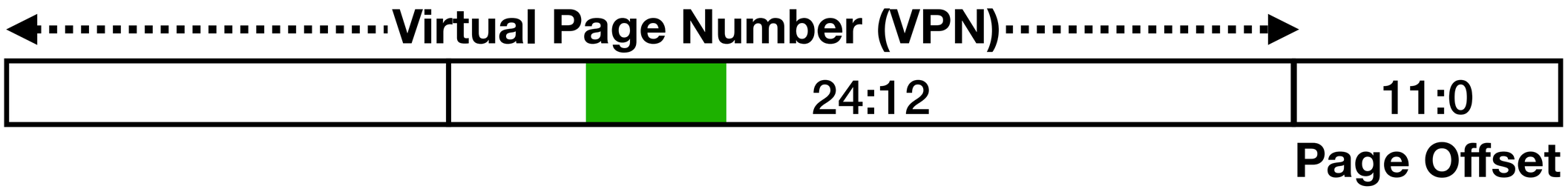




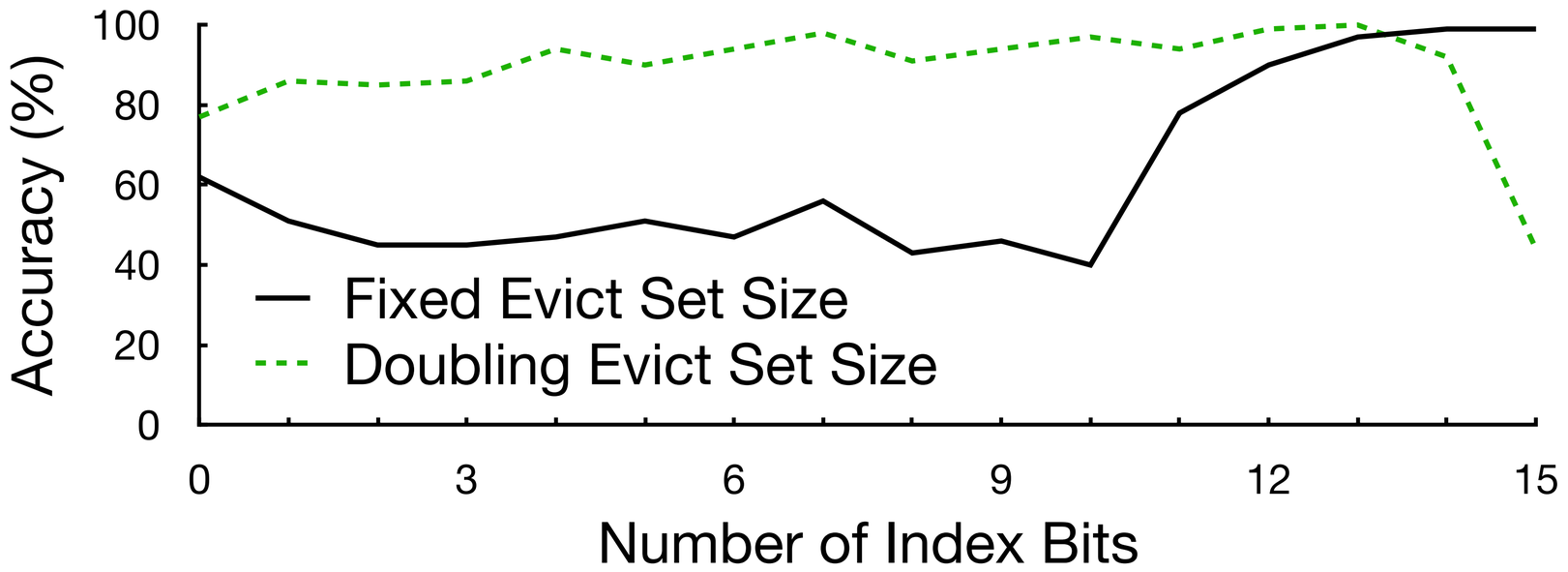
Effect of Number of Index Bits

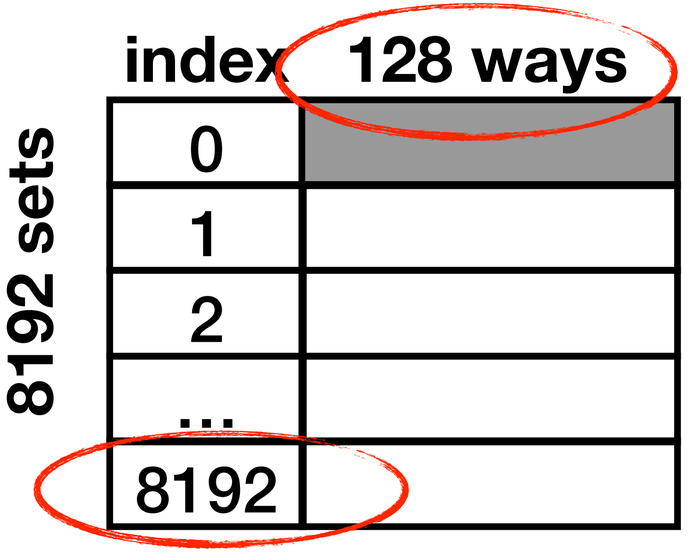


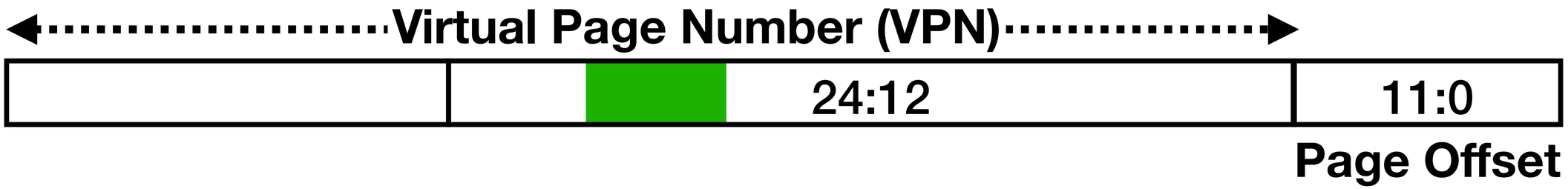




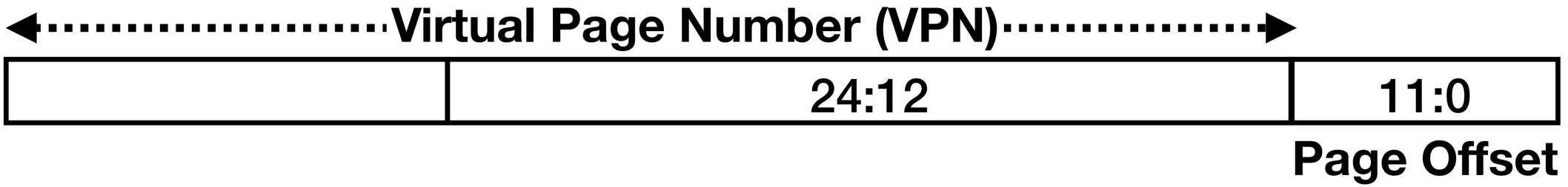
Effect of Number of Index Bits

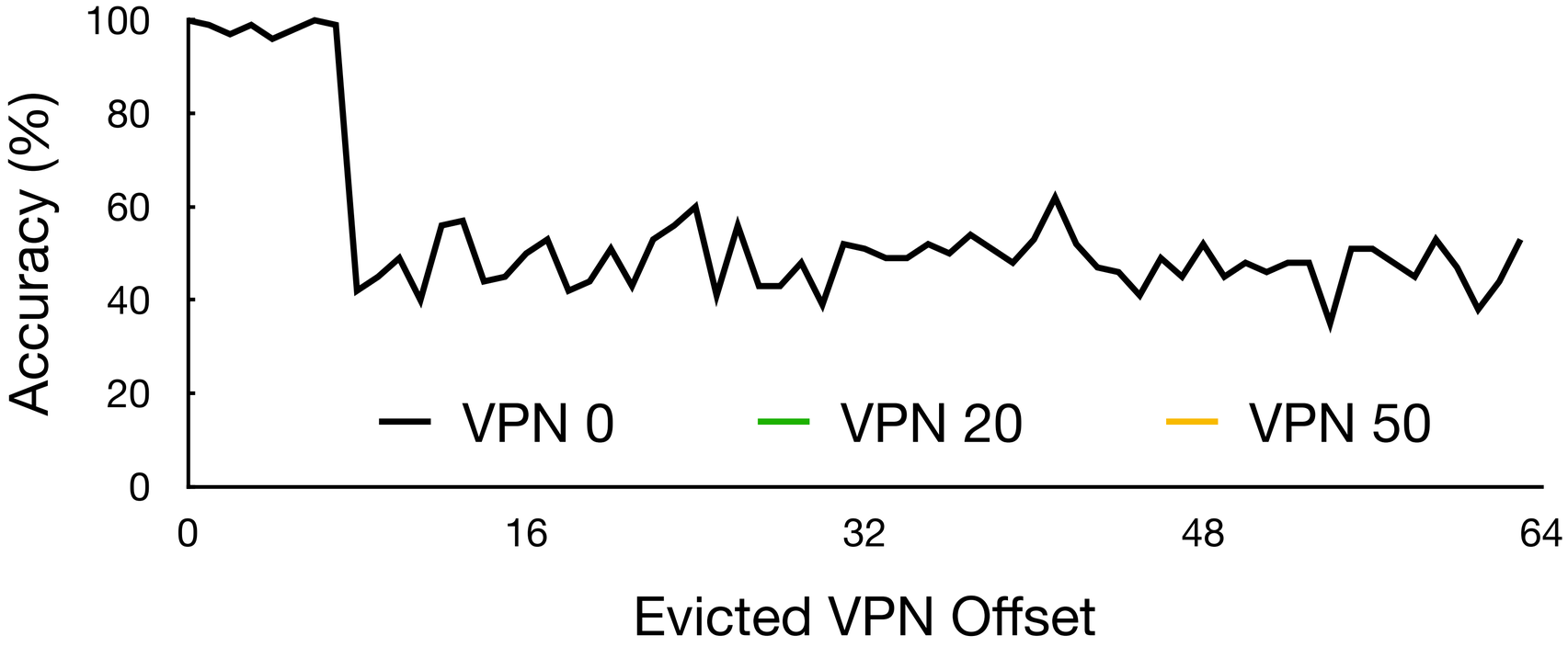


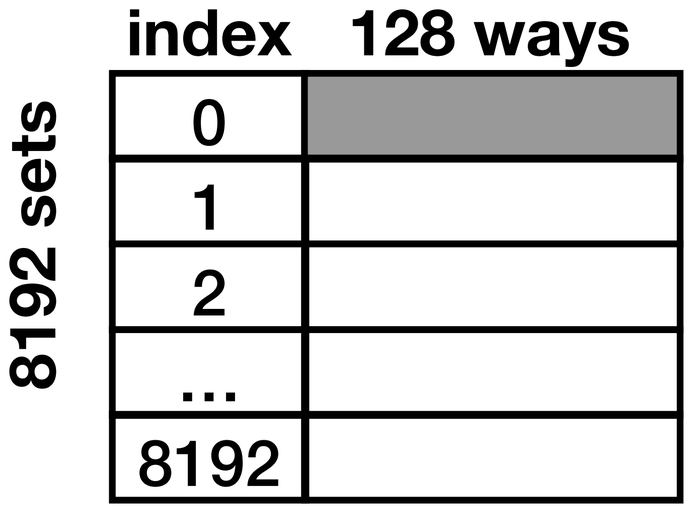




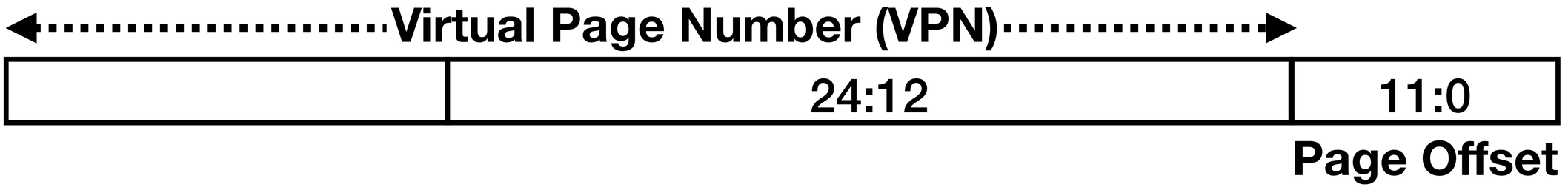
Effect of Eviction Set Offset

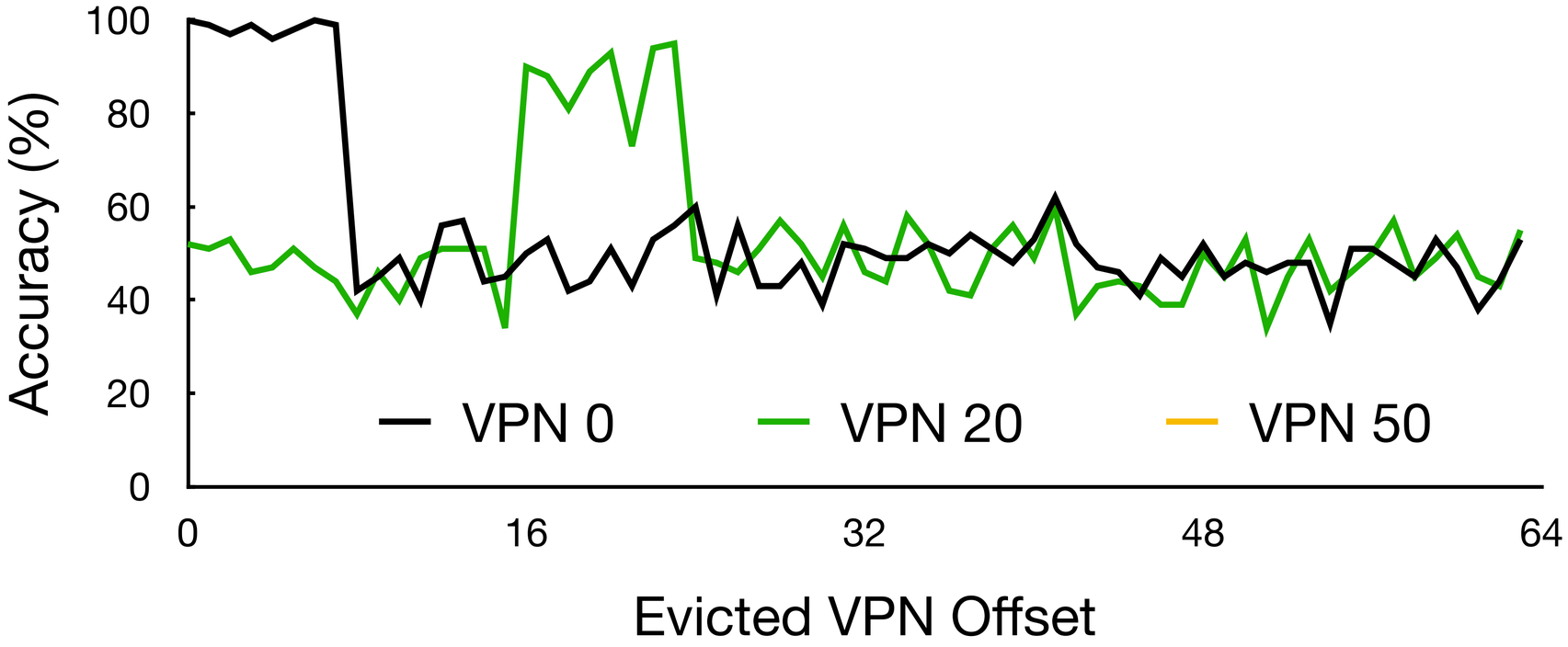


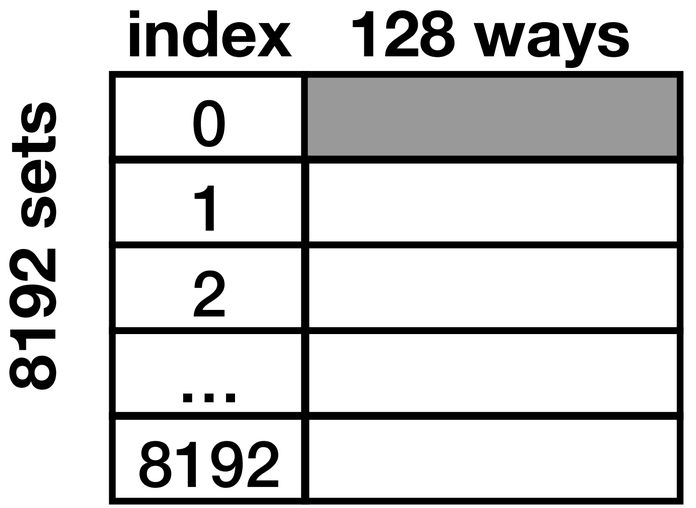




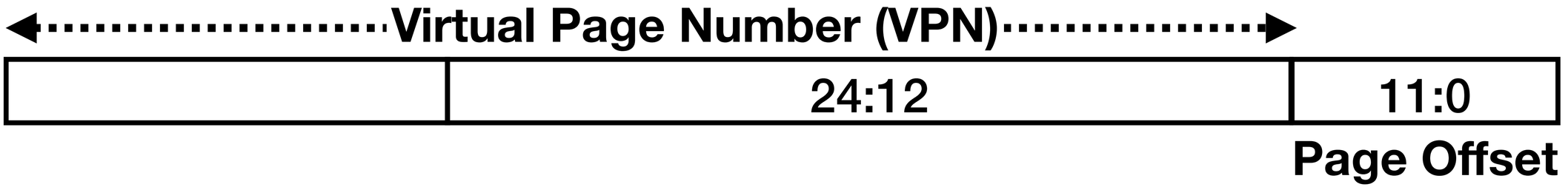
Effect of Eviction Set Offset

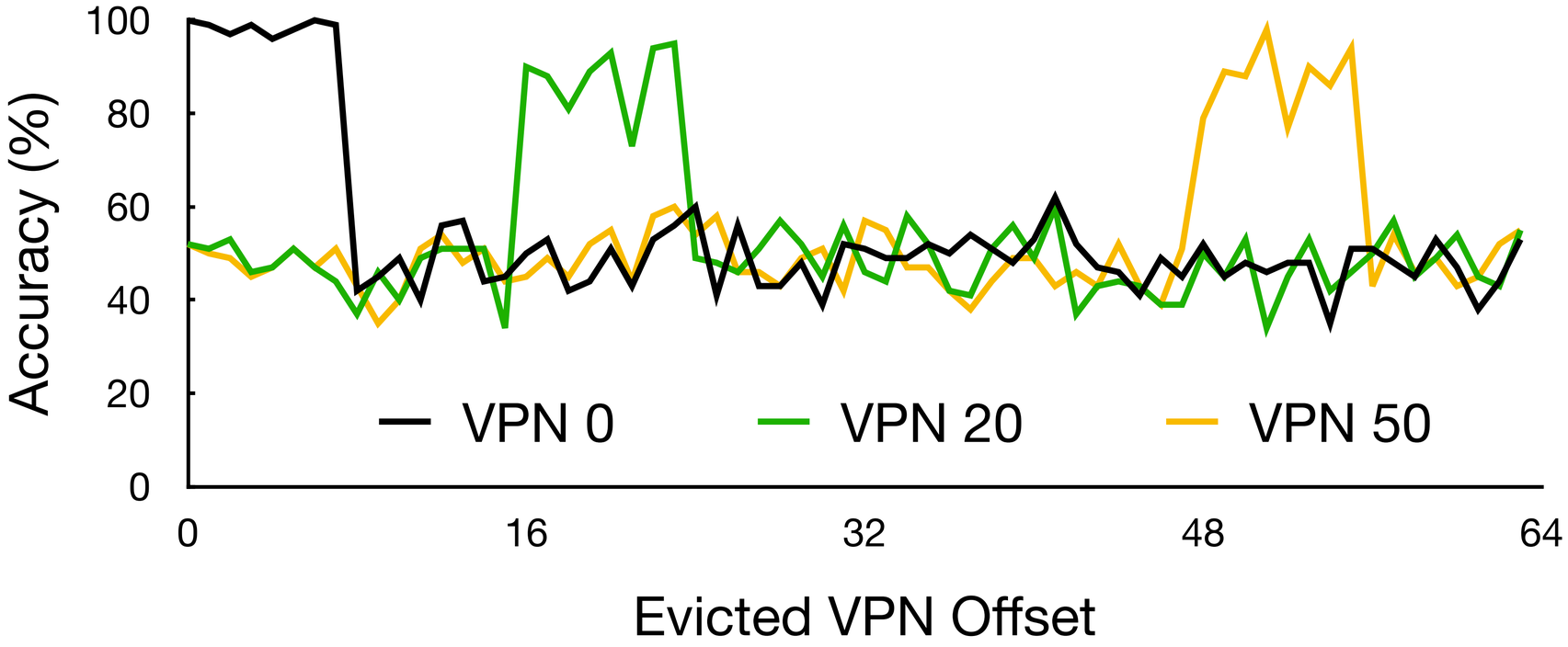


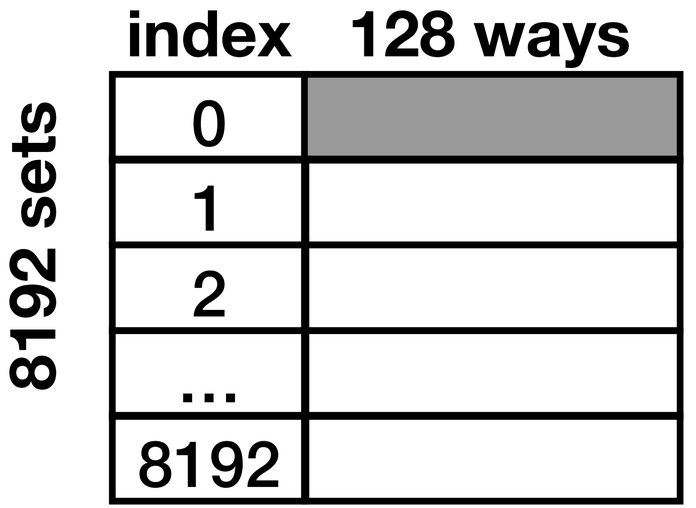




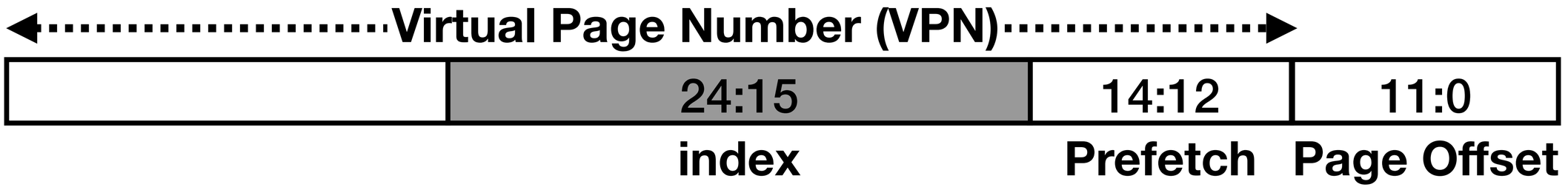
Effect of Eviction Set Offset

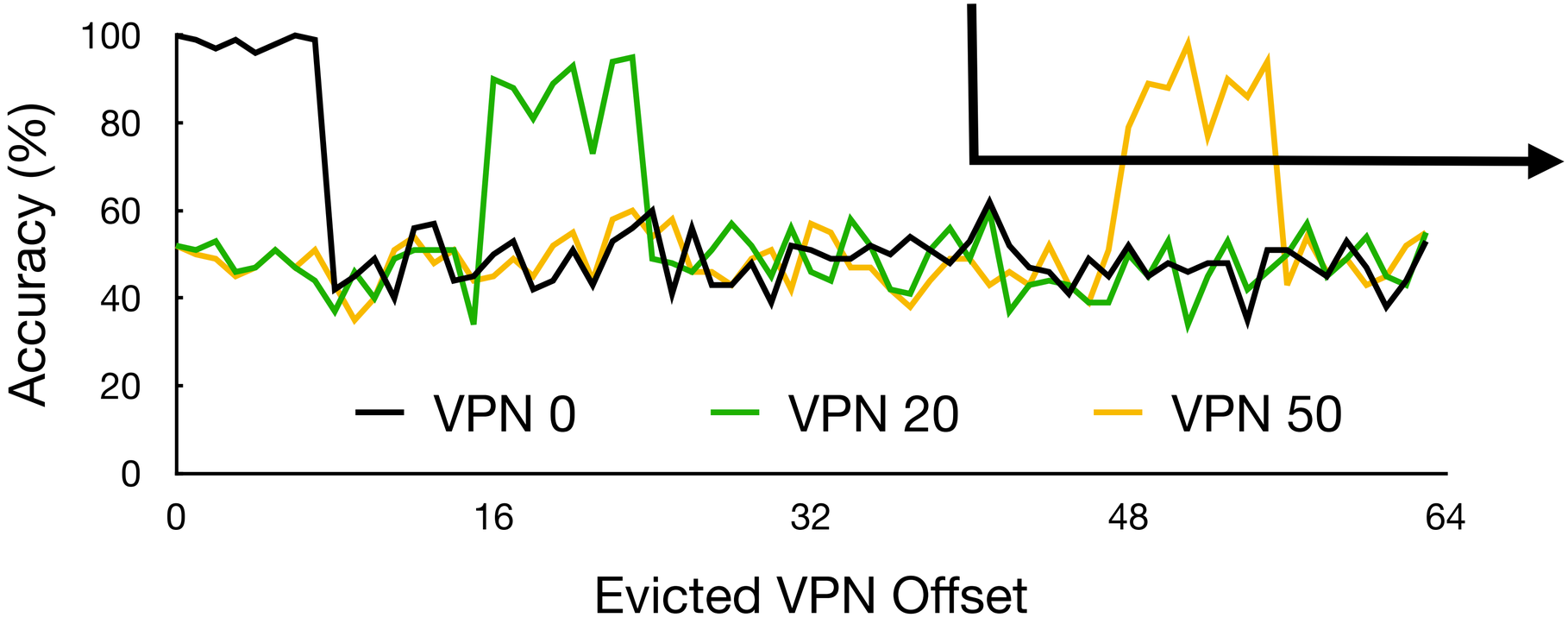


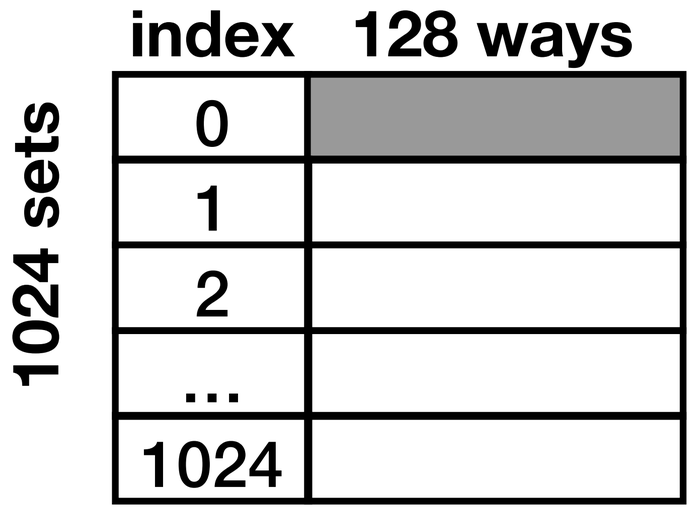




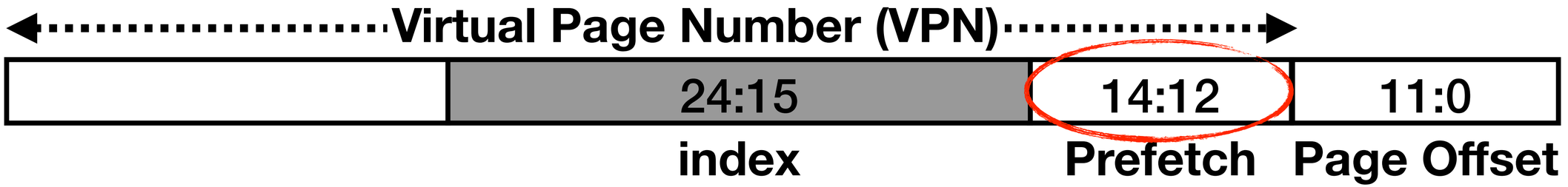
Effect of Eviction Set Offset

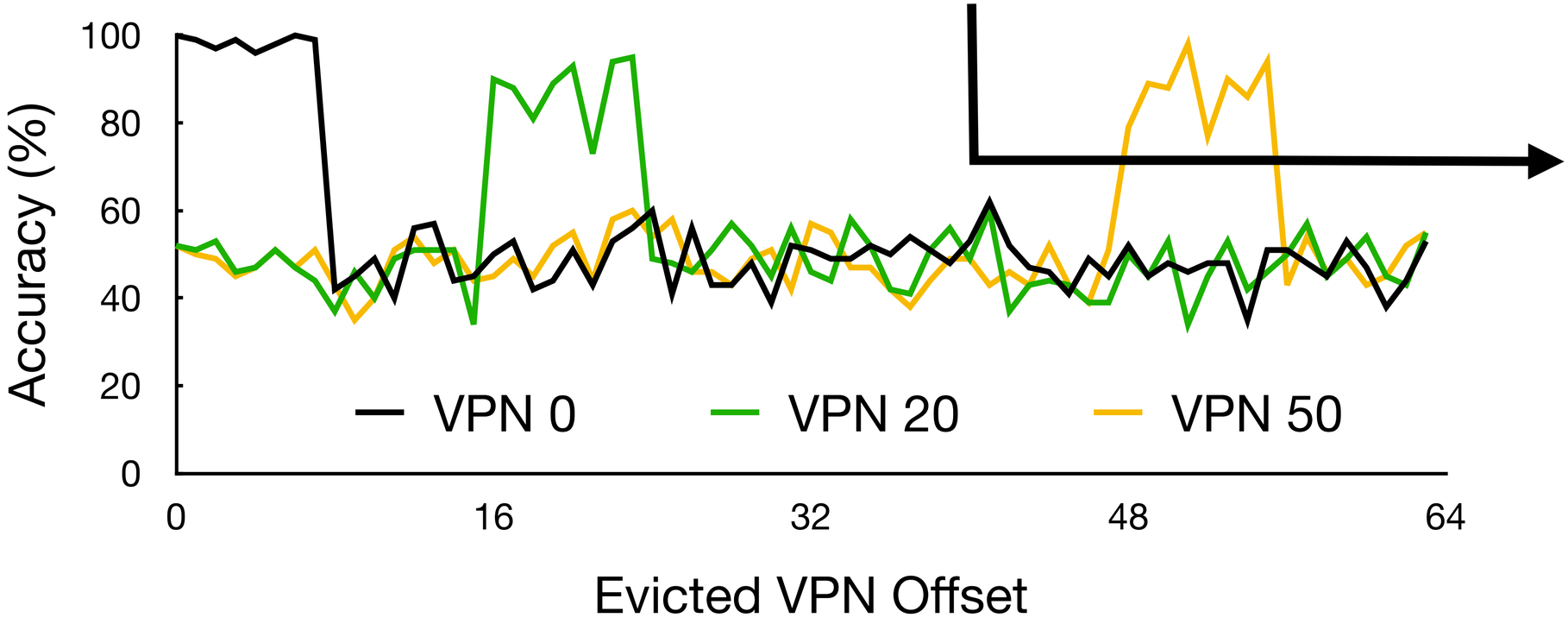


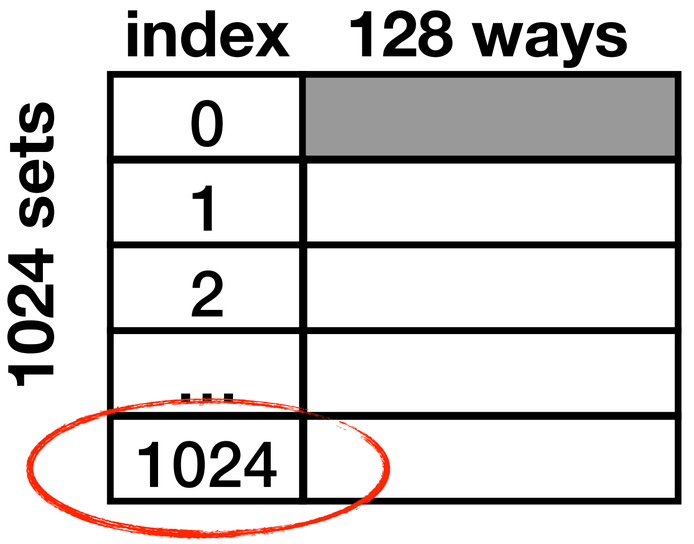




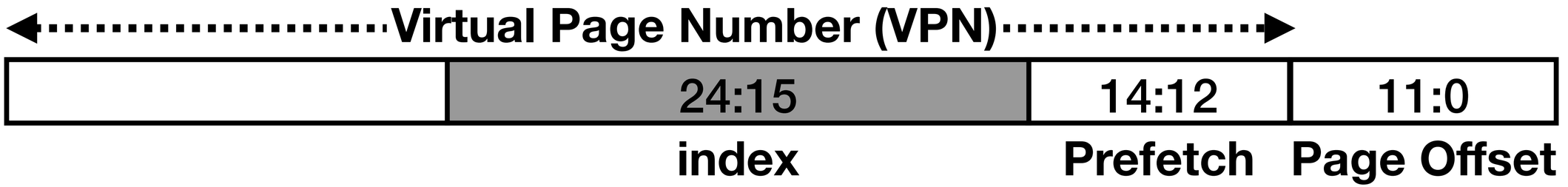
Effect of Eviction Set Offset

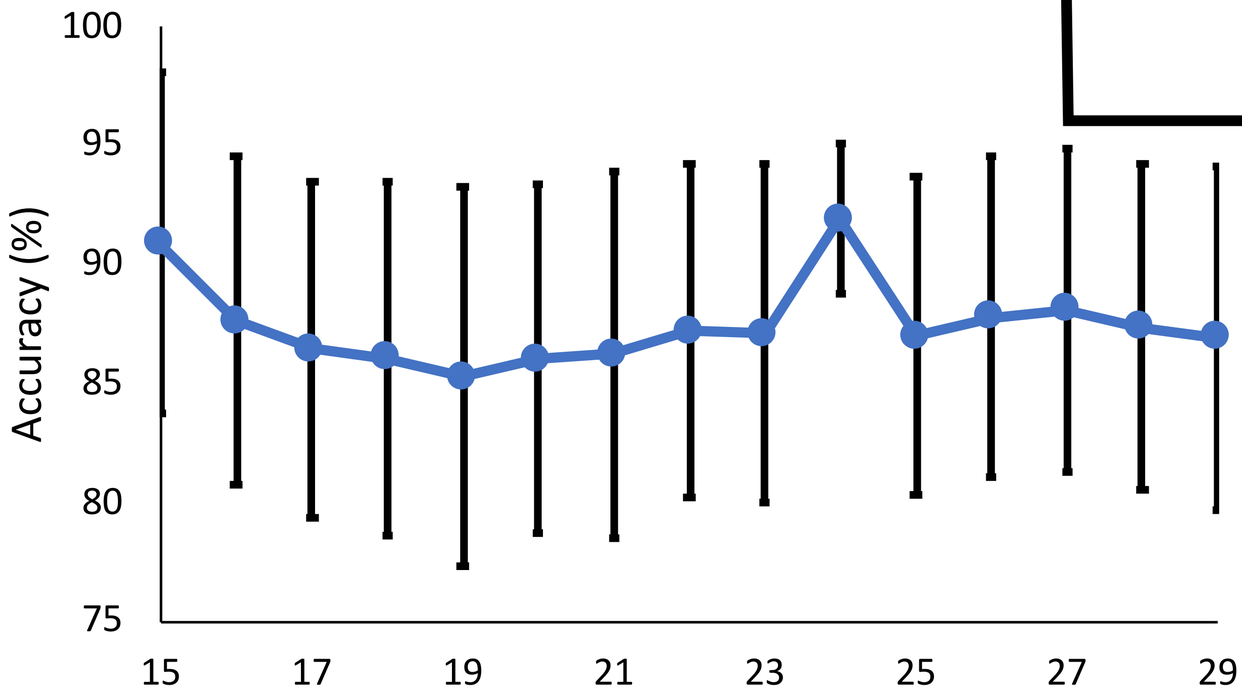


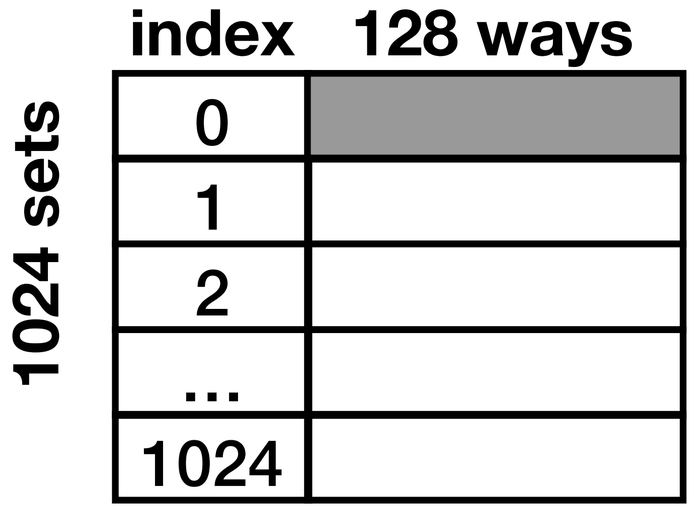




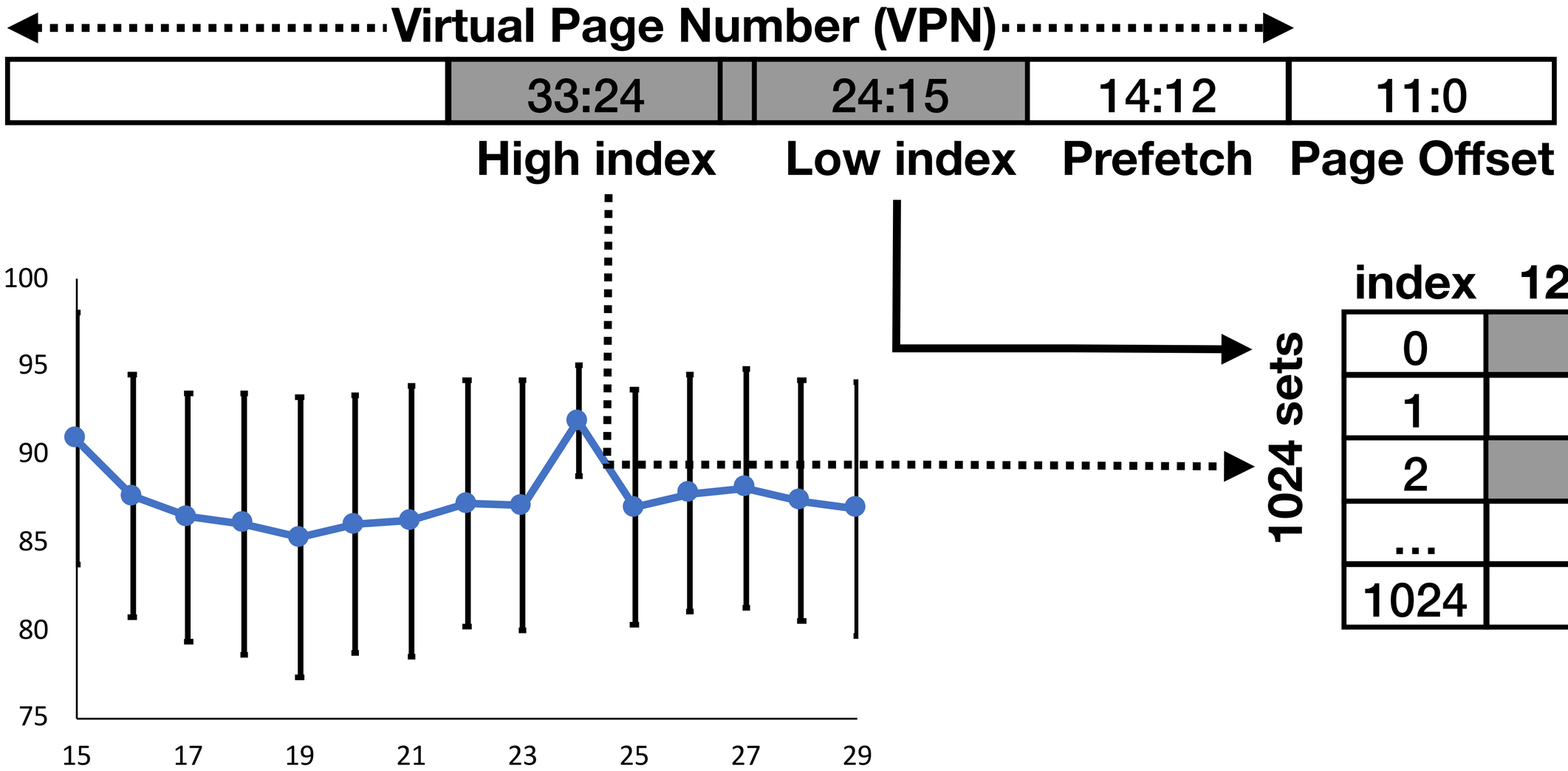
Effect of Secondary Index



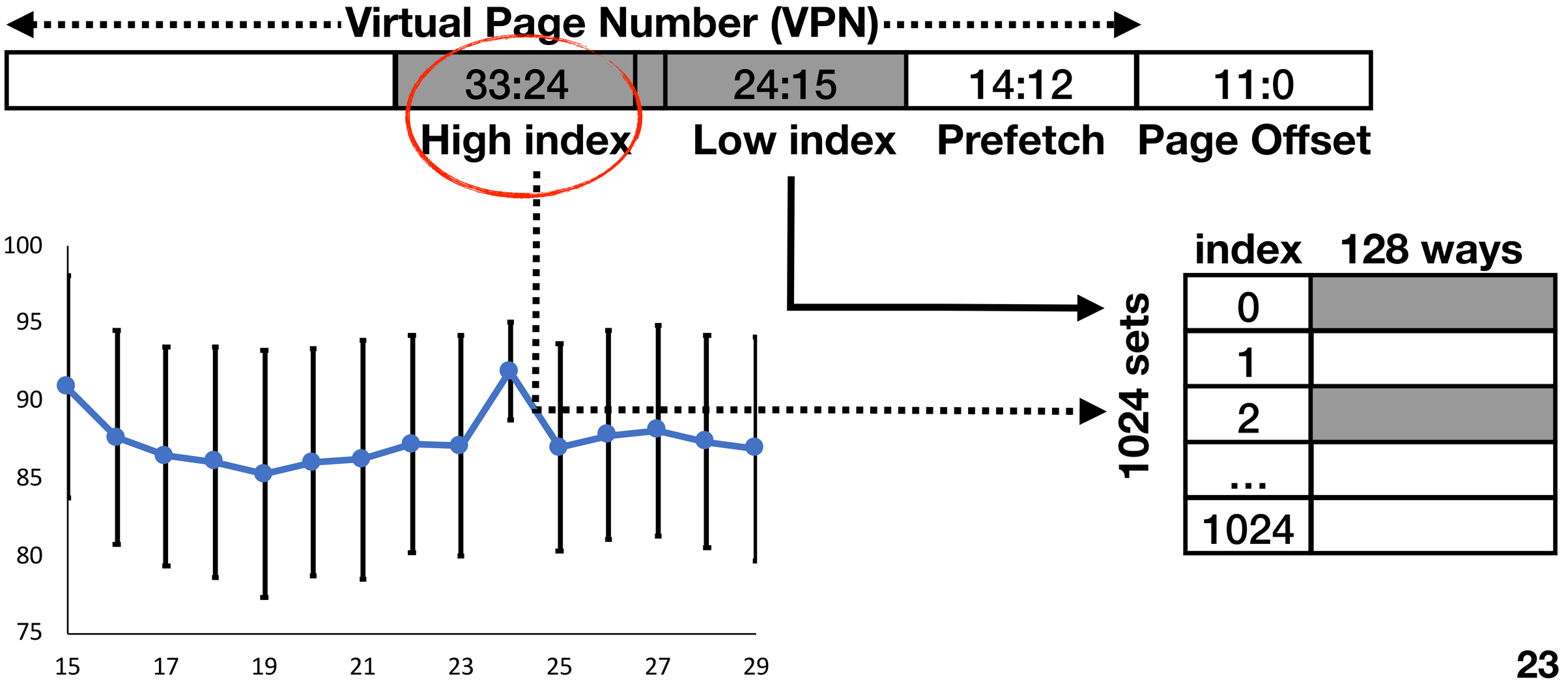




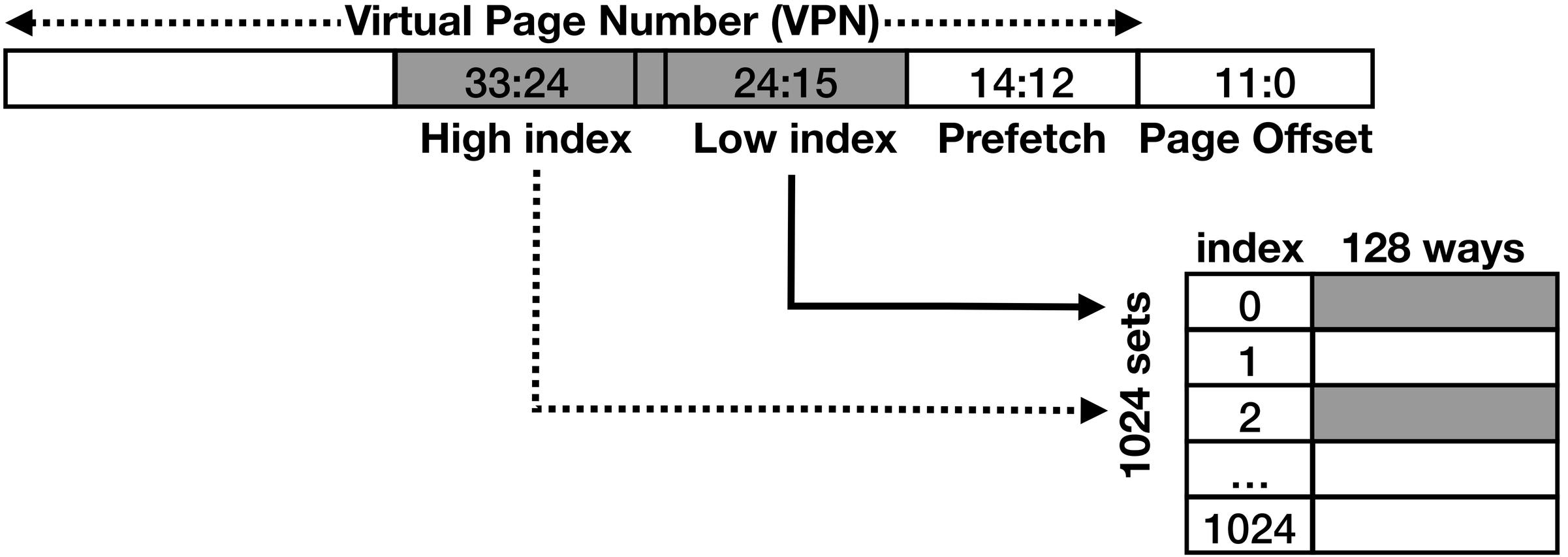
Effect of Secondary Index



Effect of Secondary Index

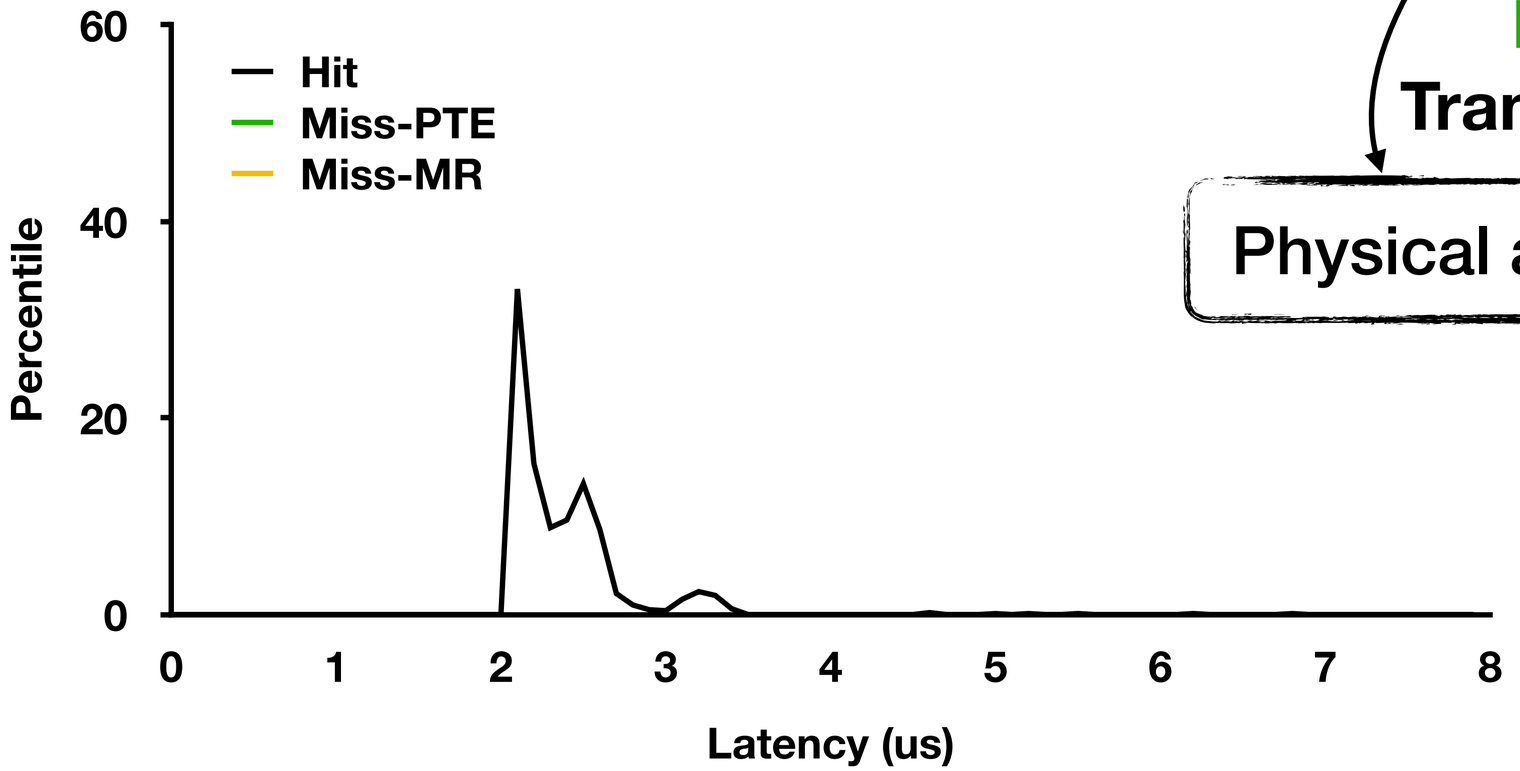


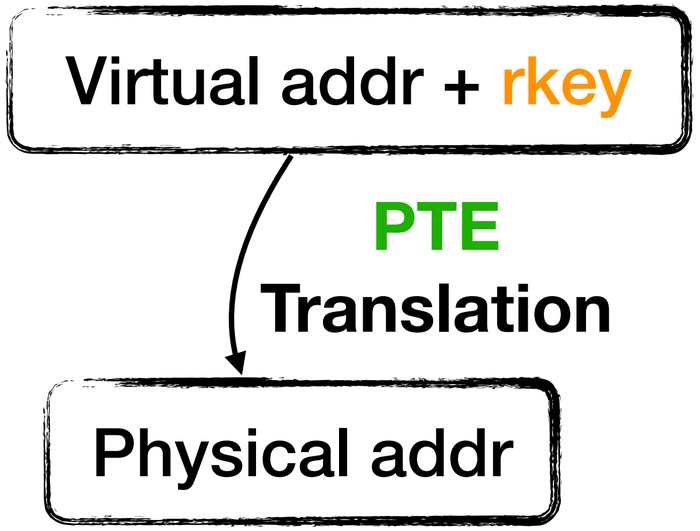
SRAM Cache Architecture



Evict MR and Evict PTE

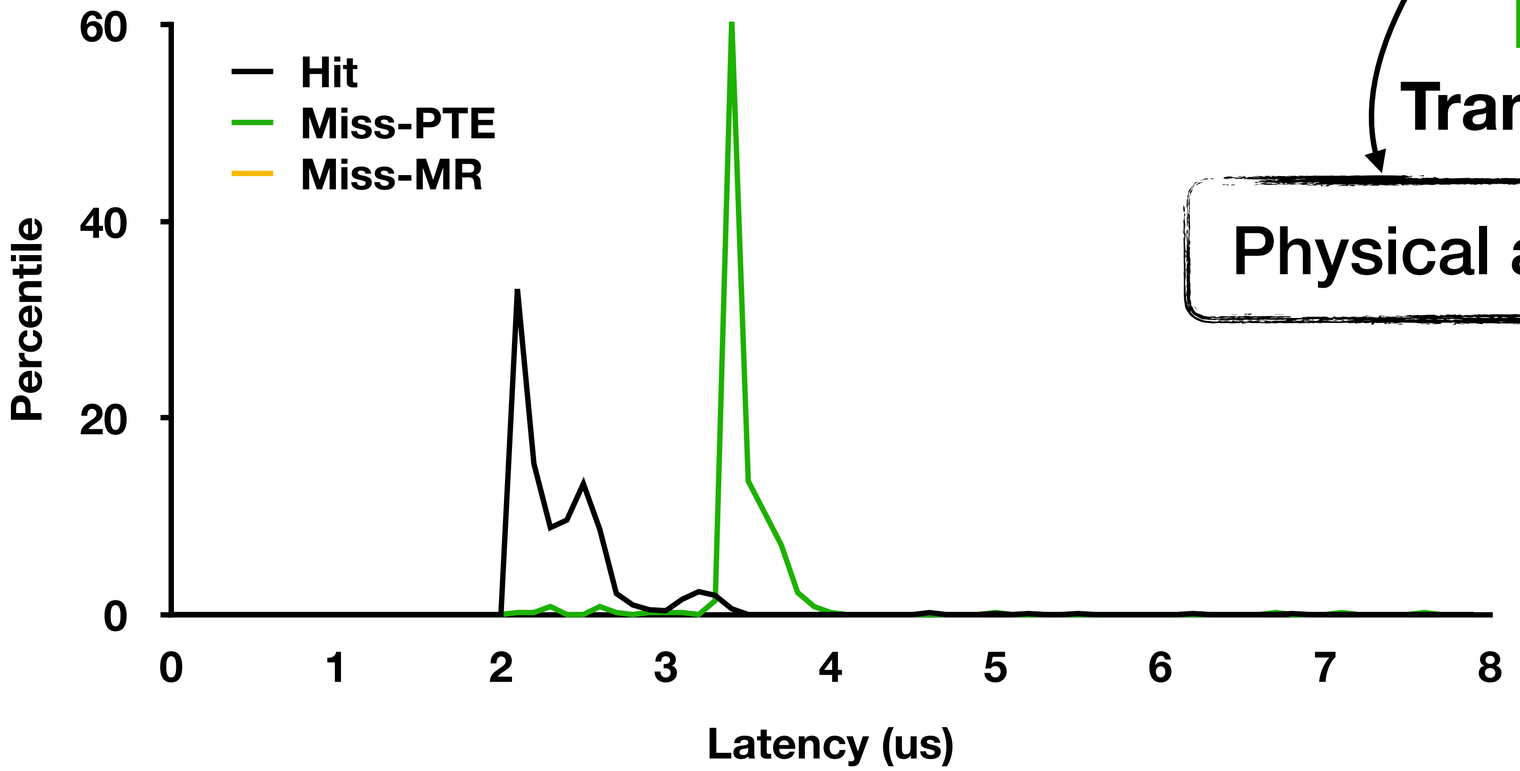
- MR and PTE
- ConnectX-4 - READ 1KB

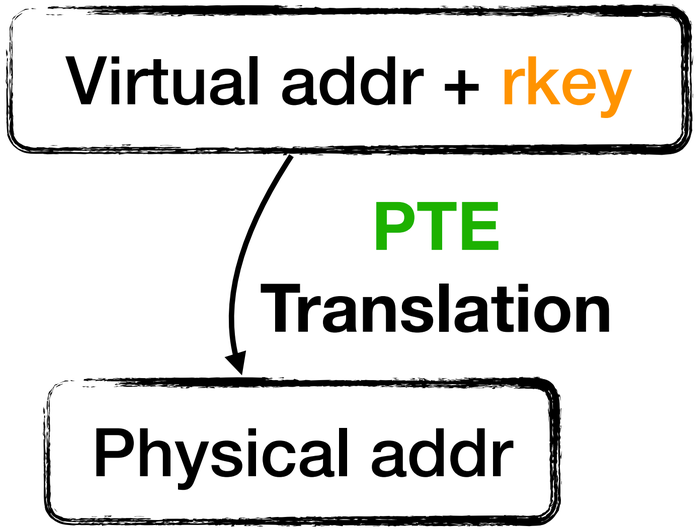




Evict MR and Evict PTE

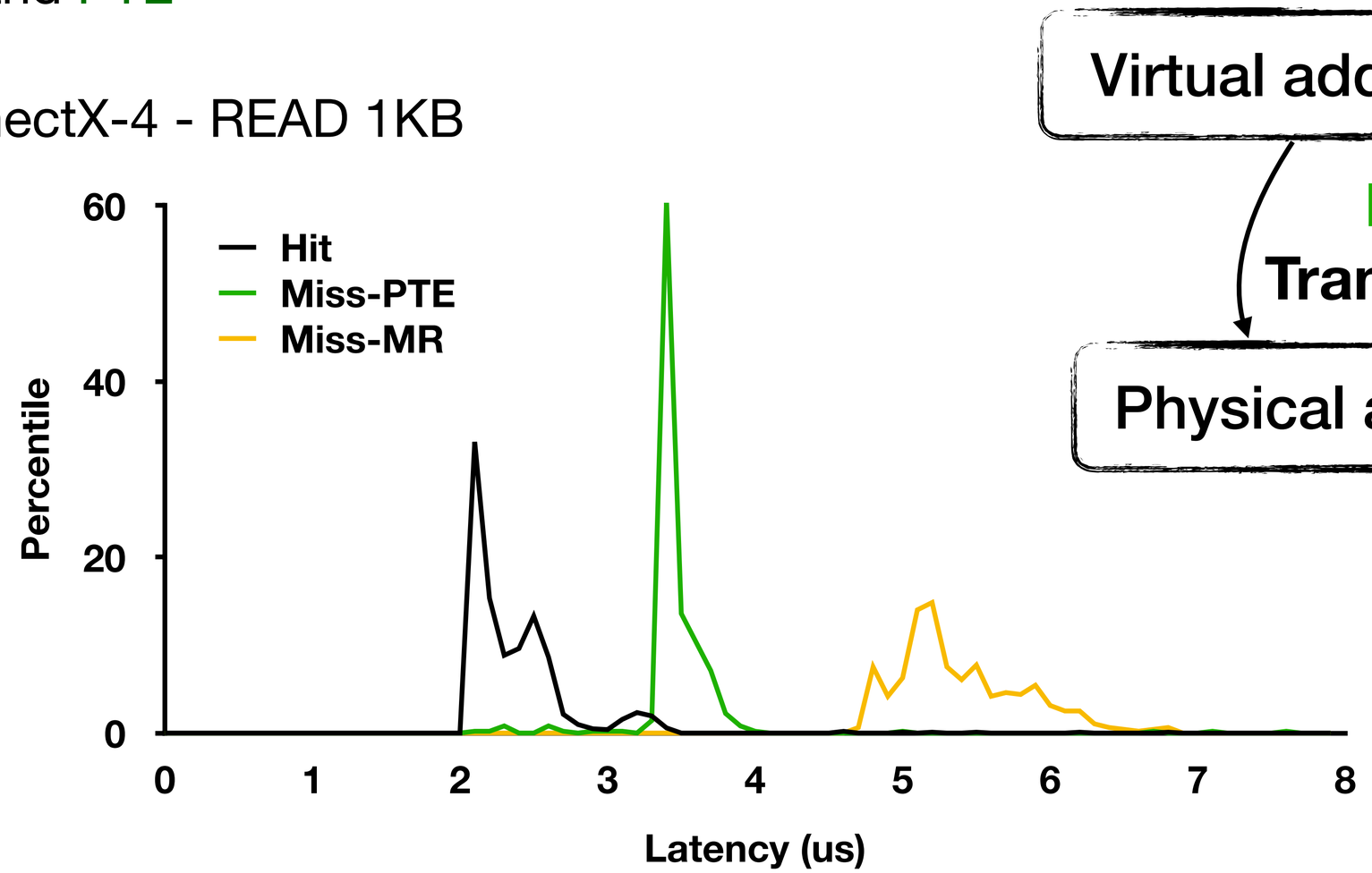
- MR and PTE
- ConnectX-4 - READ 1KB



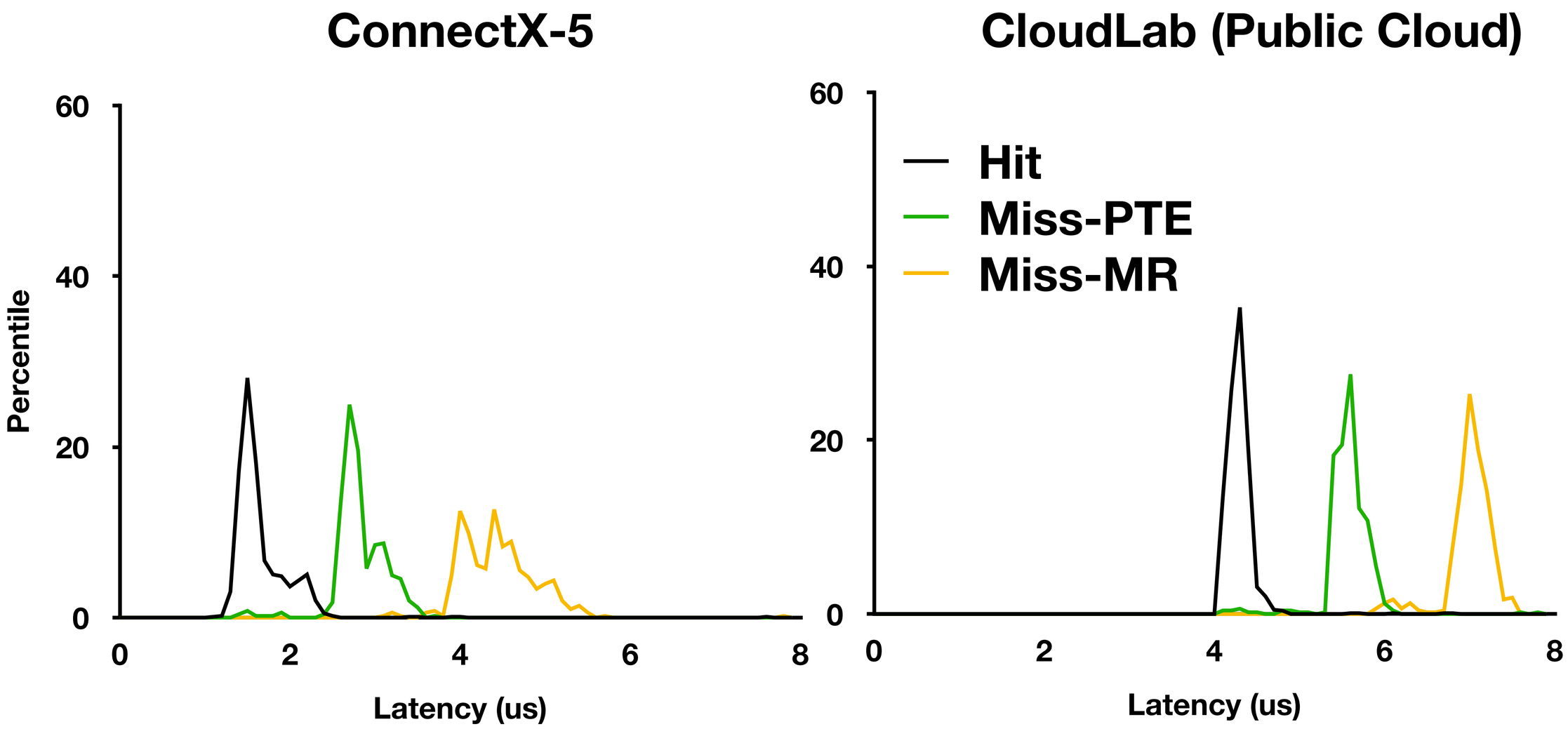


Evict MR and Evict PTE

- MR and PTE
- ConnectX-4 - READ 1KB

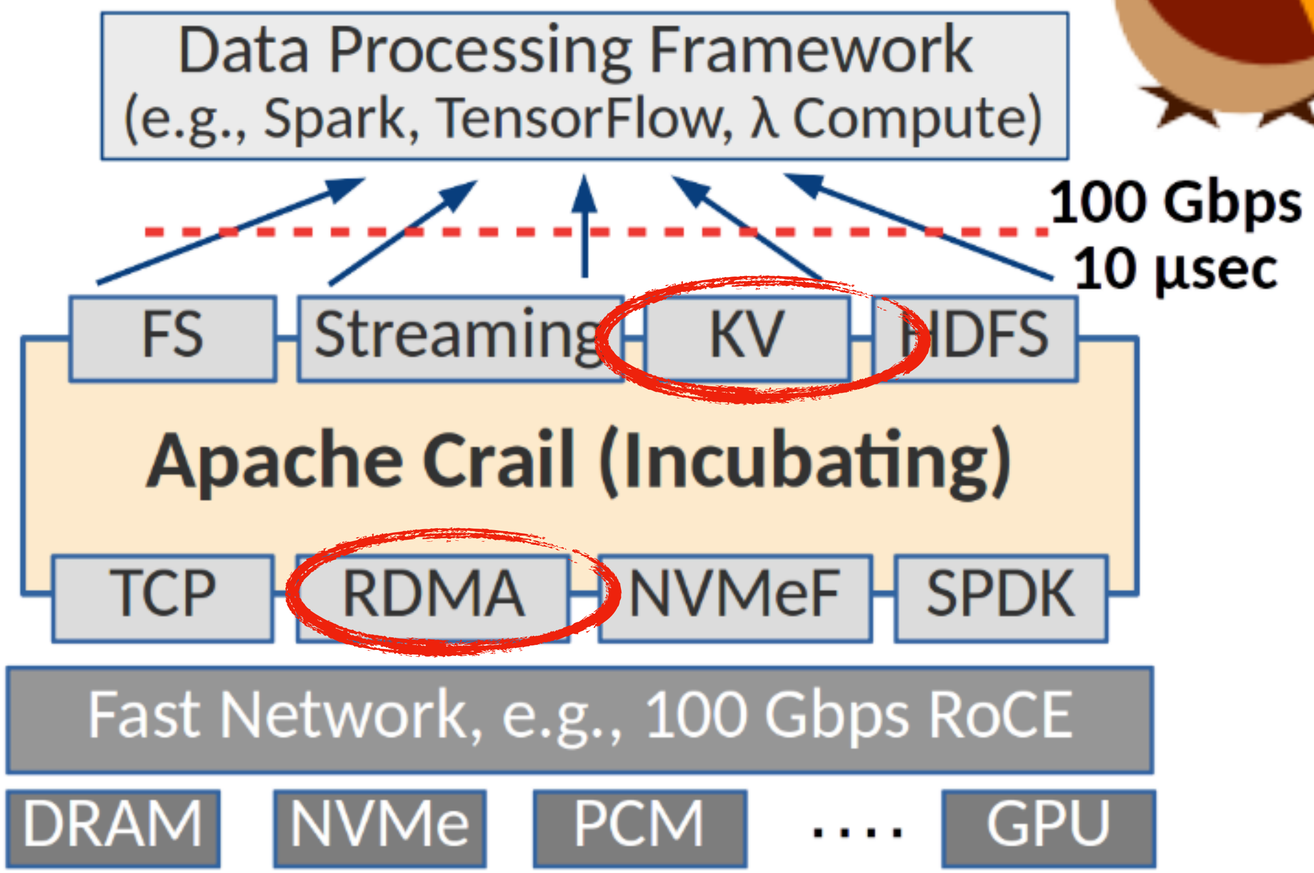


Timing Difference - Misc



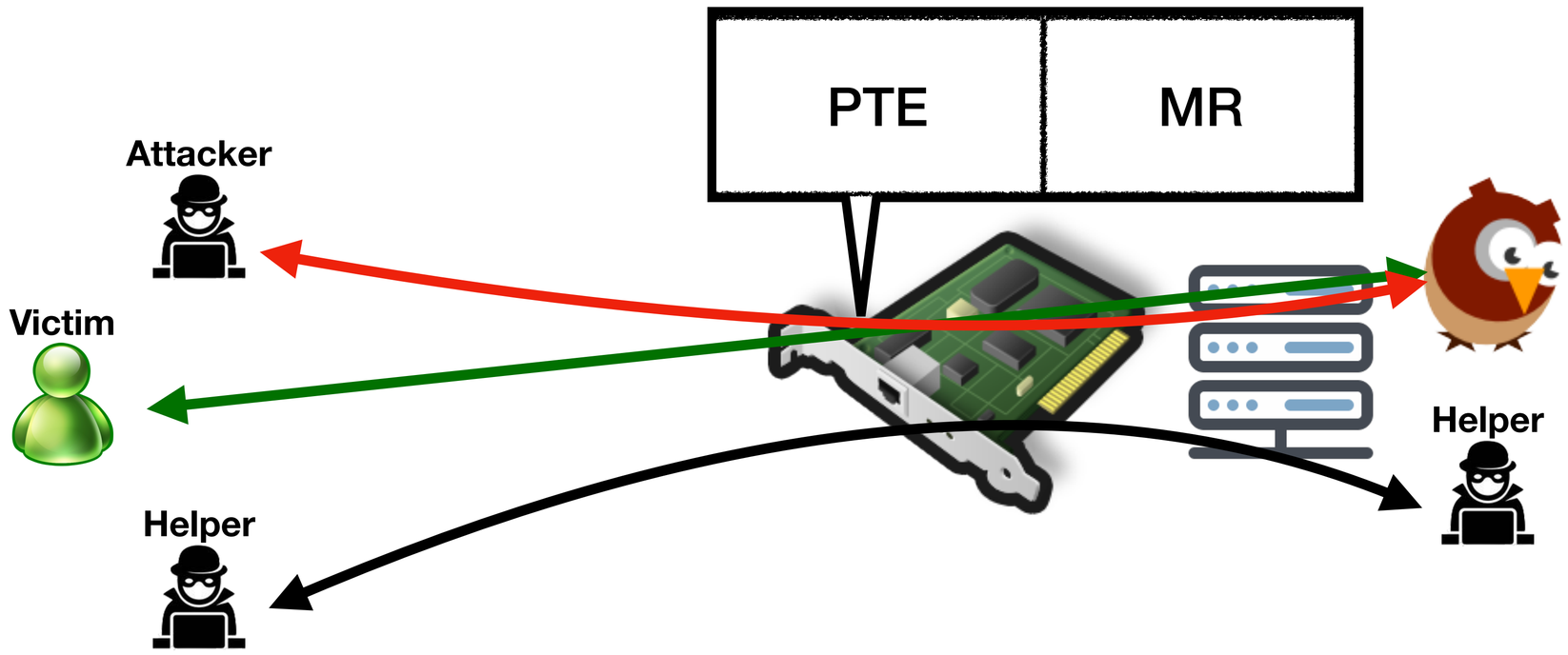
ATTACK REAL Application - Apache Crail

- Originated from a research project at IBM
- Java-based storage platform
- **Key-value store**
- **RDMA**



How to Attack Crail?

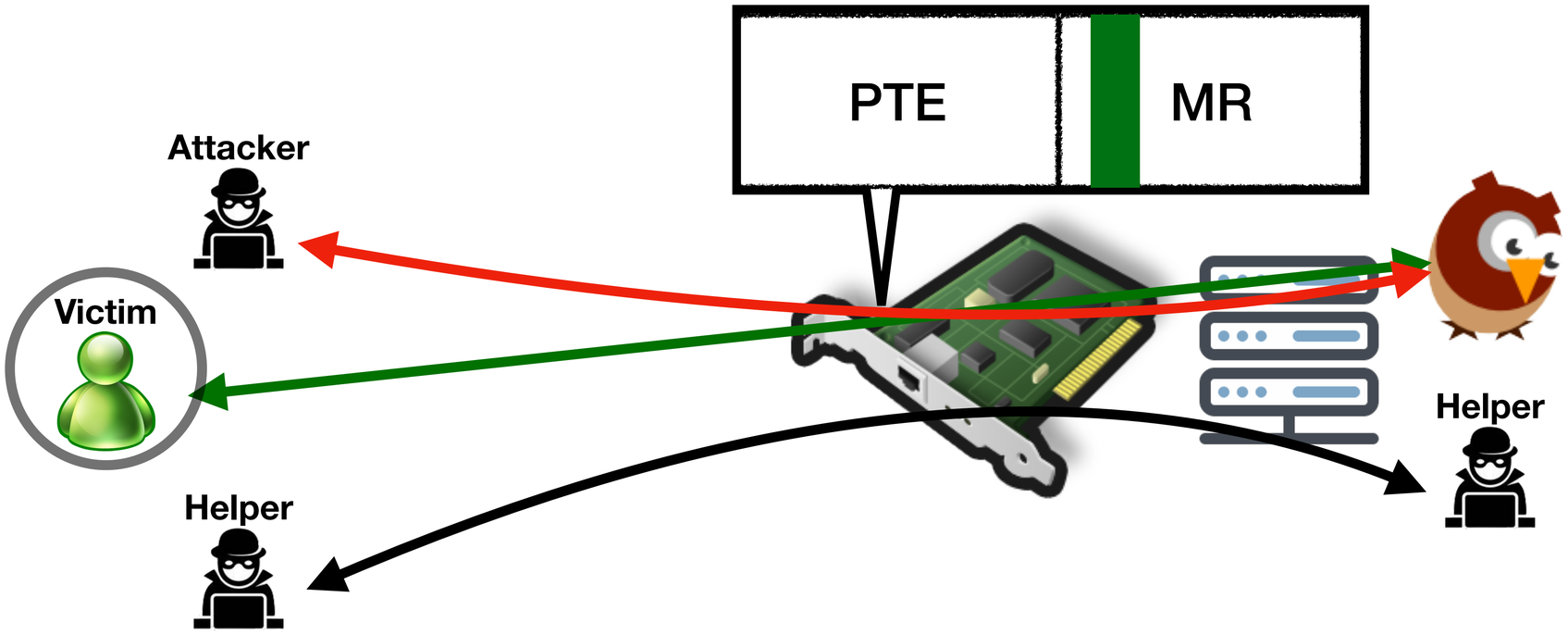
- MR-based



Need a Helper process

How to Attack Crail?

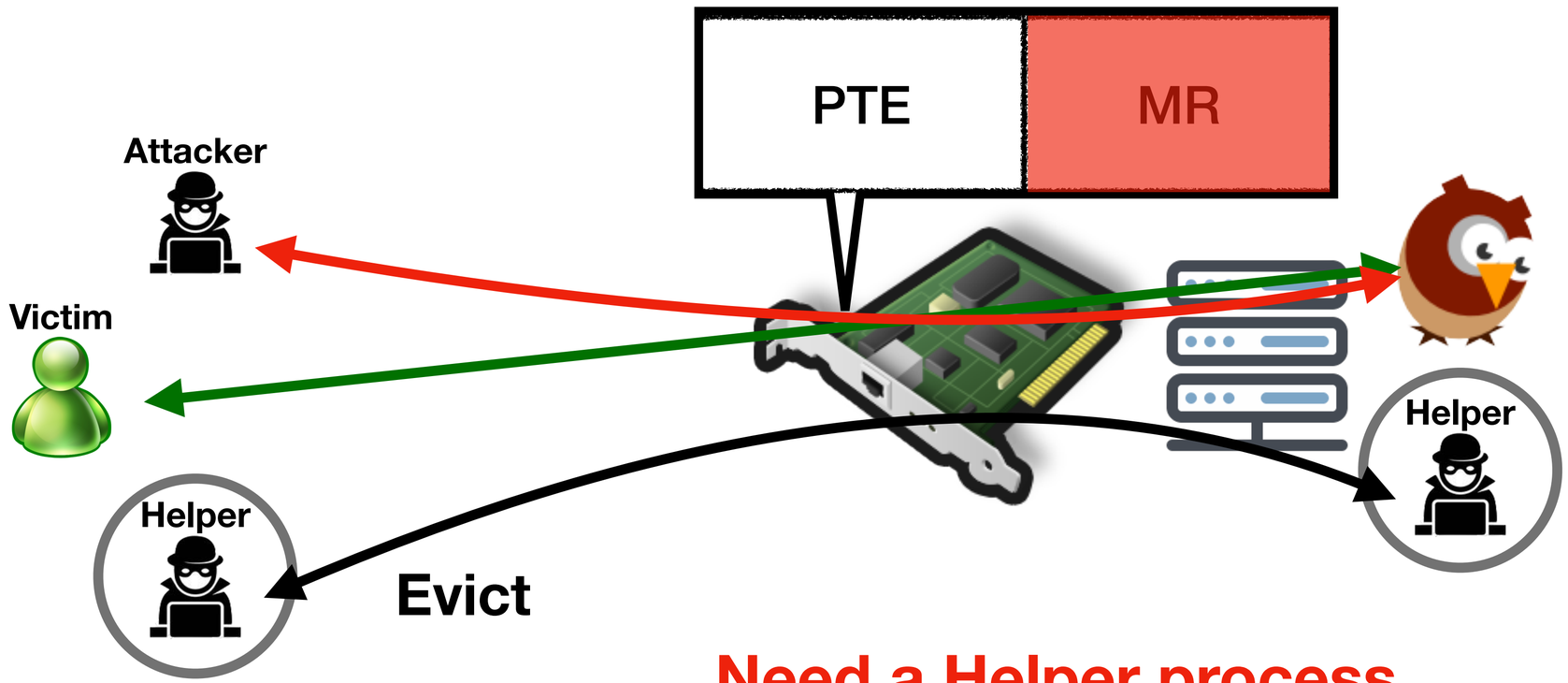
- MR-based



Need a Helper process

How to Attack Crail?

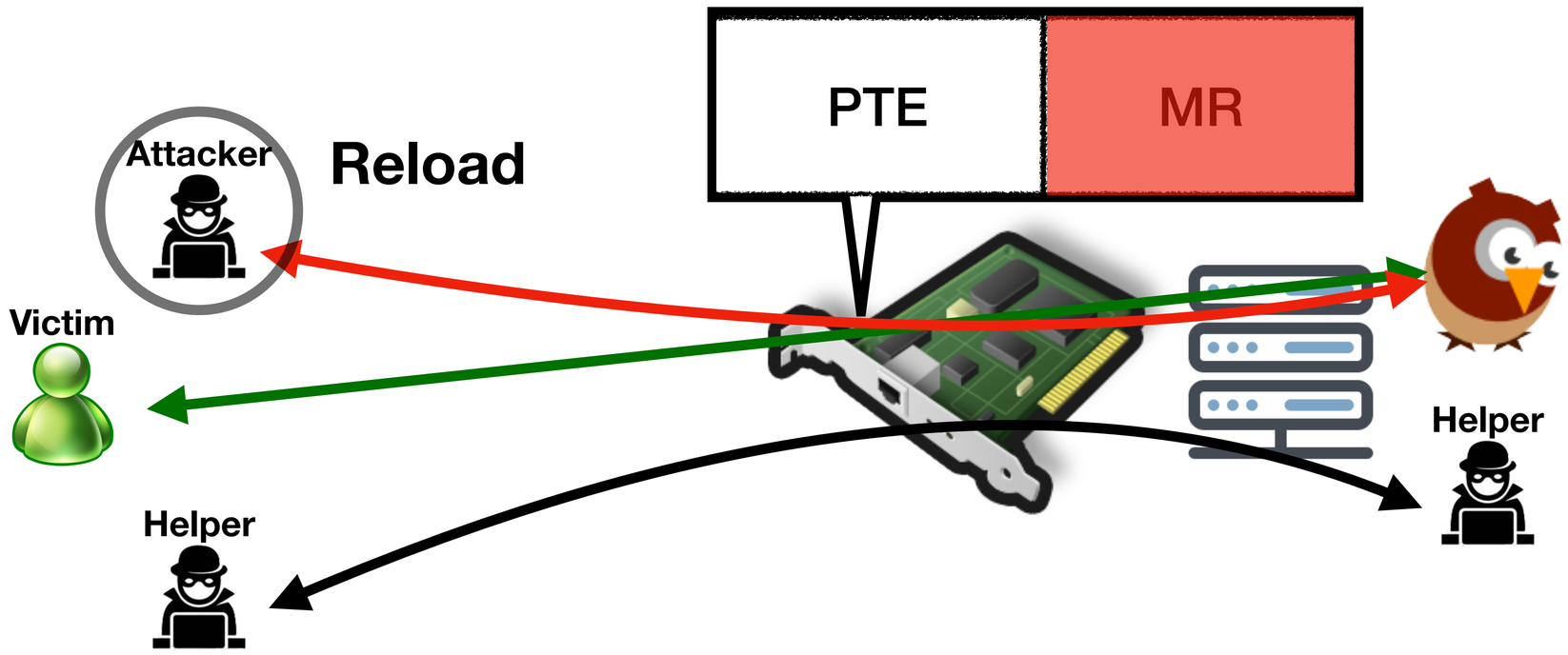
- MR-based



Need a Helper process

How to Attack Crail?

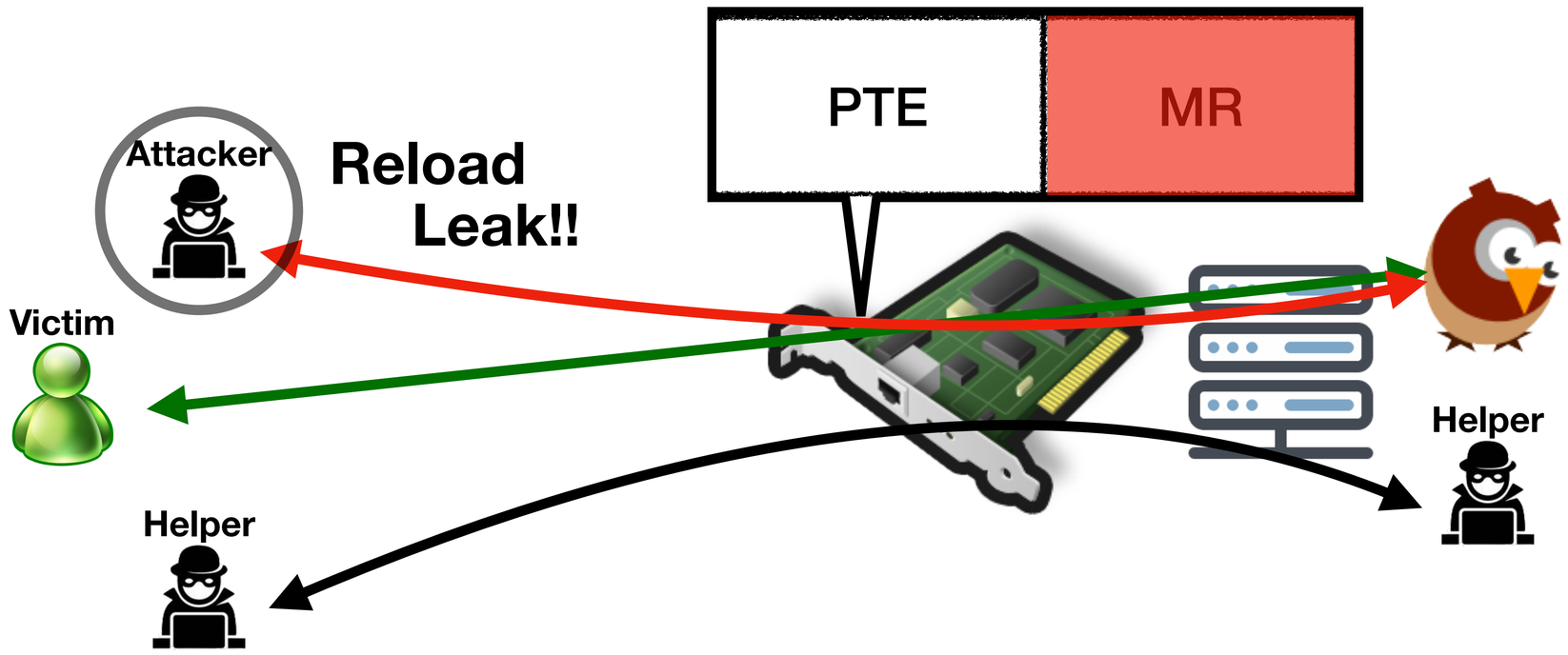
- MR-based



Need a Helper process

How to Attack Crail?

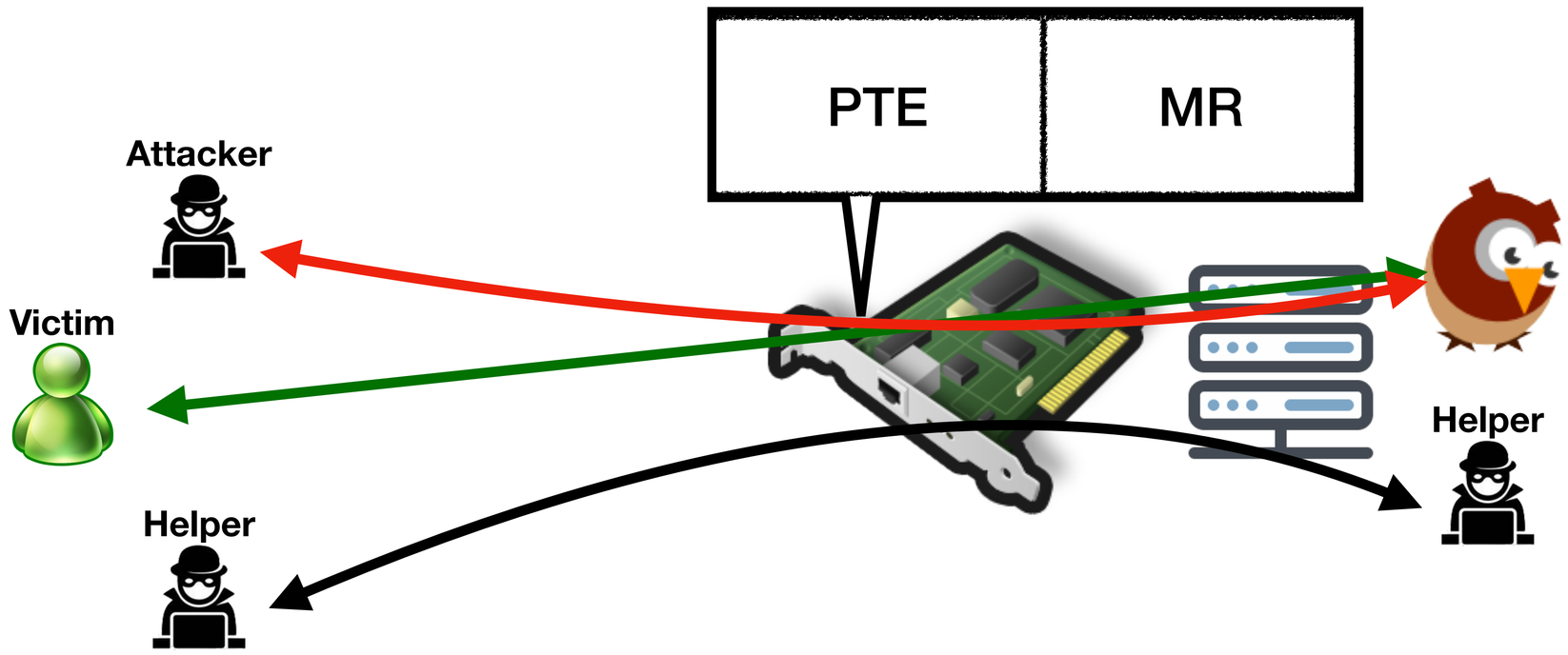
- MR-based



Need a Helper process

How to Attack Crail?

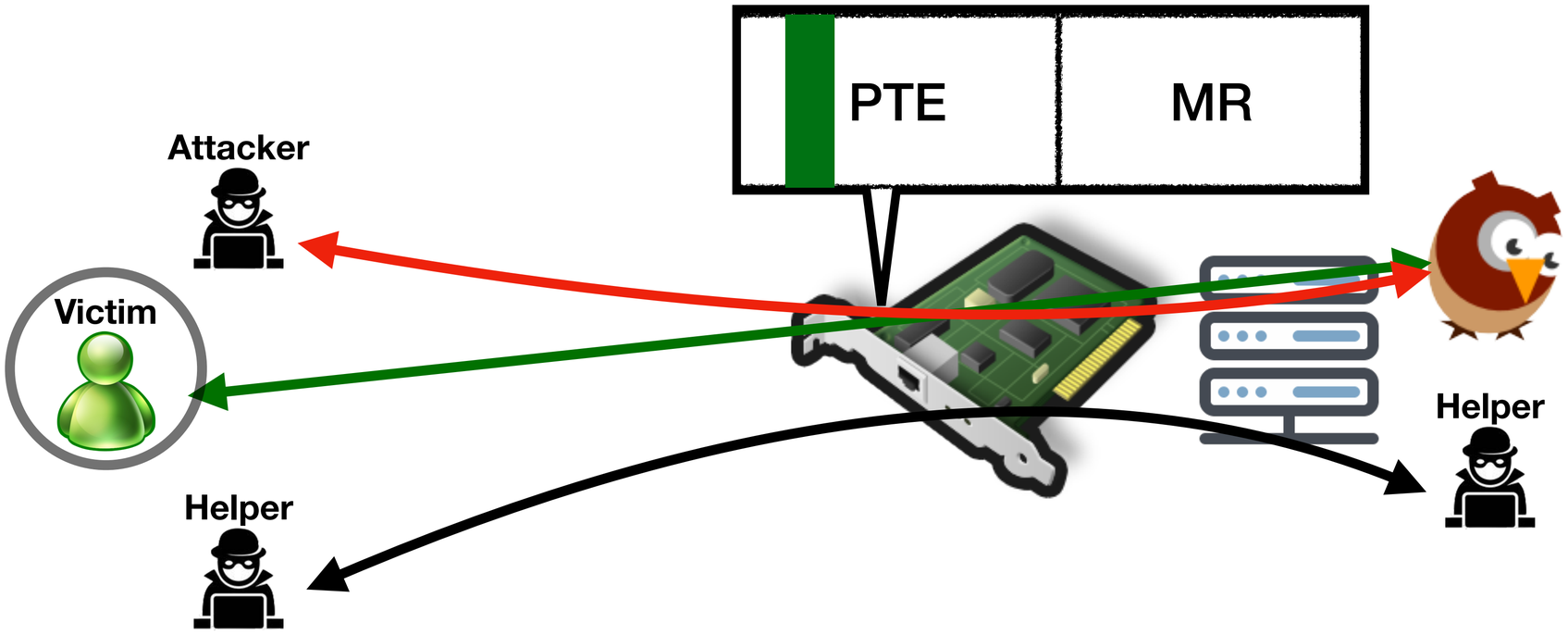
- PTE-based



Need a Helper process

How to Attack Crail?

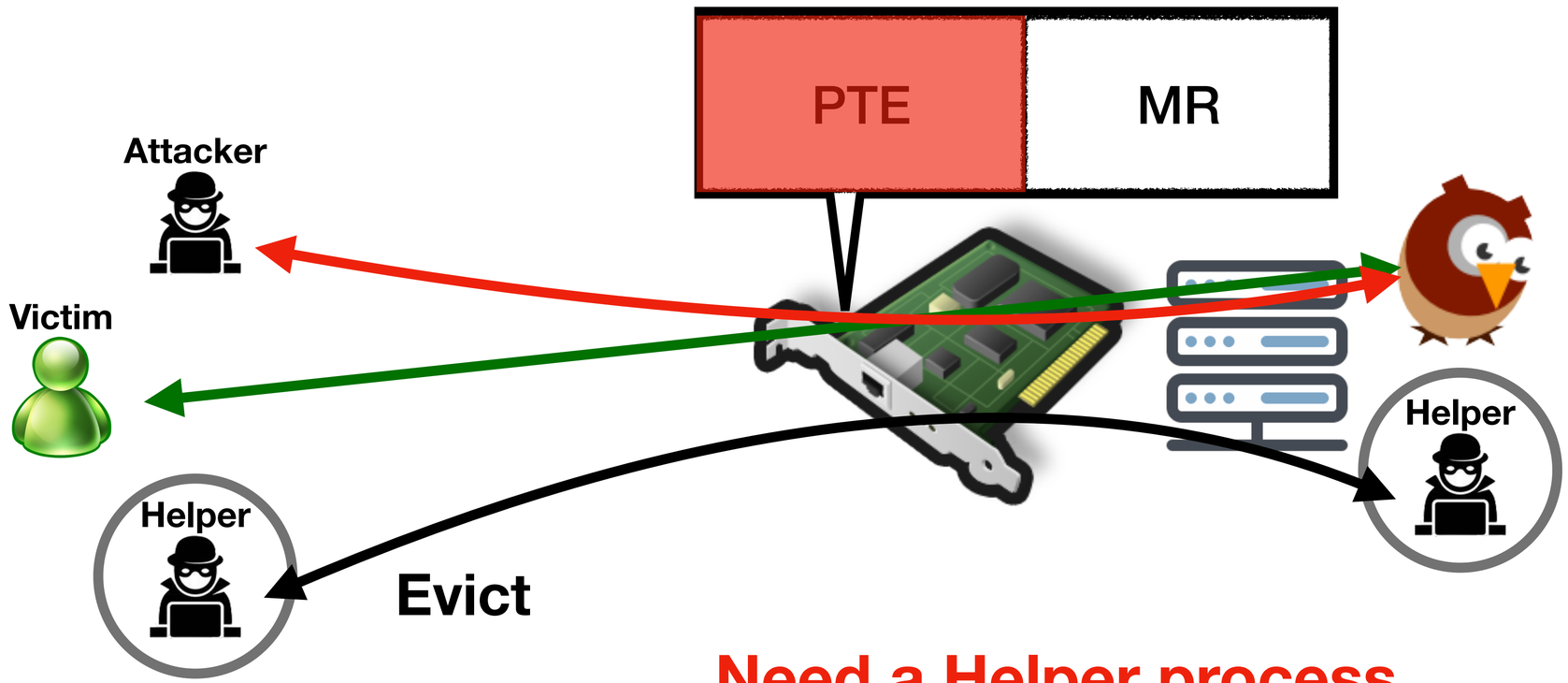
- PTE-based



Need a Helper process

How to Attack Crail?

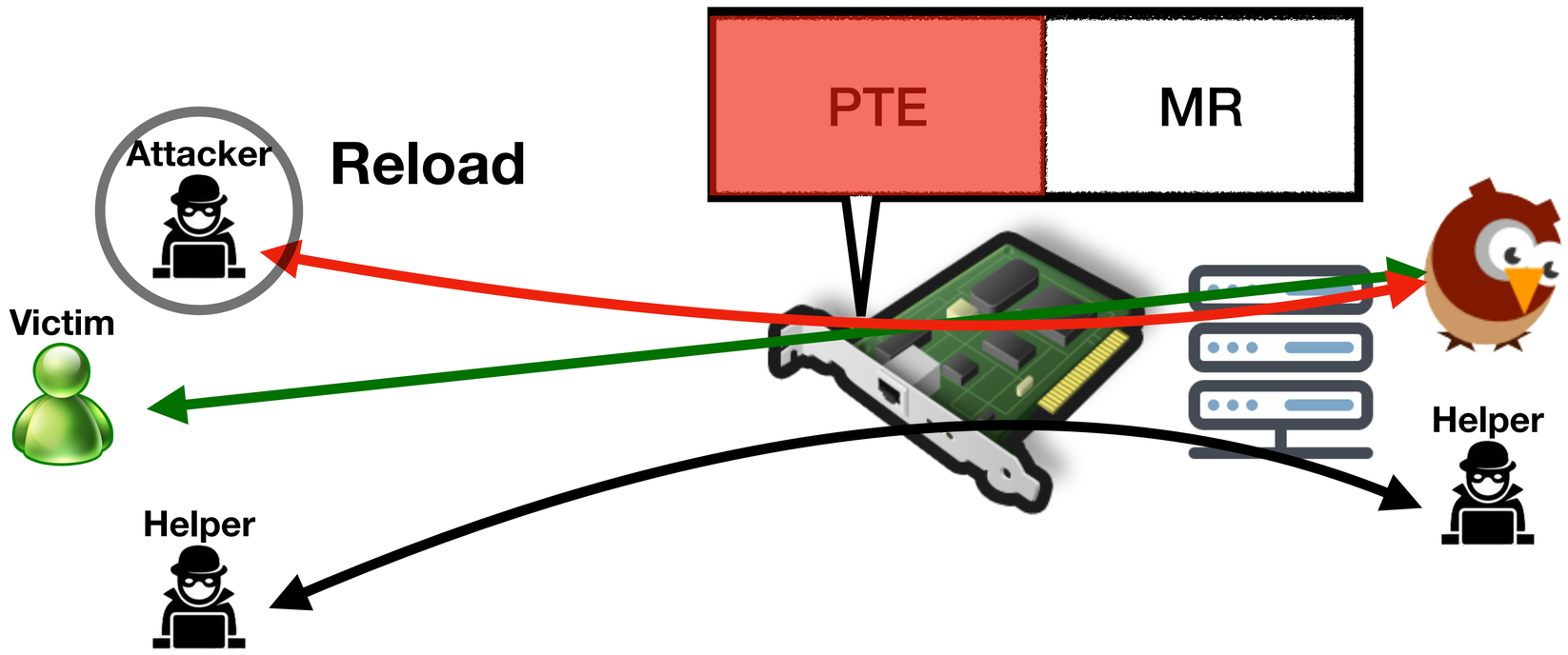
- PTE-based



Need a Helper process

How to Attack Crail?

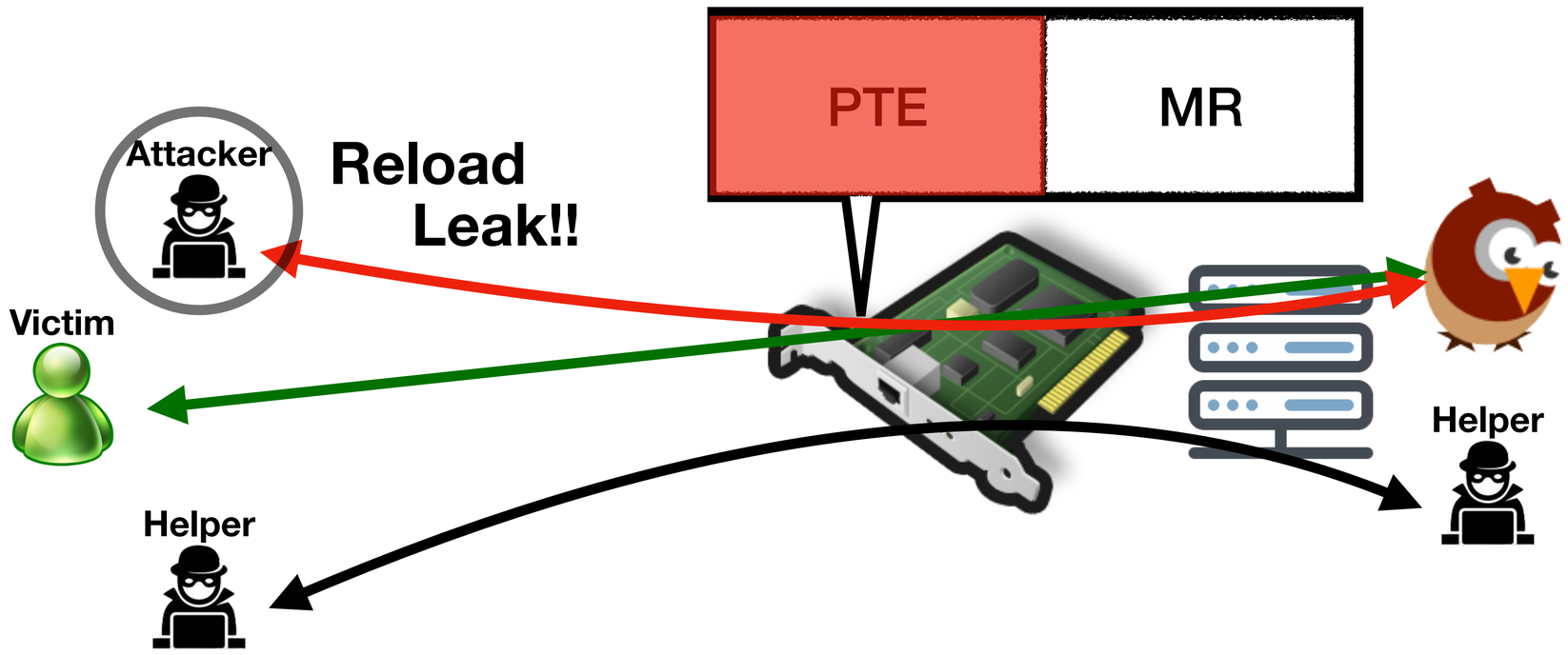
- PTE-based



Need a Helper process

How to Attack Crail?

- PTE-based

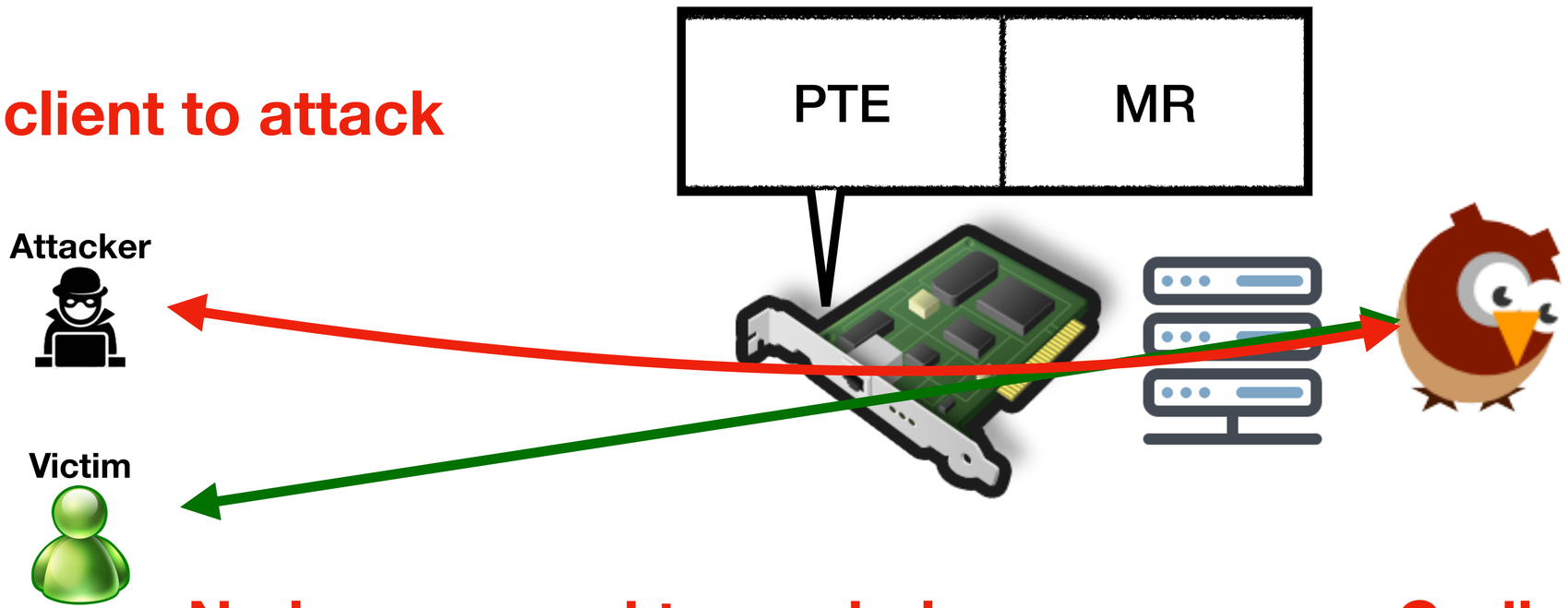


Need a Helper process

How to Attack Crail?

- Client-based

Use only Crail client to attack

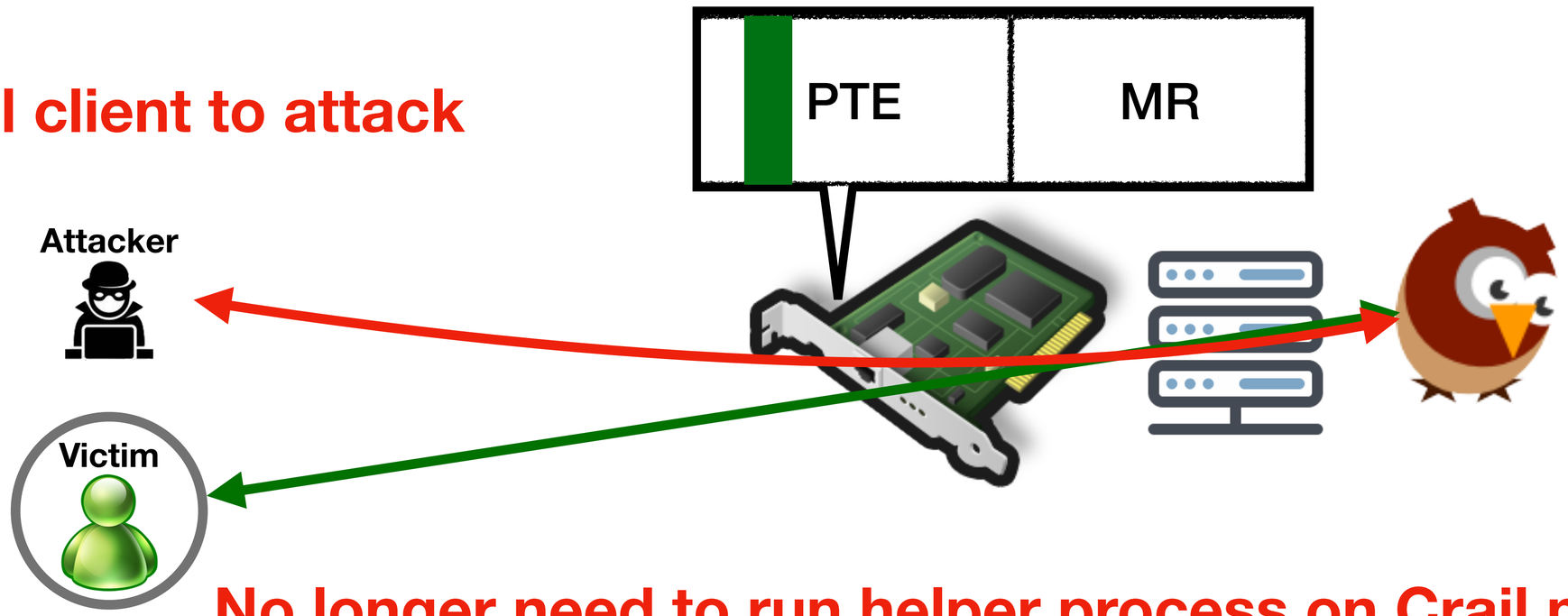


No longer need to run helper process on Crail machine

How to Attack Crail?

- Client-based

Use only Crail client to attack

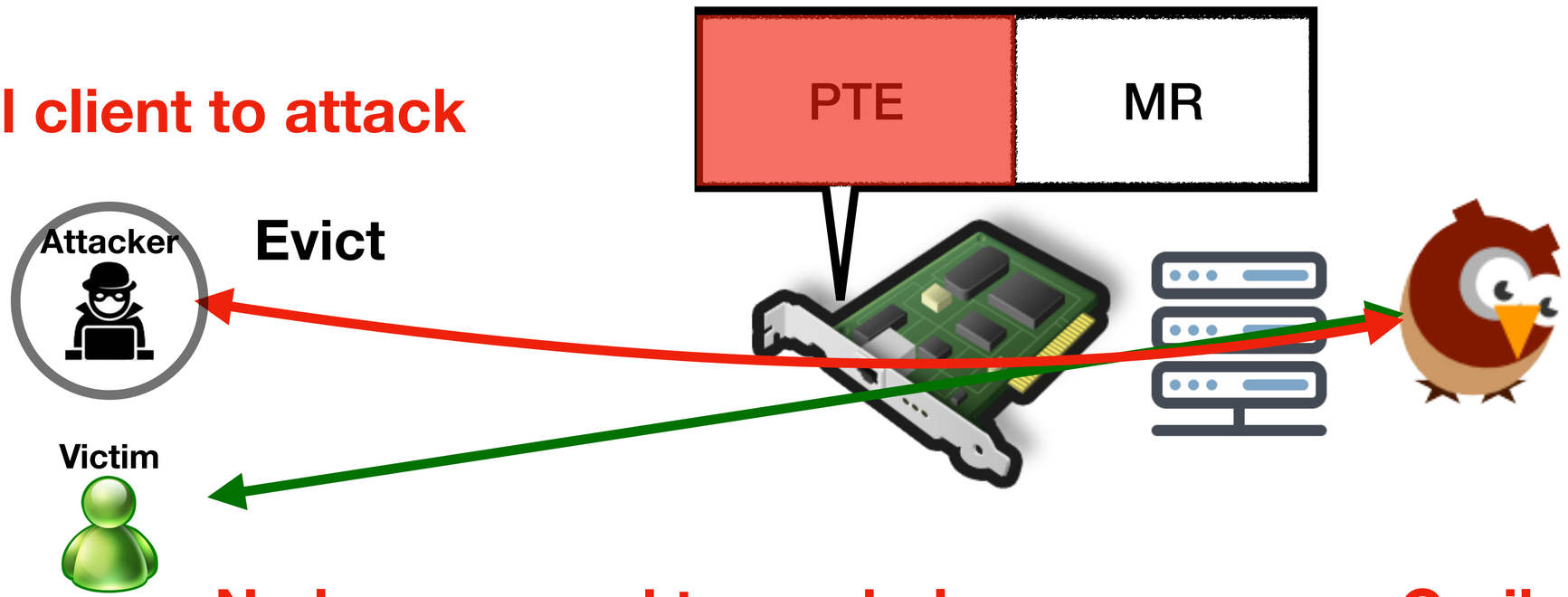


No longer need to run helper process on Crail machine

How to Attack Crail?

- Client-based

Use only Crail client to attack

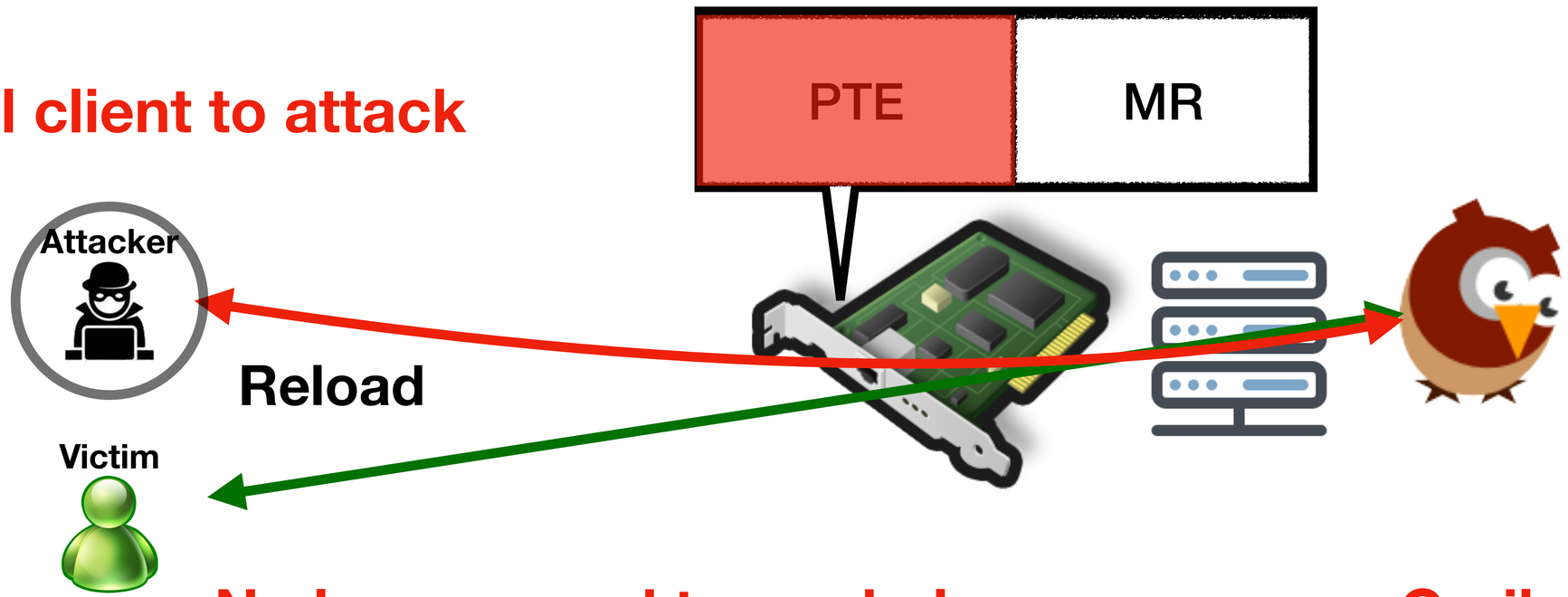


No longer need to run helper process on Crail machine

How to Attack Crail?

- Client-based

Use only Crail client to attack

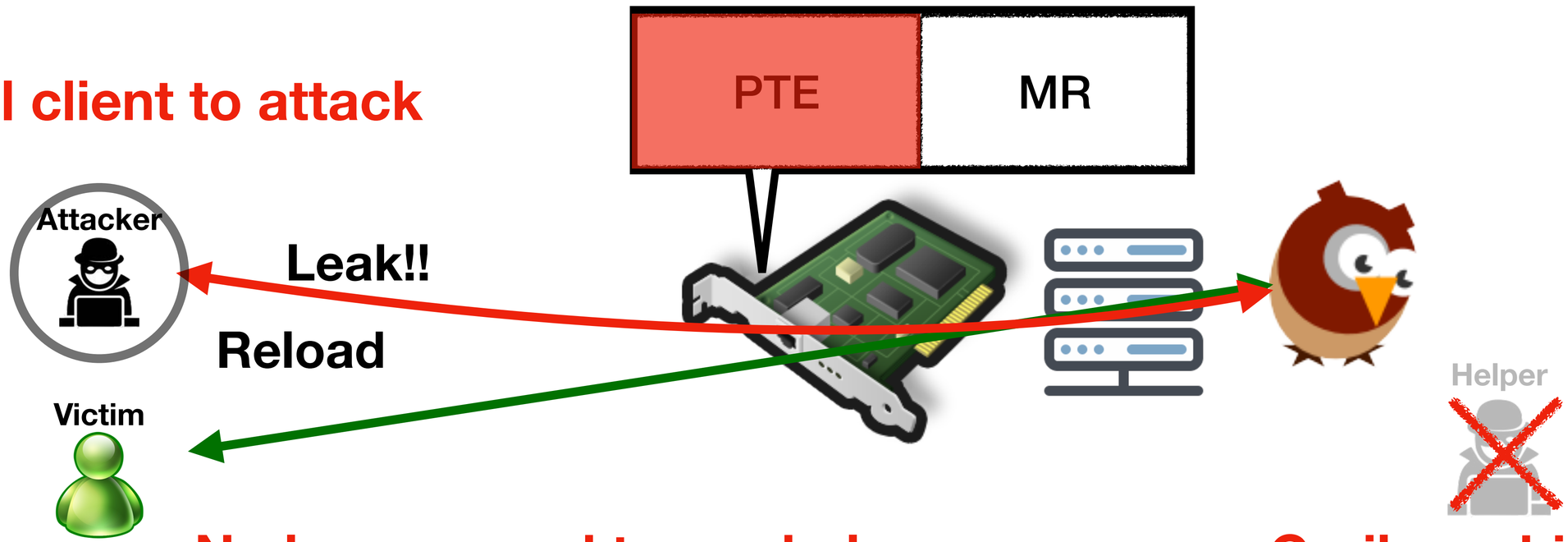


No longer need to run helper process on Crail machine

How to Attack Crail?

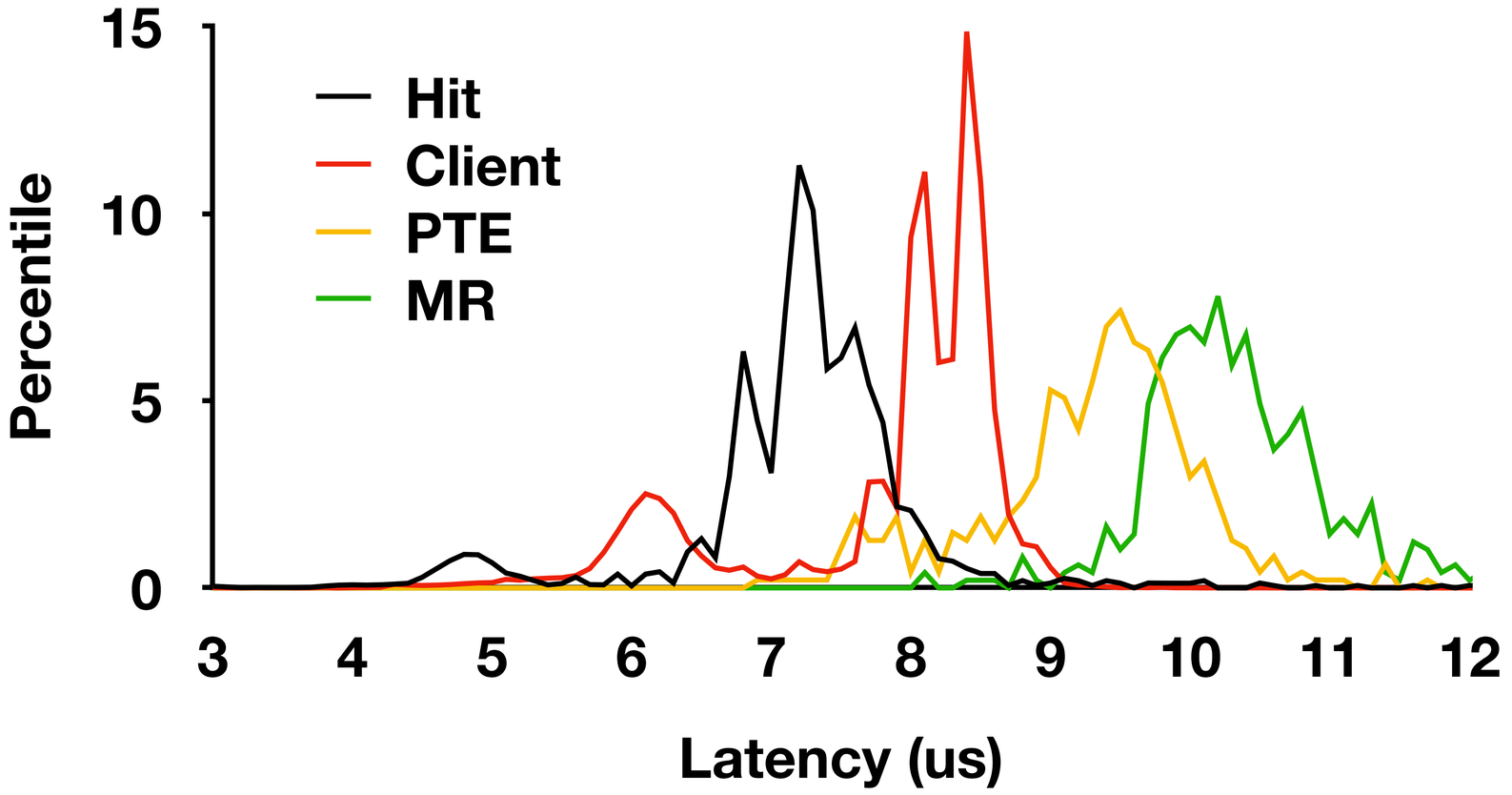
- Client-based

Use only Crail client to attack

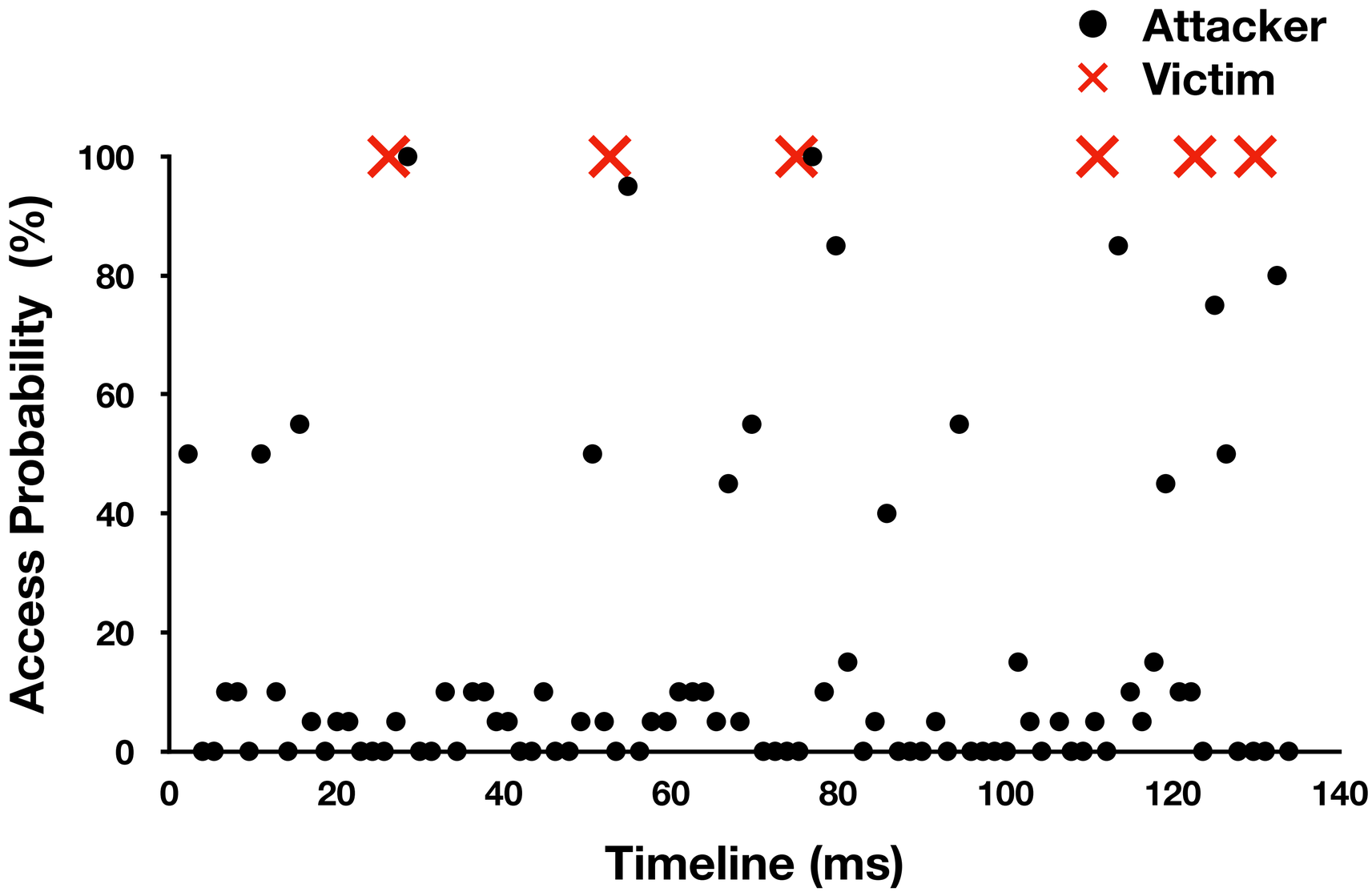


No longer need to run helper process on Crail machine

Timing Difference - Crail



Client-Based Attack



Mitigations

- **Client side**
  - Introduce noise in software

Mitigations

- **Client side**
  - Introduce noise in software
- **Network side**
  - Deploy network traffic monitoring

Mitigations

- **Client side**
  - Introduce noise in software
- **Network side**
  - Deploy network traffic monitoring
- **Server side**
  - Use huge page or no virtual memory
  - Hardware resource isolation on RNIC
  - Separate resource between different connections (protection domain)

Conclusion

- Security concerns of RDMA in datacenter
- How to add security guarantees to RDMA?
- Tradeoffs between **Performance** and **Security**

Get Pythia at: *https://github.com/Wuklab/Pythia*

wuklab.io

PURDUE UNIVERSITY

EPFL

UC San Diego