Digital Security

Introduction to software-based microarchitectural side-channel attacks

KazHackStan CYBERSEC CONFERENCE

Abc Xyz
@dura_lex

Agenda

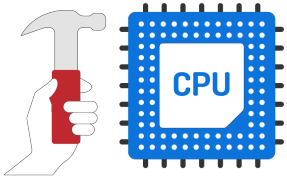
Introduction

Side-channel attacks

Typical target of a side-channel attack

Microarchitectural attacks

```
code1a:
  mov (X), %eax
  mov (Y), %ebx
  clflush (X)
  clflush (Y)
  jmp code1a
```



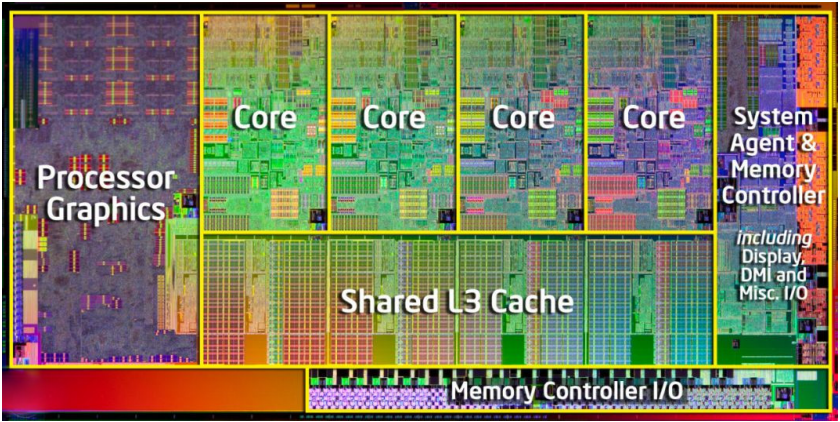
The DRAM cells get permanently damaged if hammered for a long time

Theory

Theory

CPU

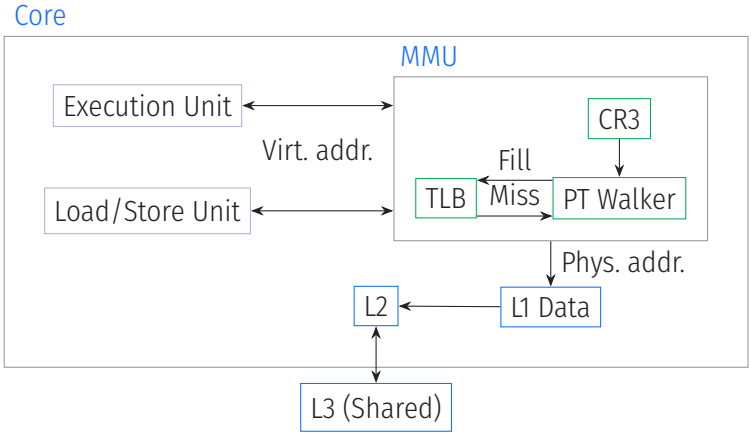
CPU



Architecture of multicore CPU

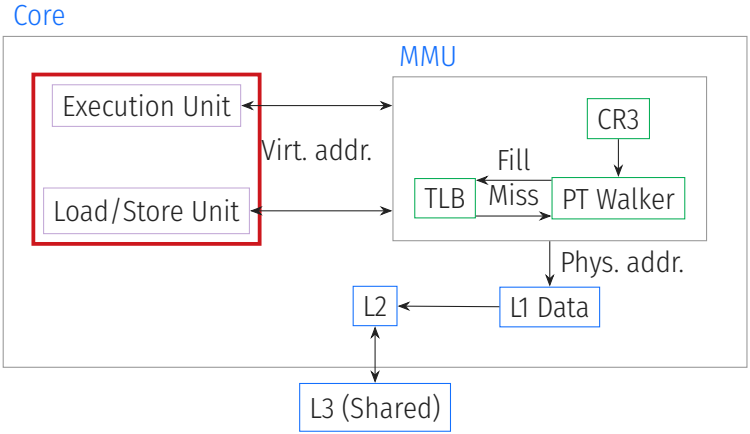
CPU

Abstract architecture of core and memory organization



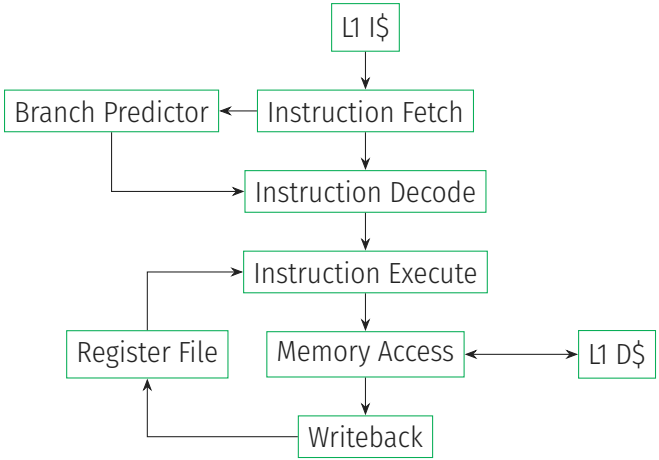
CPU

Abstract architecture of core and memory organization



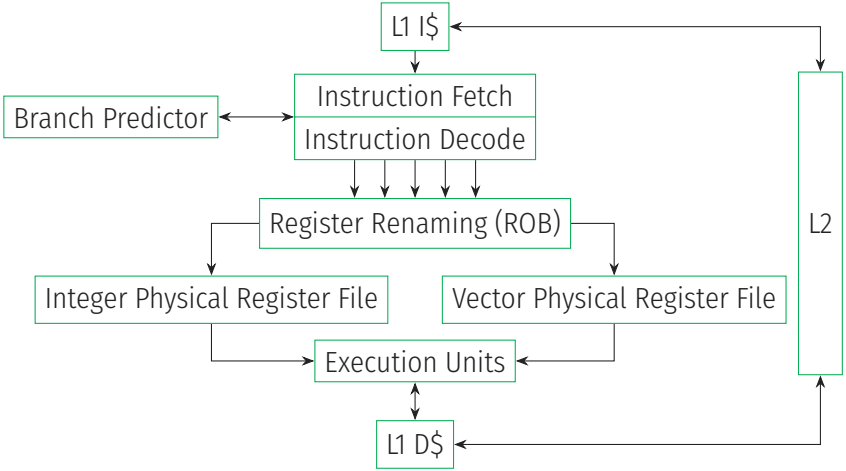
Pipelining. In-Order

Elements of a modern in-order core



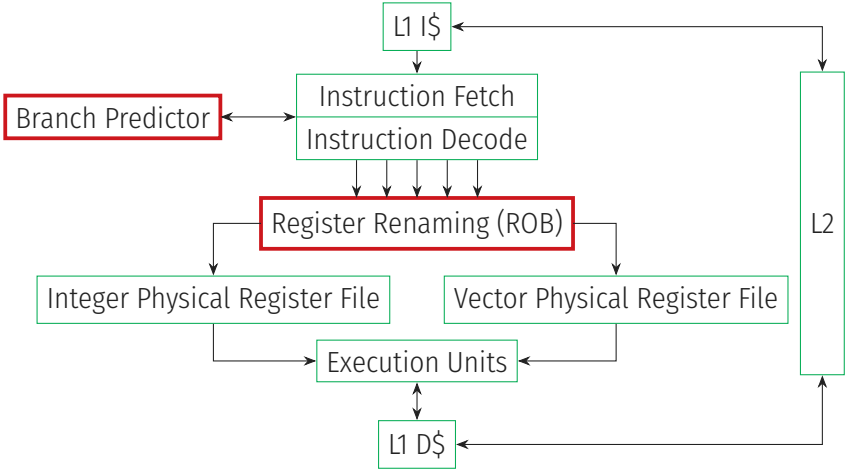
Pipelining. Out-of-Order

Elements of a modern out-of-order core

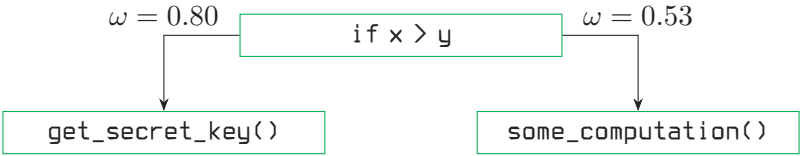


Pipelining. Out-of-Order

Elements of a modern out-of-order core

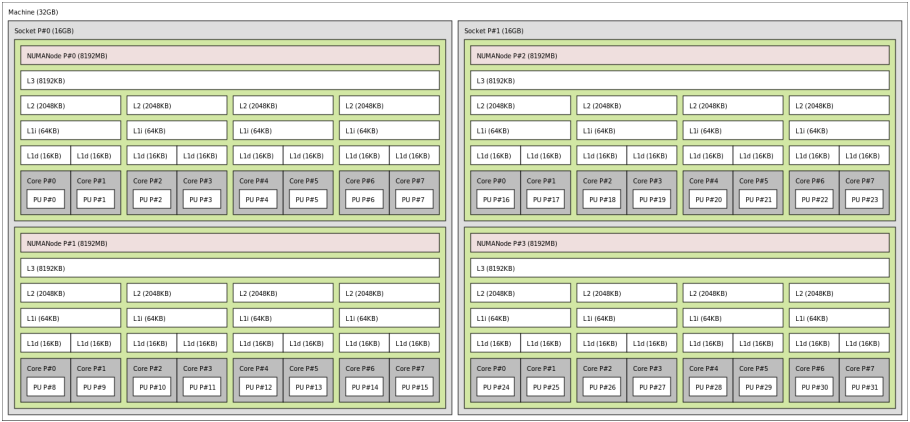


Branch Prediction and Speculation



`get_secret_key()` can be executed speculatively

Multicore



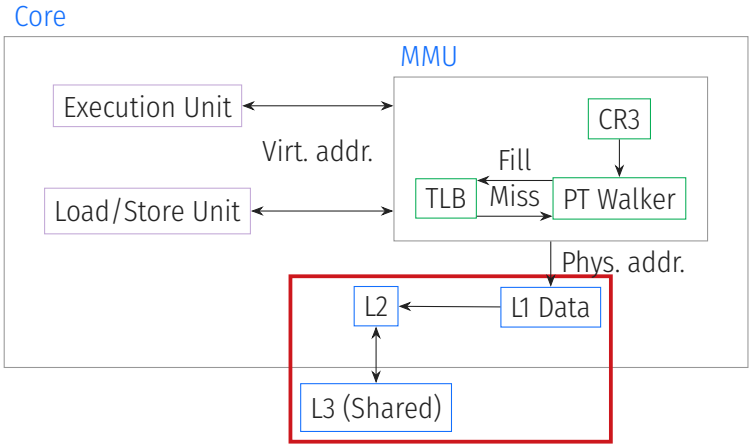
Architecture of multicore CPU AMD Bulldozer

Theory

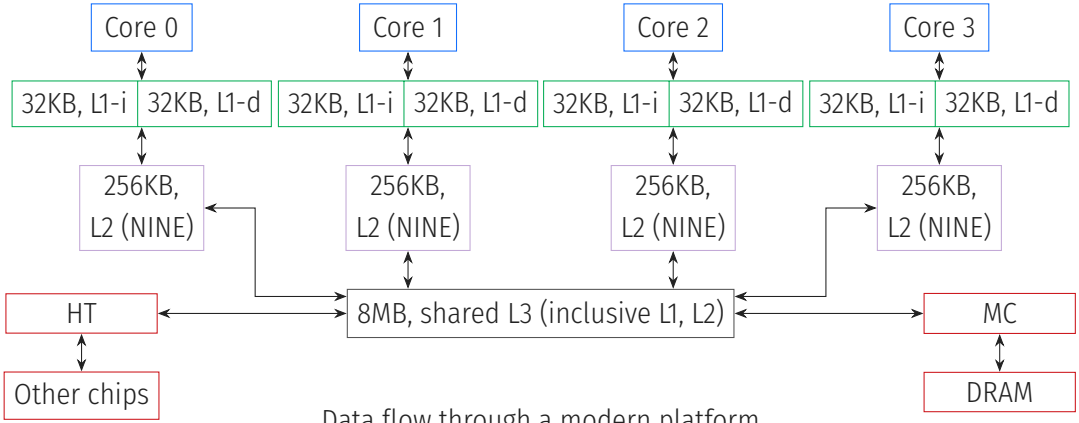
Cache

Cache

Abstract architecture of core and memory organization

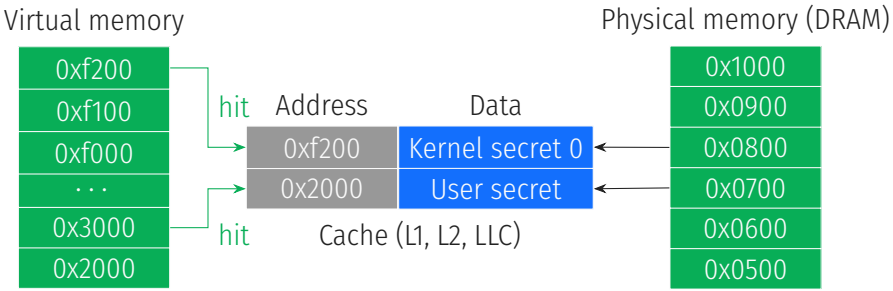


Cache



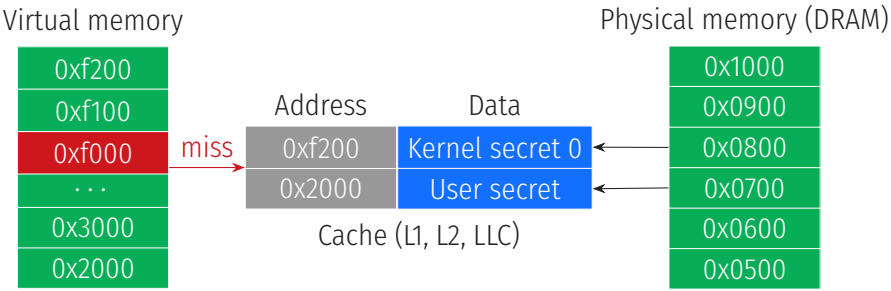
Data flow through a modern platform

Cache



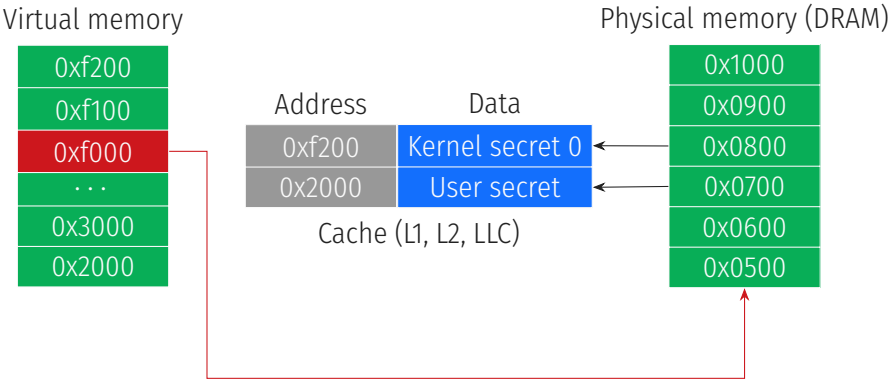
CPU cache algorithm

Cache



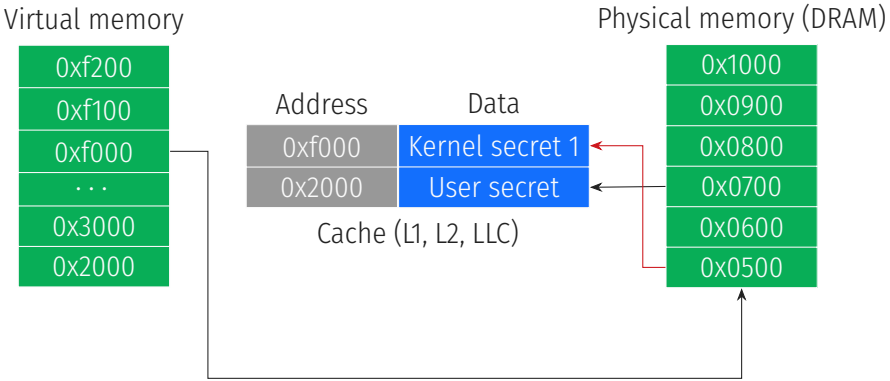
CPU cache algorithm

Cache



CPU cache algorithm

Cache

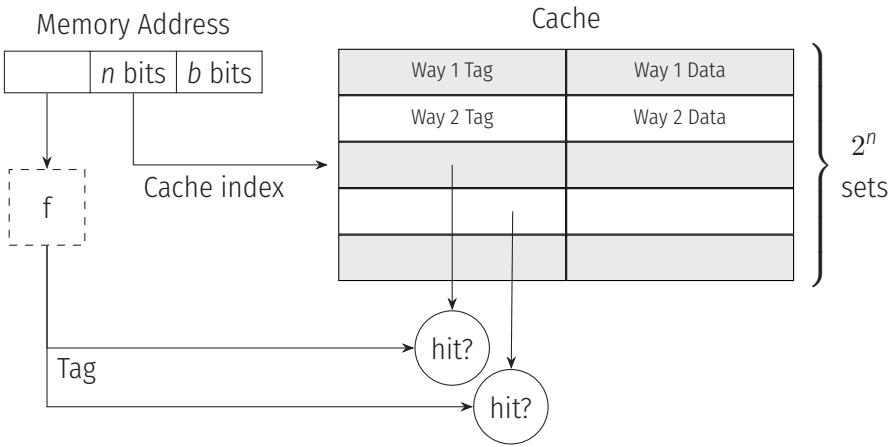


CPU cache algorithm

Types of cache

- Direct-mapped cache
- Fully-associative cache
- 2/4/8/12-way set associative cache

Two-way set associative cache



Cache replacement policies

- FIFO
- LIFO
- least recently used, LRU
- time aware least recently used, TLRU
- most recently used, MRU
- pseudo-LRU, PLRU
- random replacement, RR
- segment LRU, SLRU
- least frequently used, LFU
- least frequent recently used, LFRU
- LFU with dynamic aging, LFUDA
- low inter–reference recency set, LIRS
- adaptive replacement cache, ARC
- clock with adaptive replacement, CAR
- multi queue, MQ
- and etc.

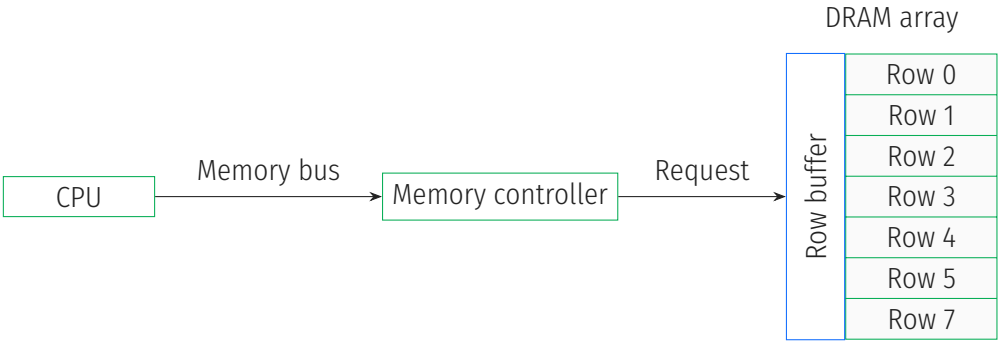
Addressing modes

- Virtually indexed, virtually tagged (VIVT)
- Physically indexed, virtually tagged (PIVT)
- Virtually indexed, physically tagged (VIPT)
- Physically indexed, physically tagged (PIPT)

Theory

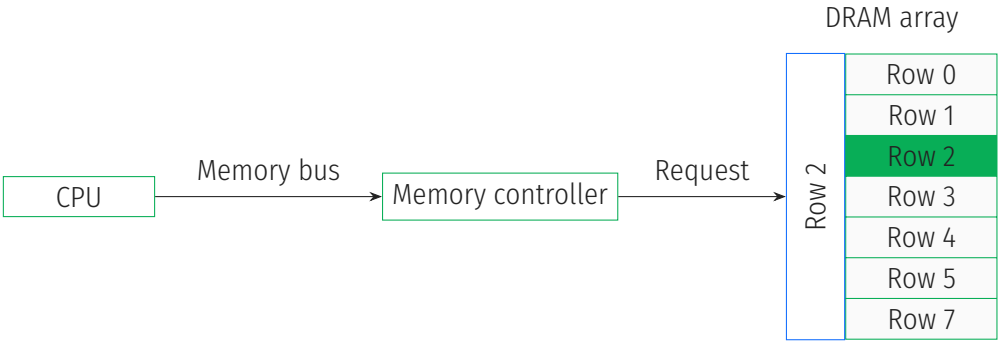
DRAM

How DRAM works



A simple computer system with a single DRAM array

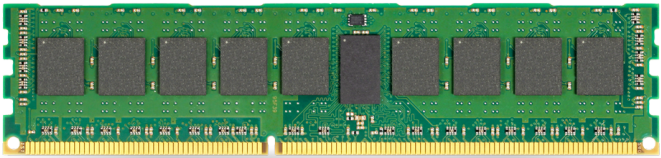
How DRAM works



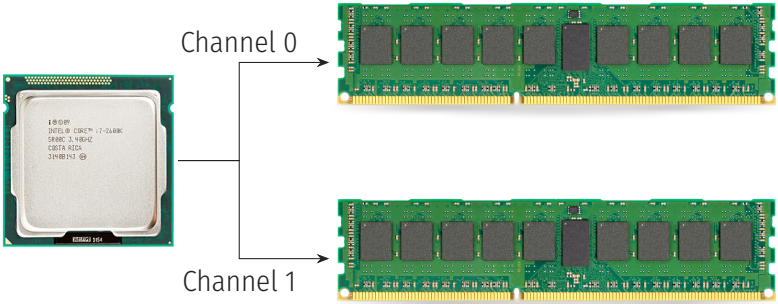
A simple computer system with a single DRAM array

DRAM organization

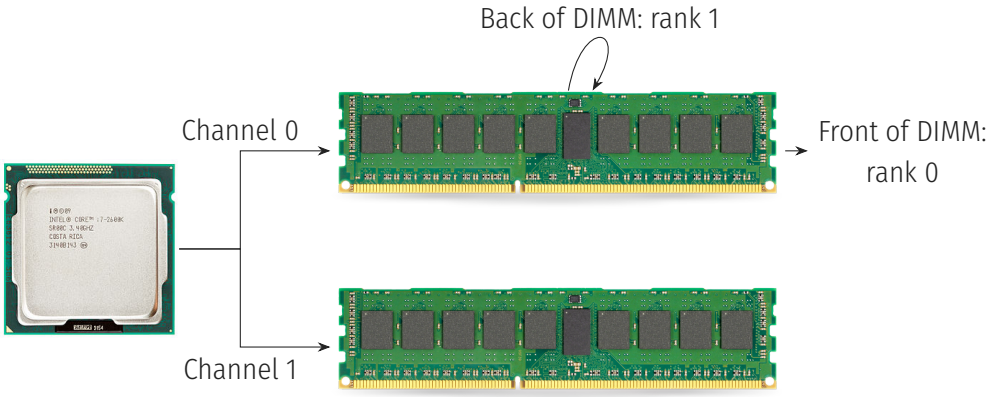
DIMM



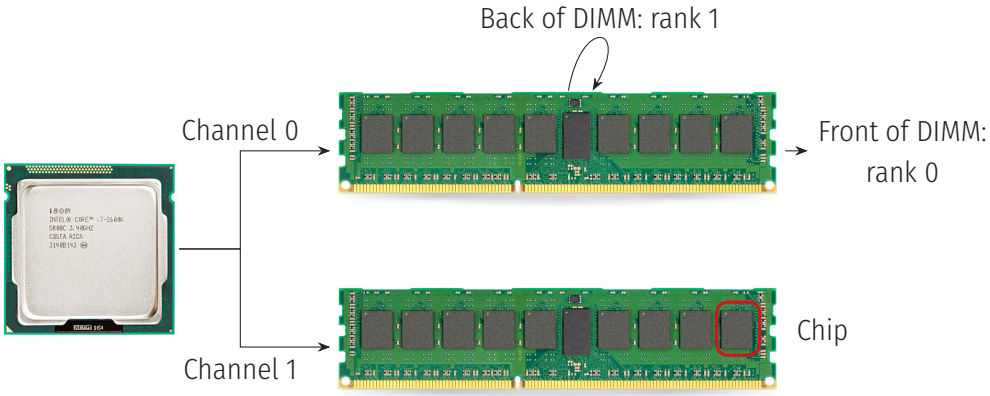
DRAM organization



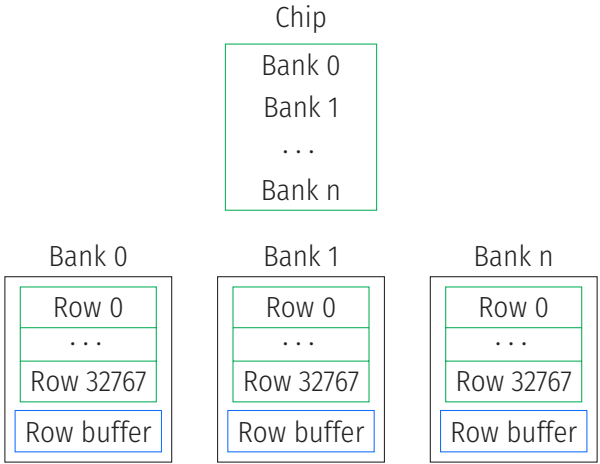
DRAM organization



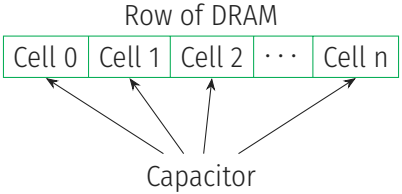
DRAM organization



DRAM organization



DRAM organization

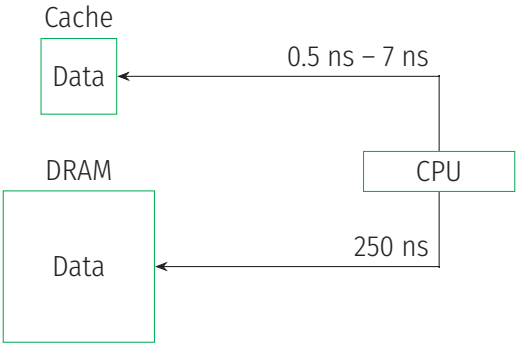


Basic attacks

Basic attacks

Cache attacks

Cache attacks



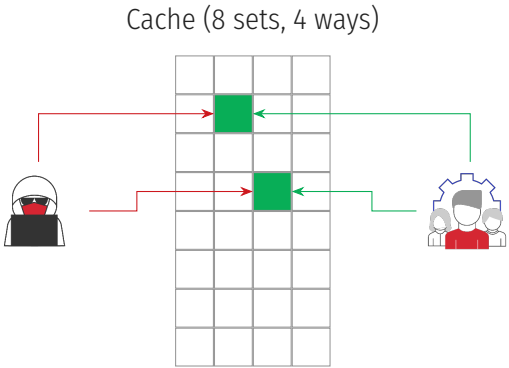
Timing attack — an attack aimed at exploiting differences in an algorithm execution time

Flush + Reload

1. Map binary (e.g., shared object) into the address space
2. Flush a cache line (code or data) from the cache
3. Schedule a victim's program
4. Check if the corresponding cache line from Step 2 has been loaded by the victim's program

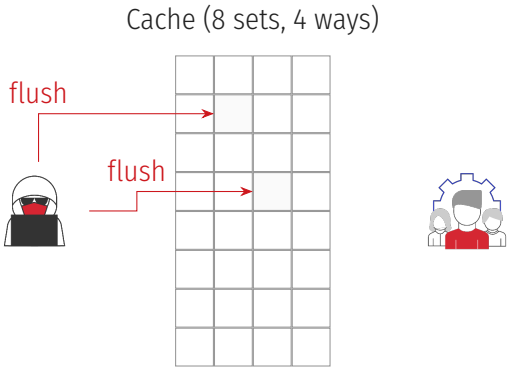
Flush + Reload

Map binary (e.g., shared object) into the address space



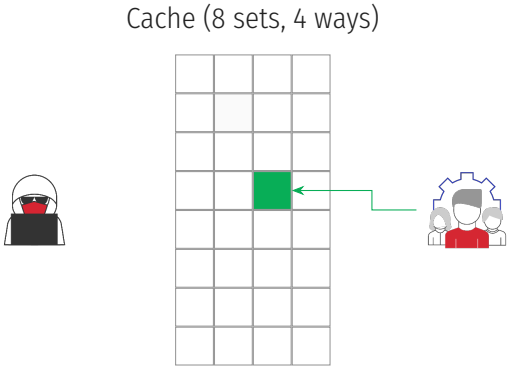
Flush + Reload

Flush a cache line (code or data) from the cache



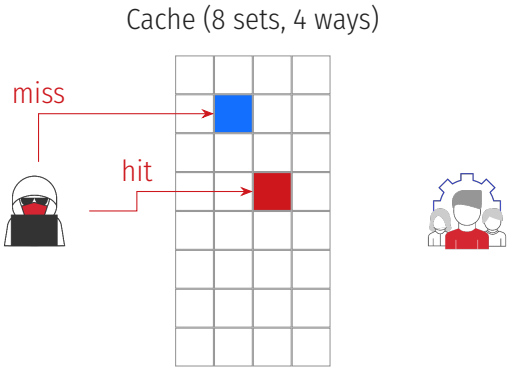
Flush + Reload

Schedule a victim's program



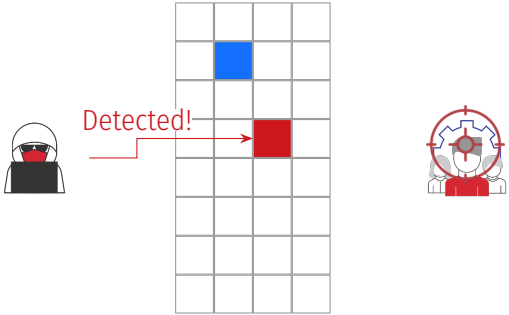
Flush + Reload

Check if the corresponding cache line from Step 2 has been loaded by the victim's program



Flush + Reload

Cache (8 sets, 4 ways)



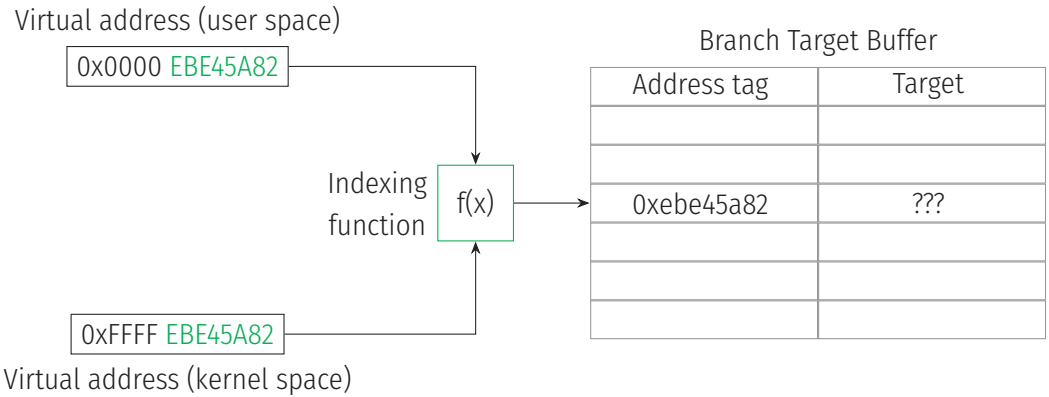
Cache attacks

- Evict + Time
- Prime + Probe
- Prime + Abort
- Flush + Flush
- Evict + Reload
- AnC (ASLR $\oplus$ Cache)
- etc.

Basic attacks

Branch-prediction attacks

Branch-prediction attacks

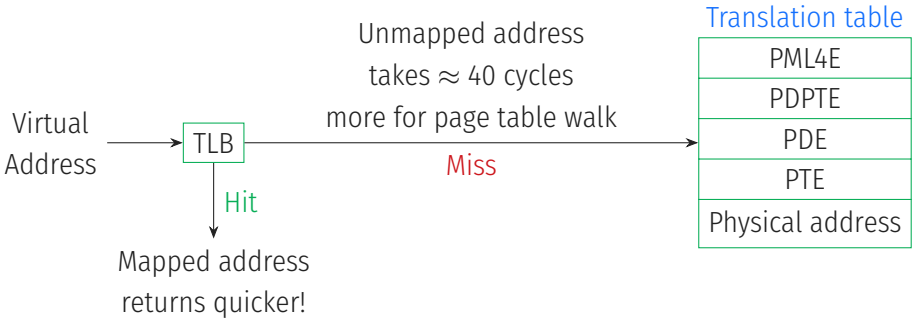


Branch Target Buffer addressing a scheme in the Haswell processor

Basic attacks

TLB-based attacks

TLB-based attacks



A translation lookaside buffer (TLB) is a memory cache that is used to reduce the time taken to access a user memory location

Basic attacks

Exception-based attacks

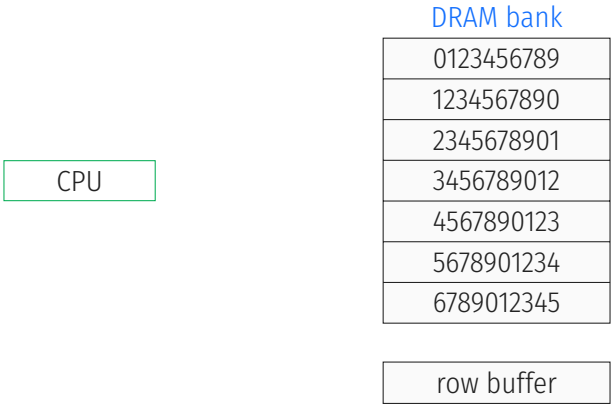
Exception-based attacks

- Scheduler interrupts
- Instruction aborts
- Page faults
- Behavioral differences (e.g, error code)

Basic attacks

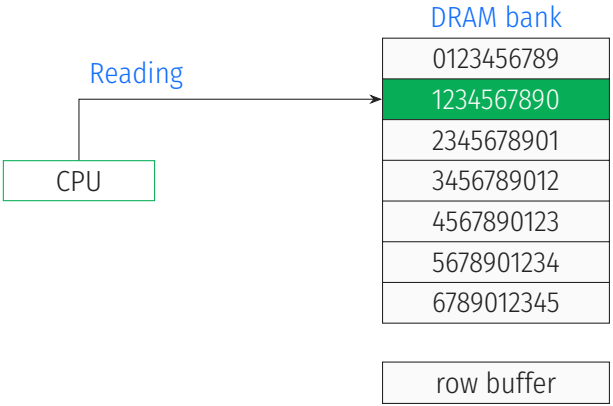
DRAM-based attacks

Reading from DRAM



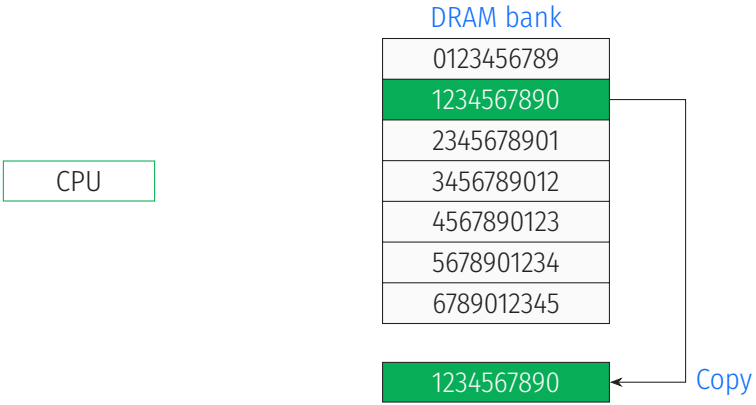
Reading from DRAM

Reading from DRAM

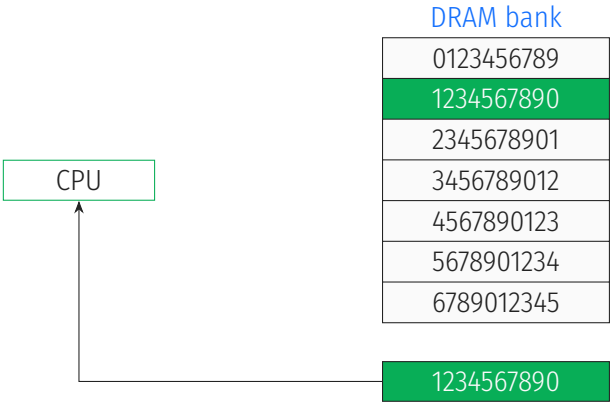


CPU reads row 1, row buffer is empty

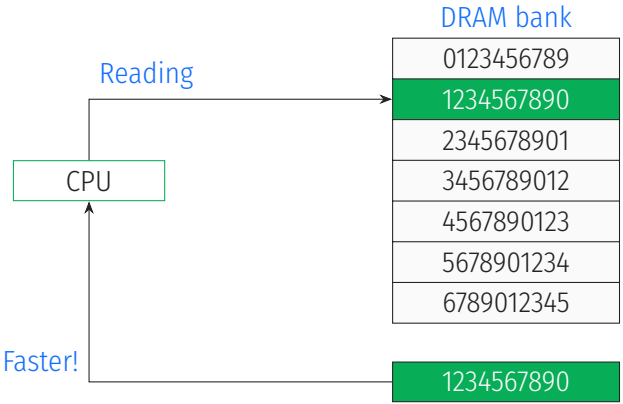
Reading from DRAM



Reading from DRAM



Reading from DRAM



CPU reads row 1, row buffer is now full

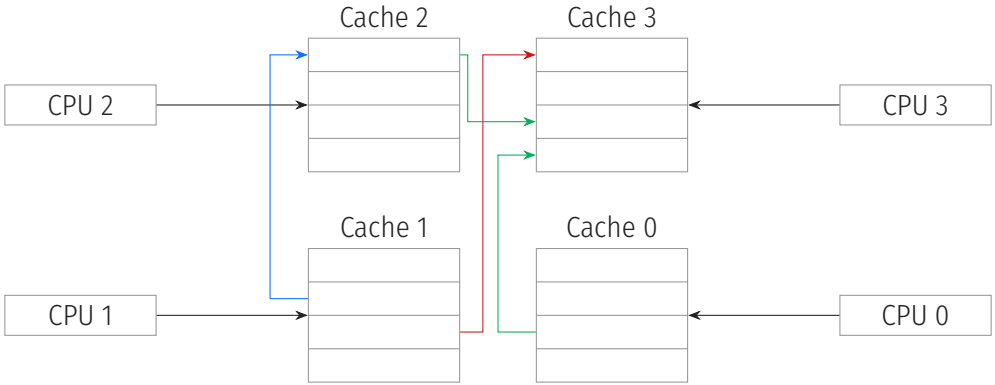
Complex DRAM-based attacks

- DRAMA
- Row hit (Flush + Reload)
- Row miss (Prime + Probe)
- etc.

Basic attacks

Covert channels

Covert channels



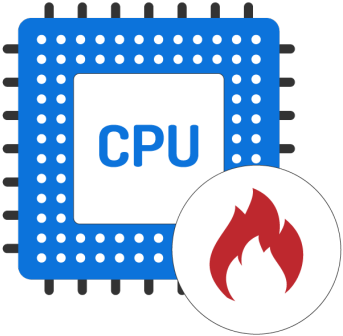
Cross-core covert channels

Covert channels

- Cache-based covert channels (shared libraries)
- Row miss attack (DRAM)
- Thermal covert channels
- Radio covert channels

Software-based Microarchitectural Fault Attacks

Software-based Microarchitectural Fault Attacks



Software-based microarchitectural fault attacks do not require physical access, but only some form of code execution on the target system

Software-based Microarchitectural Fault Attacks

Rowhammer

Rowhammer. Exploitation primitives

- Fast uncached memory access
- Physical memory massaging
- Physical memory addressing

Variations of Rowhammer

- Flip Feng Shui — targeted Rowhammer
- Throwhammer — remote Rowhammer
- Nethammer — better remote Rowhammer
- Drammer, RAMpage — exploitation ARM-based hardware
- Glitch — better exploitation ARM-based hardware

Meltdown & Spectre

Meltdown & Spectre

Derived attacks and not only

Derived attacks and not only

- MeltdownPrime & SpectrePrime
- SgxPectre
- SMM Speculative Execution Attacks
- BranchScope
- LazyFP
- ...

Derived attacks and not only

- Spectre 1.1, 1.2 (Speculative Buffer Overflows)
- SpectreRSB
- NetSpectre
- L1TF (Foreshadow)
- etc.

TotalMeltdown? And what's about other patches...

Meltdown & Spectre

Abstract example of exploitation

Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive

Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive

- Bypass out of bounds checks
- Training of branch predictor
- Speculatively read an earlier value of the data
- Pending exceptions
- Exploit branch history table
- Exploit the Return Stack Buffer
- Speculatively write to register (buffer overflow)

Abstract example of exploitation

Type of BP

Algorithm of BP

Environment of BP

The foundation of the Speculative-Based Attack tower

Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive
2. Windowing gadget

Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive
2. Windowing gadget

- Non-cached loads
- Dependency chain of loads
- Dependency chain of integer ALU operations

Abstract example of exploitation

Search/create gadgets

Contents of cache

Type of BP

Algorithm of BP

Environment of BP

The Speculative-Based Attack tower

Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive
2. Windowing gadget
3. Disclosure gadget

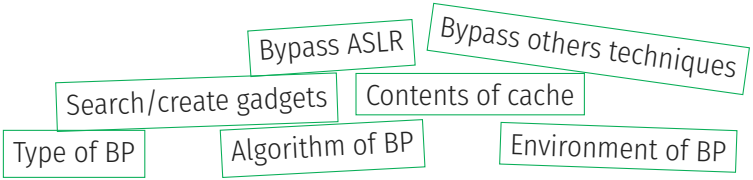
Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive
2. Windowing gadget
3. Disclosure gadget

- ASLR
- CFI
- SMAP
- DEP/NX
- retpoline
- and others.

Abstract example of exploitation



The Speculative-Based Attack **Babel** tower

Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive
2. Windowing gadget
3. Disclosure gadget
4. Disclosure primitive

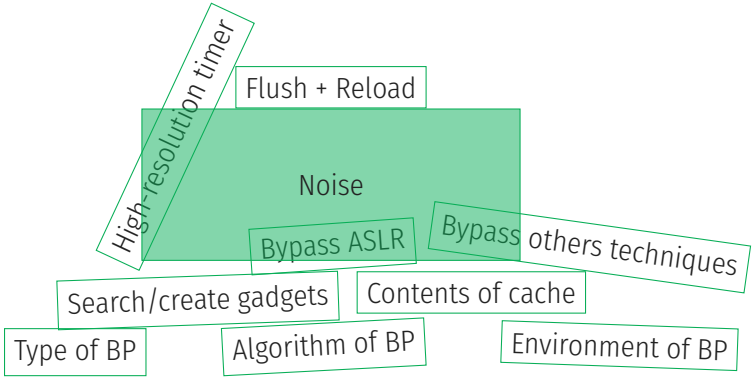
Abstract example of exploitation

Four components of speculative techniques

1. Speculation primitive
2. Windowing gadget
3. Disclosure gadget
4. Disclosure primitive

- Architecture of cache
- Replacement policies
- Exclusive and inclusive
- Type of cache attack
- Noise
- High-resolution timer
- and etc.

Abstract example of exploitation



The Speculative-Based Attack **Babel** tower

Summary

Summary

- Software-based microarchitectural attacks has become very popular

Summary

- Software-based microarchitectural attacks has become very popular
- Requires a lot of resources to develop a working exploit

Summary

- Software-based microarchitectural attacks has become very popular
- Requires a lot of resources to develop a working exploit
- Microarchitectural attacks may be automated

Summary

- Software-based microarchitectural attacks has become very popular
- Requires a lot of resources to develop a working exploit
- Microarchitectural attacks may be automated
- Many attacks have not yet been disclosed

Summary

- Software-based microarchitectural attacks has become very popular
- Requires a lot of resources to develop a working exploit
- Microarchitectural attacks may be automated
- Many attacks have not yet been disclosed
- Countermeasures come with a performance impact

Questions?

References i

- D. Gruss, “Software-based Microarchitectural Attacks.”
- M. Lipp and D. Gruss, “ARMageddon: Cache Attacks on Mobile Devices.”
- D. Page, “MASCAB: a Micro-Architectural Side-Channel Attack Bibliography.”
- P. Pessl and D. Gruss, “DRAMA: Exploiting DRAM Addressing for Cross-CPU Attacks.”
- H. Bos and Y. Fratantonio, “Drammer: Determenistic Rowhammer Attacks on Mobile Platforms.”
- Microsoft, “Mitigating speculative execution side channel hardware vulnerabilities.”
- J. Horn, “Reading privileged memory with a side-channel.”

References ii

- D. Gruss and M. Lipp, “KASLR is Dead: Long Live KASLR.”
- D. Gruss and C. Maurice, “Flush+Flush: A Fast and Stealthy Cache Attack.”
- F. Liu and Y. Yarom, “Last-Level Cache Side-Channel Attacks are Practical.”
- C. Trippel, D. Lustig, and M. Martonosi, “MeltdownPrime and SpectrePrime: Automatically-Synthesized Attacks Exploiting Invalidation-Based Coherence Protocols.”
- M. Schwarz, C. Maurice, D. Gruss, and S. Mangard, “Fantastic Timers and Where to Find Them: High-Resolution Microarchitectural Attacks in JavaScript.”
- M. Lipp and M. T. Aga, “Nethammer: Inducing Rowhammer Faults through Network Requests.”

References iii

- A. Tatar and R. Krishnan, “Throwhammer: Rowhammer Attacks over the Network and Defenses.”
- G. Camurati and S. Poeplau, “Screaming Channels: When Electromagnetic Side Channels Meet Radio Transceivers.”
- J. Stecklina and T. Prescher, “LazyFP: Leaking FPU Register State using Microarchitectural Side-Channels.”
- M. Guri and A. Kachlon, “GSMem: Data Exfiltration from Air-Gapped Computers over GSM Frequencies.”
- D. Sullivan, O. Arias, T. Meade, and Y. Jin, “Microarchitectural Minefields: 4K-Aliasing Covert Channel and Multi-Tenant Detection in IaaS Clouds.”

References iv

- B. Gras, K. Razavi, E. Bosman, H. Bos, and C. Giuffrida, “ASLR on the Line: Practical Cache Attacks on the MMU.”
- S. van Schaik, C. Giuffrida, H. Bos, and K. Razavi, “Malicious Management Unit: Why Stopping Cache Attacks in Software is Harder Than You Think.”
- D. Gruss and A. Fogh, “Prefetch Side-Channel Attacks: Bypassing SMAP and Kernel ASLR.”
- E. M. Koruyeh and K. N. Khasawneh, “Spectre Returns! Speculation Attacks using the Return Stack Buffer.”
- G. Maisuradze and C. Rossow, “ret2spec: Speculative Execution Using Return Stack Buffers.”

References v

- G. Chen and S. Chen, “SgxPectre Attacks: Leaking Enclave Secrets via Speculative Execution.”
- M. Lipp and M. Schwarz, “Meltdown.”
- P. Kocher and D. Genkin, “Spectre Attacks: Exploiting Speculative Execution.”
- ARM, “Cache Speculation Side-channels.”
- M. Schwarz, M. Schwarzl, M. Lipp, and D. Gruss, “NetSpectre: Read Arbitrary Memory over Network.”
- S. D’Antoine, “Out-of-Order Execution and Its Applications.”
- V. Kiriansky and C. Waldspurger, “Speculative Buffer Overflows: Attacks and Defenses.”

References vi

- B. Gras, K. Razavi, H. Bos, and C. Giuffrida, “Translation Leak-aside Buffer: Defeating Cache Side-channel Protections with TLB Attacks.”
- C. Disselkoen, D. Kohlbrenner, L. Porter, and D. Tullsen, “Prime+Abort: A Timer-Free High-Precision L3 Cache Attack using Intel TSX.”
- M. Lipp and M. Schwarz, “Meltdown & Spectre Side-channels considered hARMful.”
- J. Masters, “Exploiting modern microarchitectures: Meltdown, Spectre, and other attacks.”
- M. Lipp, “Cache attacks on ARM.”
- D. Evtyushkin, D. Ponomarev, and N. Abu-Ghazaleh, “Jump over ASLR: attacking branch predictors to bypass ASLR.”