

FIC
2020

International Cybersecurity Forum

28th, 29th & 30th January 2020

LILLE GRAND PALAIS

Quelle confiance
peut-on placer dans
les plateformes
matérielles qui
exécutent nos
applications?

Guillaume Hiet
Clémentine Maurice

#FIC2020



You
Tube

in f

@FIC_eu . Forum FIC

WWW.FORUM-FIC.COM





Guillaume Hiet

Assistant professor, CentraleSupélec
@GuillaumeHiet

Clémentine Maurice

CNRS researcher
@BloodyTangerine



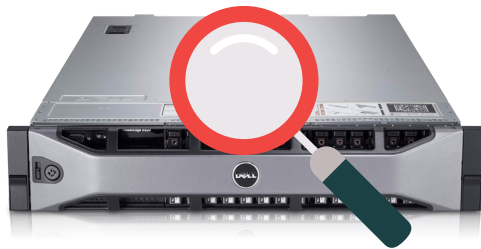
Introduction







server



server



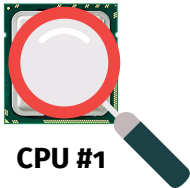
CPU #1

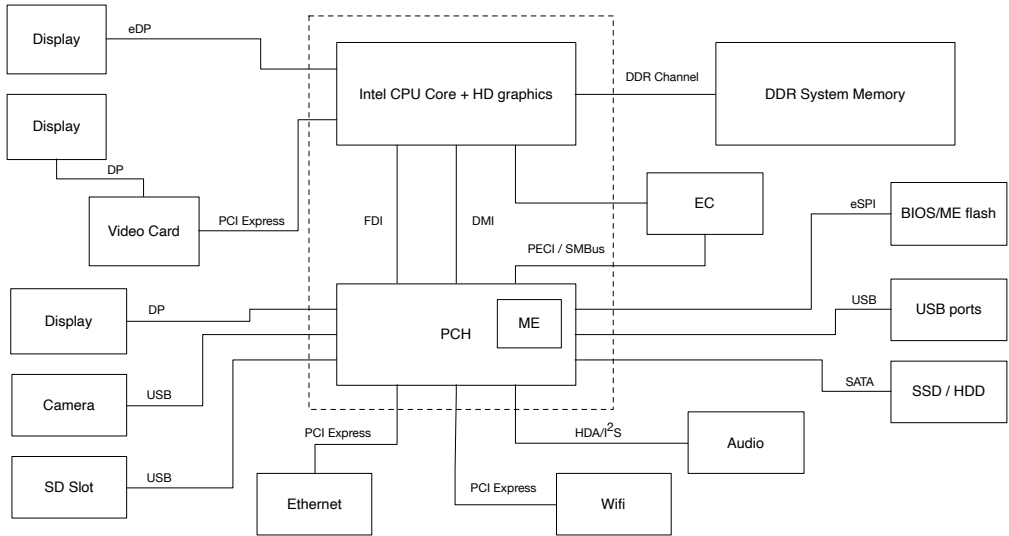


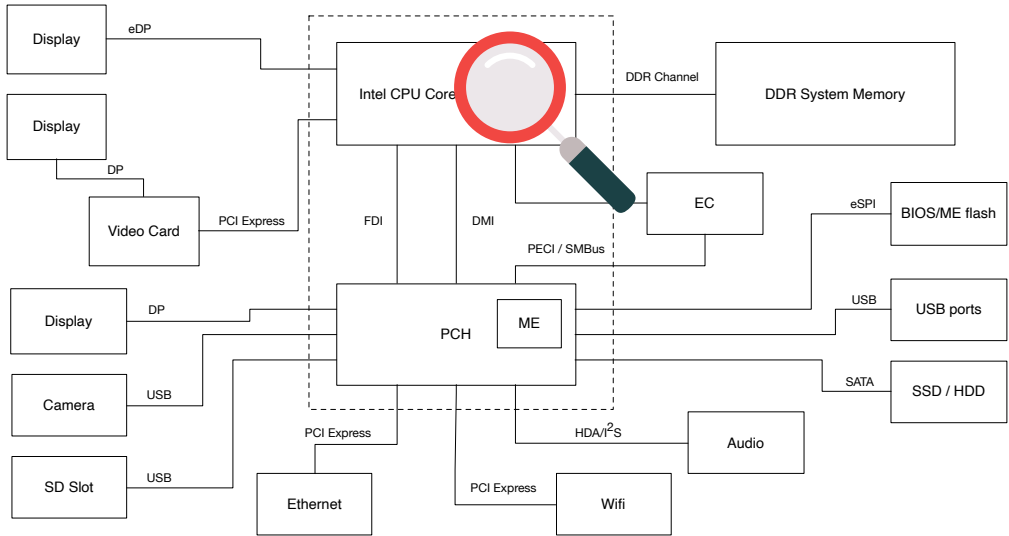
CPU #2

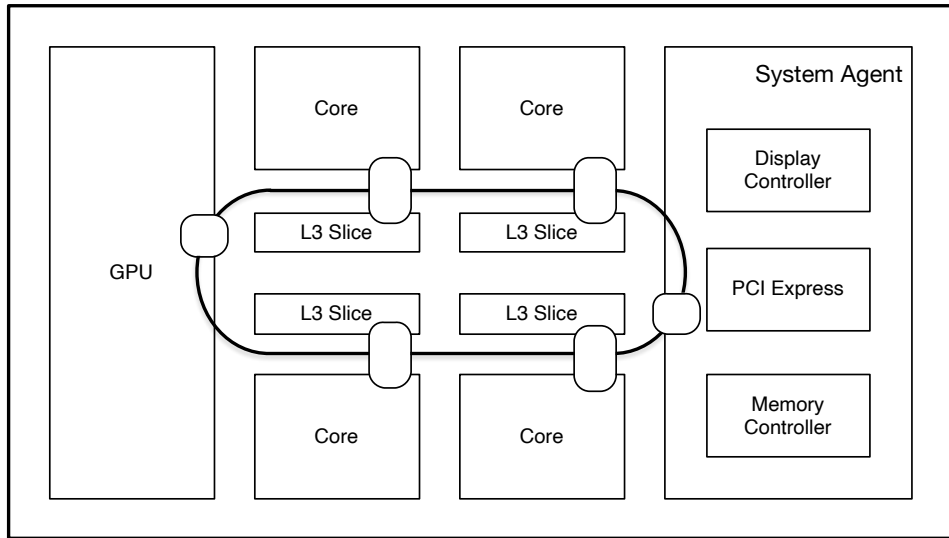


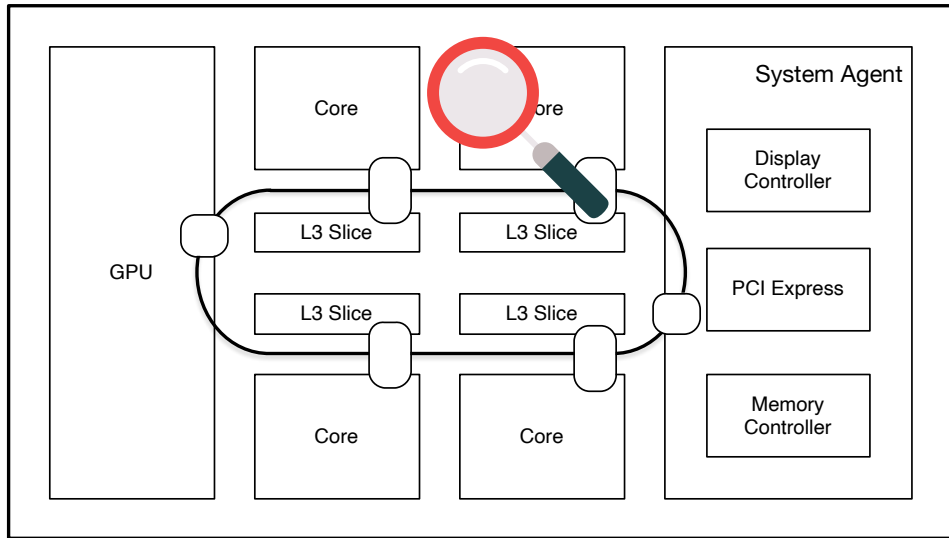
DRAM

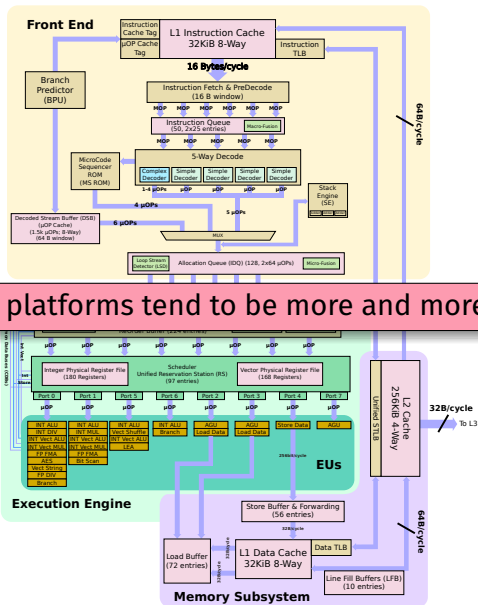










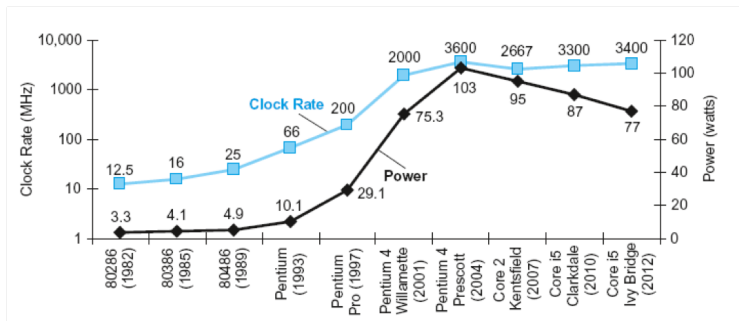


Hardware platforms tend to be more and more complex

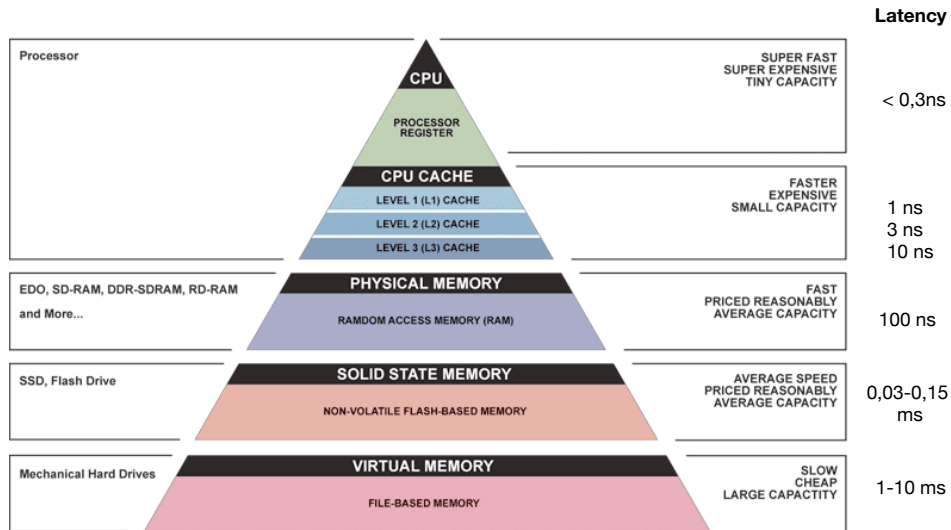
Performance summary

$$\text{CPU time} = \frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Clock cycles}}{\text{Instruction}} \times \frac{\text{Seconds}}{\text{Clock cycle}}$$

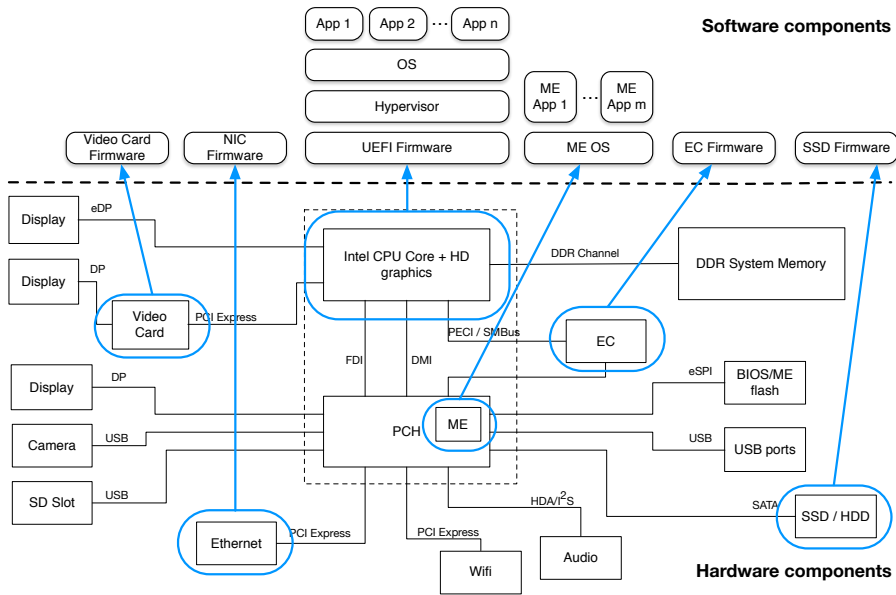
How to increase the performances?

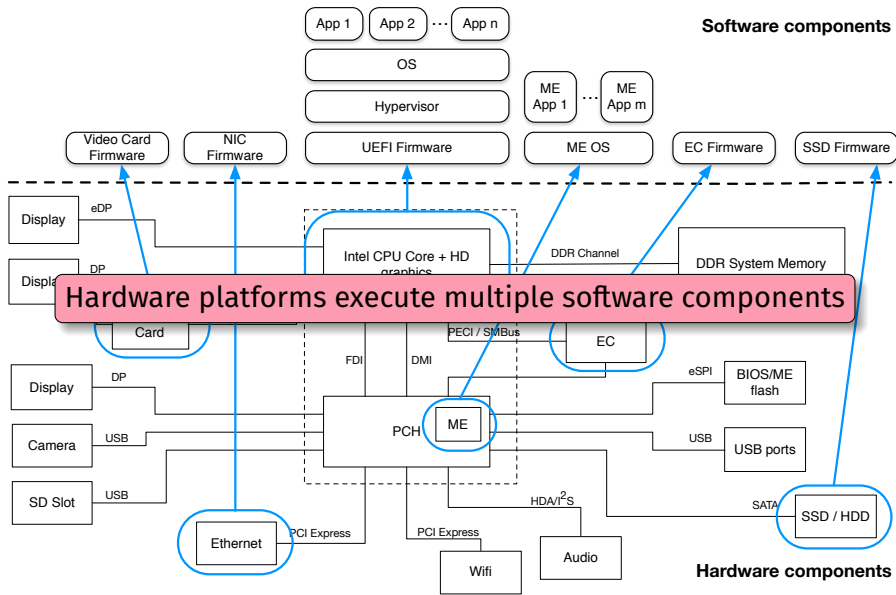


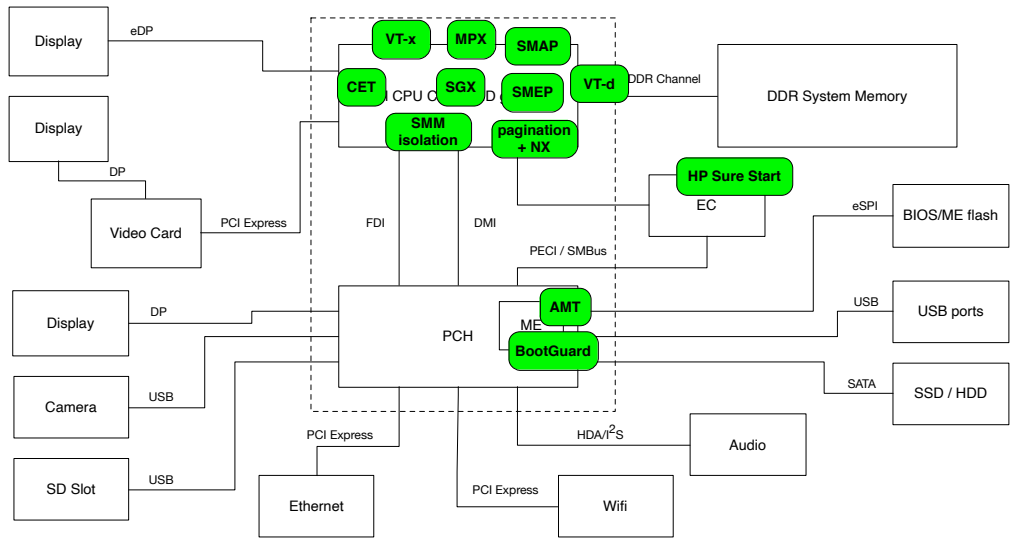
Why do we need caches?

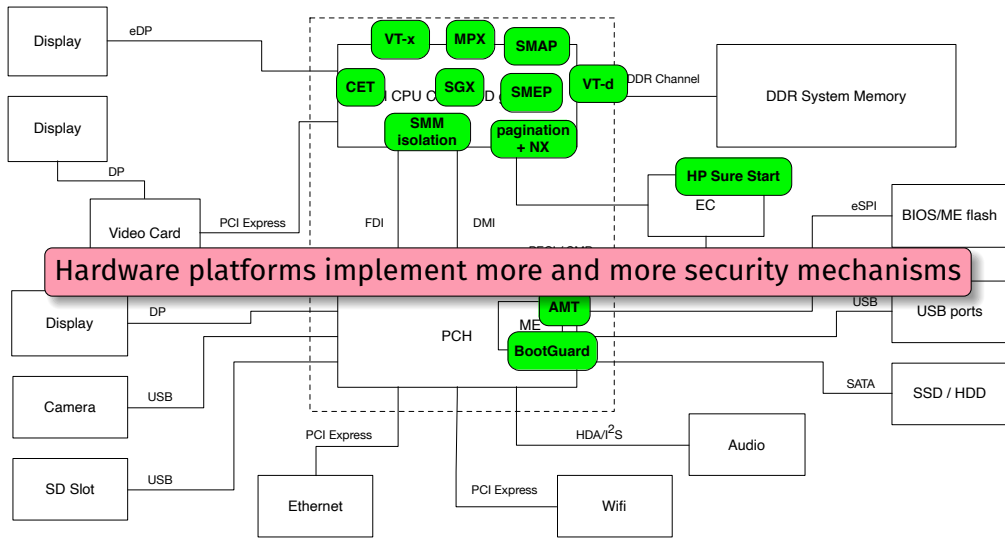


▲ Simplified Computer Memory Hierarchy
Illustration: Ryan J. Leng

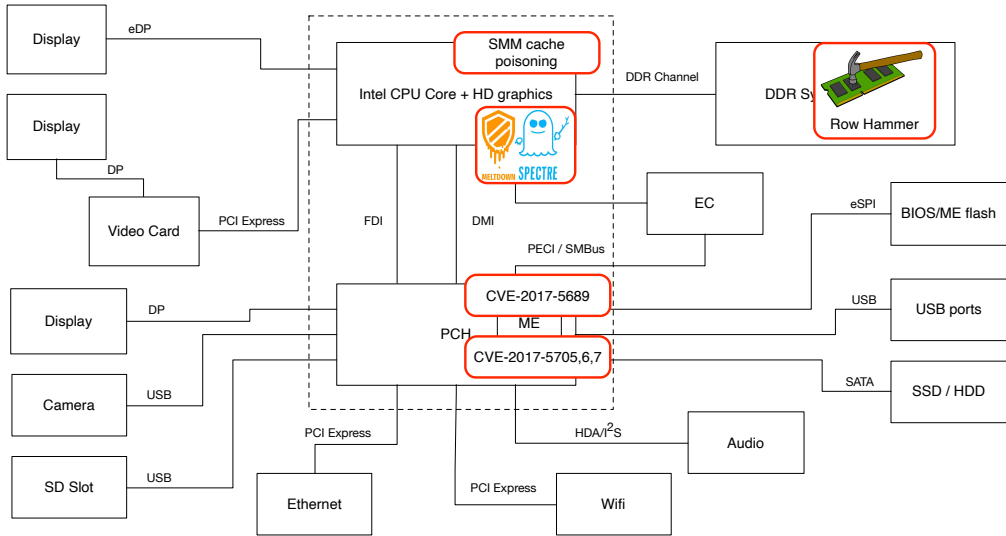


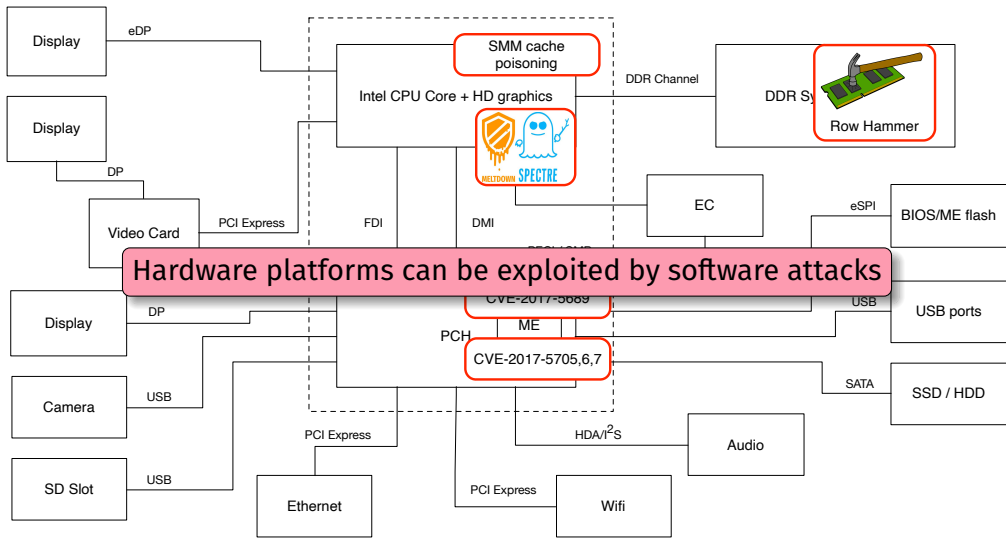






Hardware platforms implement more and more security mechanisms







1. Introduction
2. Side-channel attacks on SGX
3. Detection of Attacks Against the System Management Mode
4. SILM Thematic Semester
5. Conclusion

Side-channel attacks on SGX



- ▶ channels that are **outside of the functional specification**, i.e., that are not supposed to carry useful information



- ▶ channels that are **outside of the functional specification**, i.e., that are not supposed to carry useful information
- ▶ however they can **leak secret information**, exploiting implementations



- ▶ channels that are **outside of the functional specification**, i.e., that are not supposed to carry useful information
- ▶ however they can **leak secret information**, exploiting implementations
- ▶ can be performed by **software**, e.g., measuring memory accesses



- ▶ channels that are **outside of the functional specification**, i.e., that are not supposed to carry useful information
 - ▶ however they can **leak secret information**, exploiting implementations
 - ▶ can be performed by **software**, e.g., measuring memory accesses
- attacker monitors which cache lines are accessed, **not the content**

- ▶ **sandboxes** assume trusted system and **untrusted application**
- protects system from harm
- e.g., JavaScript in the browser





- ▶ **sandboxes** assume trusted system and **untrusted application**
 - protects system from harm
 - e.g., JavaScript in the browser
- ▶ **TEEs** assume an **untrusted system** and a trusted application
 - isolates the application
 - e.g., cloud, DRM, app that manipulates secrets...

Threat model

- ▶ malicious OS
 - ▶ only the CPU is trusted
- TEE memory is encrypted → inaccessible to the OS



Threat model

- ▶ malicious OS
 - ▶ only the CPU is trusted
- TEE memory is encrypted → inaccessible to the OS



Implementations

- ▶ Intel: SGX
- ▶ ARM and AMD: TrustZone

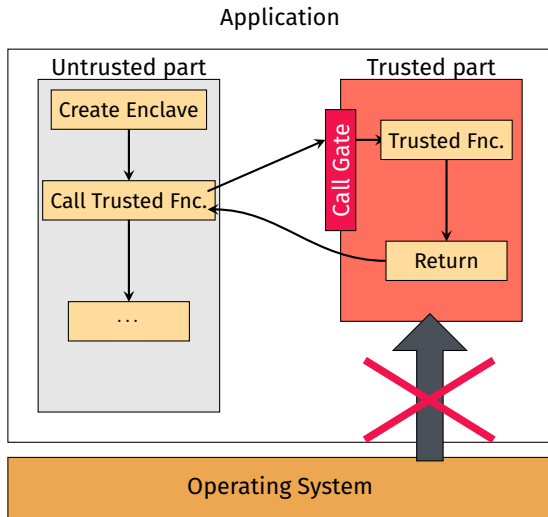
Threat model

- ▶ malicious OS
 - ▶ only the CPU is trusted
- TEE memory is encrypted → inaccessible to the OS



Implementations

- ▶ Intel: SGX
- ▶ ARM and AMD: TrustZone





Previous work showed that code inside an enclave
is not protected from side channels from outside

Previous work showed that code inside an enclave is not protected from side channels from outside

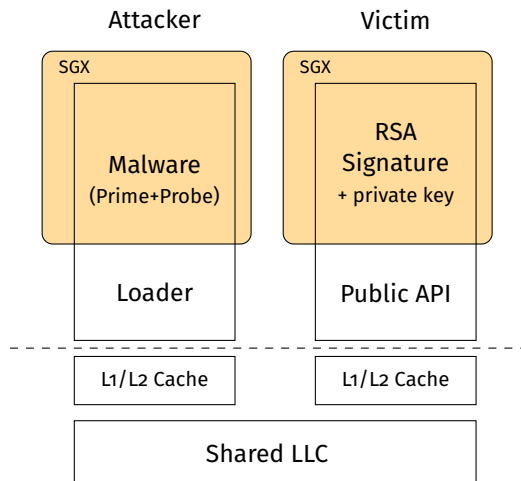


What if the enclave contains malicious code?

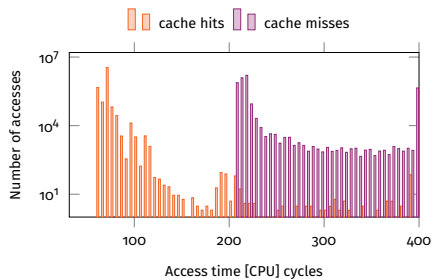
Classical exploits cannot be mounted within SGX:

- ▶ no syscalls
- ▶ no shared memory/libraries
- ▶ no interprocess communication
- ▶ blocked instructions



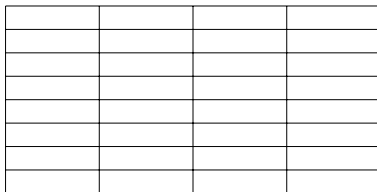


- ▶ exploits **timing differences** between:
 - cached data (fast)
 - uncached data (slow)
- ▶ targets a cache set
- ▶ works **across CPU cores** (shared LLC)

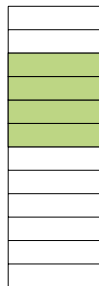




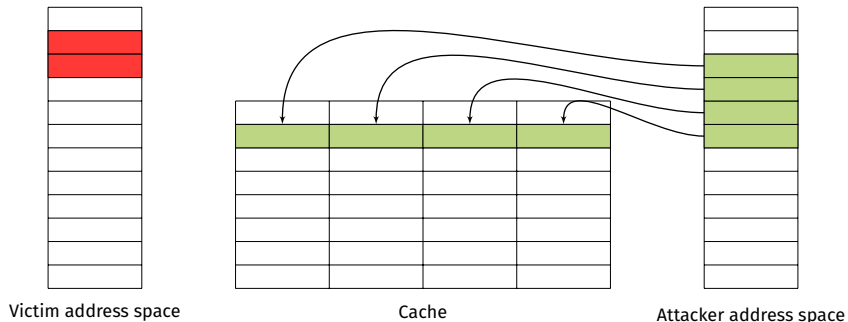
Victim address space



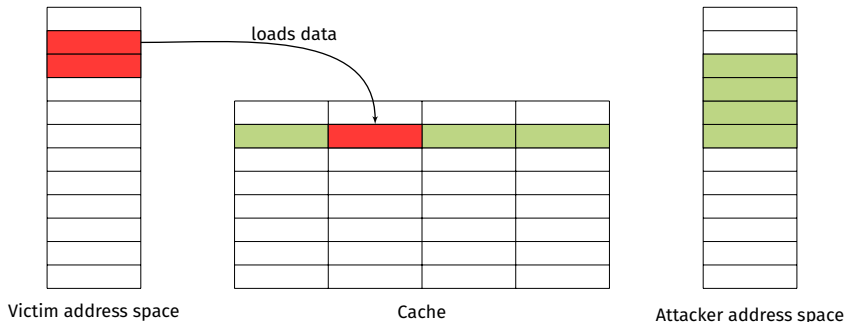
Cache



Attacker address space

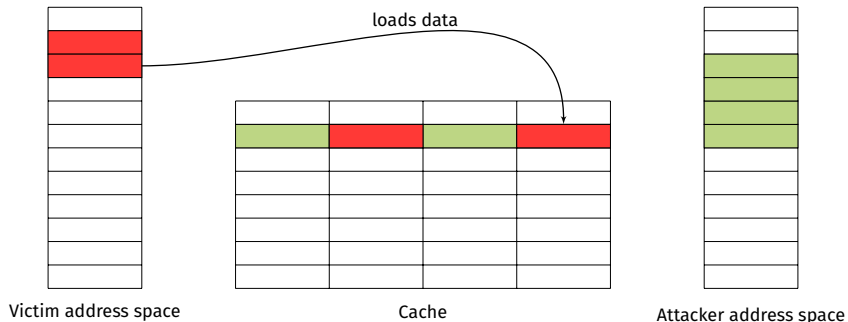


Step 1: Attacker **primes**, i.e., fills, the cache (no shared memory)



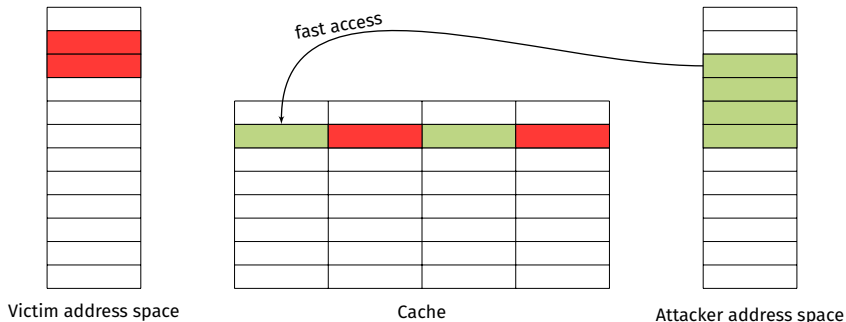
Step 1: Attacker **primes**, i.e., fills, the cache (no shared memory)

Step 2: Victim evicts cache lines while running



Step 1: Attacker **primes**, i.e., fills, the cache (no shared memory)

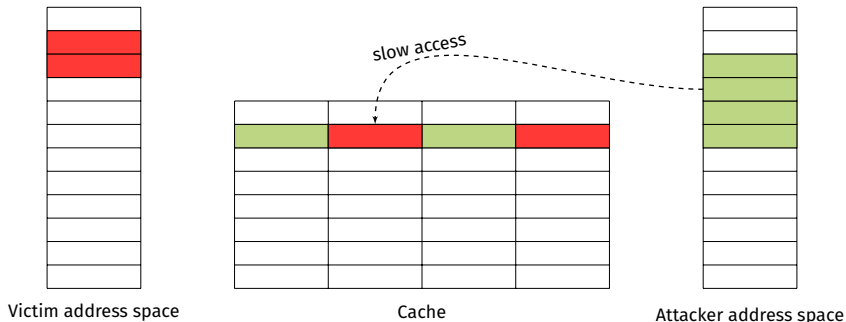
Step 2: Victim evicts cache lines while running



Step 1: Attacker **primes**, i.e., fills, the cache (no shared memory)

Step 2: Victim evicts cache lines while running

Step 3: Attacker **probes** data to determine if set has been accessed



Step 1: Attacker **primes**, i.e., fills, the cache (no shared memory)

Step 2: Victim evicts cache lines while running

Step 3: Attacker **probes** data to determine if set has been accessed



mbedTLS version 2.3.0 (fixed since)

Algorithm 1: Square-and-multiply exponentiation

Input: base b , exponent e , modulus n

Output: $b^e \pmod n$

$X \leftarrow 1$

for $i \leftarrow \text{bitlen}(e)$ **downto** 0 **do**

$X \leftarrow \text{multiply}(X, X)$

if $e_i = 1$ **then**

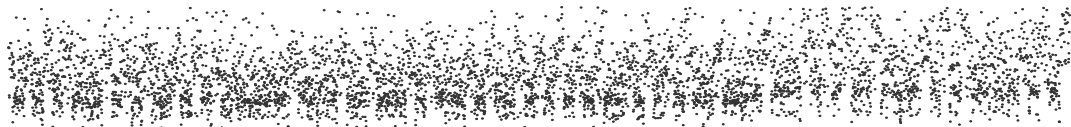
$X \leftarrow \text{multiply}(X, b)$

end

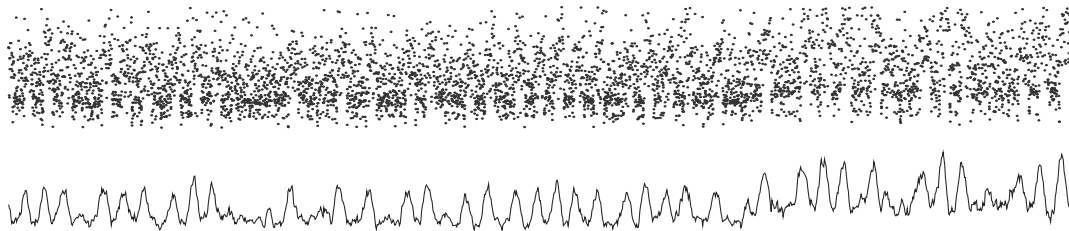
end

return X

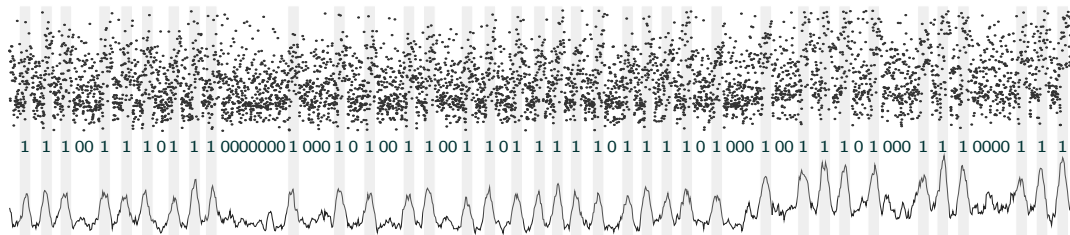
- ▶ raw Prime+Probe trace on the buffer holding the multiplier b



- ▶ raw Prime+Probe trace on the buffer holding the multiplier b
- ▶ processed with a simple moving average



- ▶ raw Prime+Probe trace on the buffer holding the multiplier b
 - ▶ processed with a simple moving average
 - ▶ allows to clearly recover the **bits of the secret exponent**
- 96% of a 4096-bit RSA key from a single trace, full key using 11 traces



We need to differentiate between events a few nanoseconds apart:

- ▶ `rdtsc` instruction does that very well
- ▶ unprivileged...



We need to differentiate between events a few nanoseconds apart:

- ▶ `rdtsc` instruction does that very well
- ▶ unprivileged...
- ▶ ... but not permitted inside SGX enclaves



We need to differentiate between events a few nanoseconds apart:

- ▶ `rdtsc` instruction does that very well
- ▶ unprivileged...
- ▶ ... but not permitted inside SGX enclaves



Build your own timer!

- ▶ start a thread that continuously **increments a global variable**
- ▶ the global variable is our timestamp

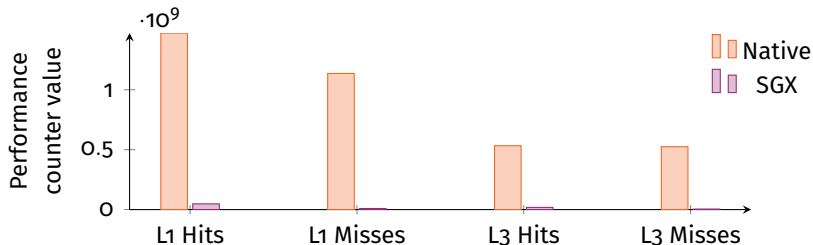


- ▶ OS cannot inspect the enclave: traditional methods fail



- ▶ OS cannot inspect the enclave: traditional methods fail
- ▶ **performance counters** can detect cache attacks → high rate of cache misses

- ▶ OS cannot inspect the enclave: traditional methods fail
- ▶ **performance counters** can detect cache attacks → high rate of cache misses
- ▶ however they are **disabled inside the enclave**





- ▶ **hardware optimizations** (e.g., cache) can create side channels to exploit weak software implementations

- ▶ **hardware optimizations** (e.g., cache) can create side channels to exploit weak software implementations
- ▶ SGX does not protect from side-channel attacks
 - **enclaves are not a magic protection**
 - Intel: "this is not in the threat model"

- ▶ **hardware optimizations** (e.g., cache) can create side channels to exploit weak software implementations
- ▶ SGX does not protect from side-channel attacks
 - **enclaves are not a magic protection**
 - Intel: "this is not in the threat model"
- ▶ SGX can actually **protect malware**
 - security features can be abused too
 - rethink enclave protection?

Detection of Attacks Against the System Management Mode

Where can we find firmware?

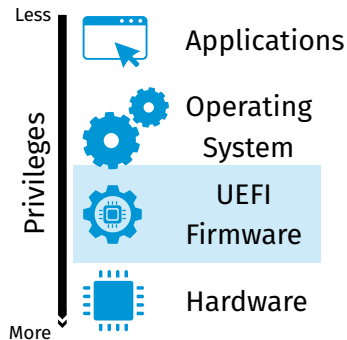
- ▶ Hard disks, network cards, etc.
- ▶ **Motherboards BIOS/UEFI**

What is it?

- ▶ Low-level software
- ▶ Tightly linked to hardware
- ▶ Stored into a dedicated flash memory

Boot time vs Runtime

- ▶ Early execution and configuration
- ▶ Highly privileged **runtime software**



BIOSs are often written in unsafe languages (i.e., C & assembly)

- ▶ Memory safety errors (e.g., use after free or buffer overflow)
- ▶ BIOSs are not exempt from vulnerabilities (Kallenberg et al. 2013; Bazhaniuk et al. 2015)

Why compromise a BIOS?

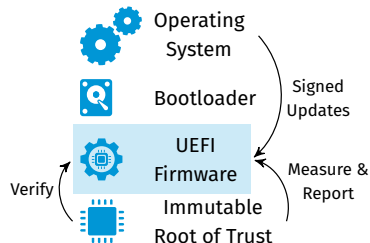
- ▶ Malware can be hard to detect (stealth)
- ▶ Malware can be persistent (survives even if the HDD/SSD is changed)

What do we want?

- ▶ Boot time integrity
- ▶ **Runtime integrity** (some platforms are rarely rebooted)

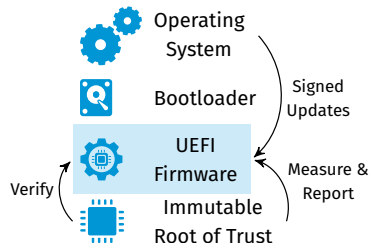
Boot time

- ▶ Signed updates
- ▶ Signature verification before executing
- ▶ Measurements and reporting to a TPM chip
- ▶ Immutable hardware root of trust



Boot time

- ▶ Signed updates
- ▶ Signature verification before executing
- ▶ Measurements and reporting to a TPM chip
- ▶ Immutable hardware root of trust



Runtime

Isolation of critical **runtime services** available while the OS is running:

- ▶ BIOS update, power management, UEFI variables handling, etc.

→ our focus is on the **System Management Mode (SMM)**



SMM is the **highest privileged** execution mode for x86 processors.

How to enter the SMM?

- ▶ Trigger a System Management Interrupt (SMI)
- ▶ SMIs code & data are stored in a protected memory region: System Management RAM (SMRAM)

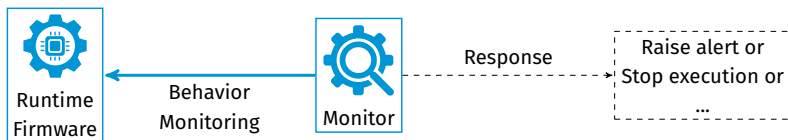
BIOS code is not exempt from vulnerabilities affecting SMM

(Bazhaniuk et al. 2015; Bulygin et al. 2017; Pujos 2016)

Why is it interesting for an attacker?

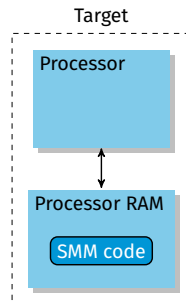
- ▶ Only mode that can write to the flash containing the BIOS
- ▶ Arbitrary code execution in SMM gives full control of the platform

Our goal is to detect attacks that modify the **expected behavior** of the SMM by **monitoring** its behavior **at runtime**.

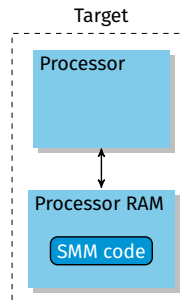
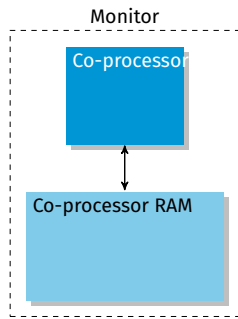


Such goal raises the following questions:

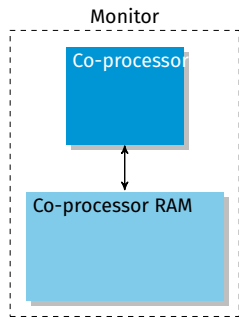
- ▶ How to ensure the integrity of the monitor?
- ▶ How to define a correct behavior?
- ▶ How to monitor?



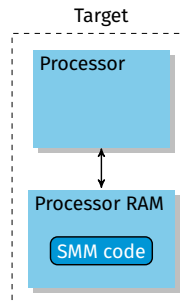
How to ensure the integrity of the monitor?

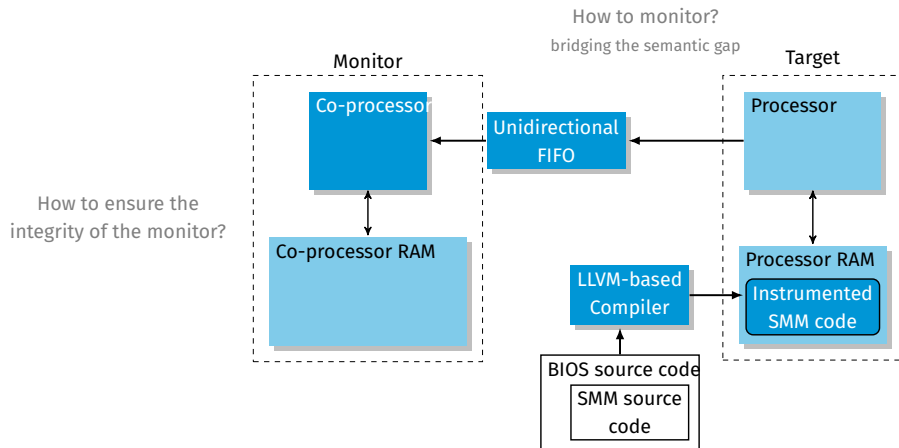


How to ensure the integrity of the monitor?

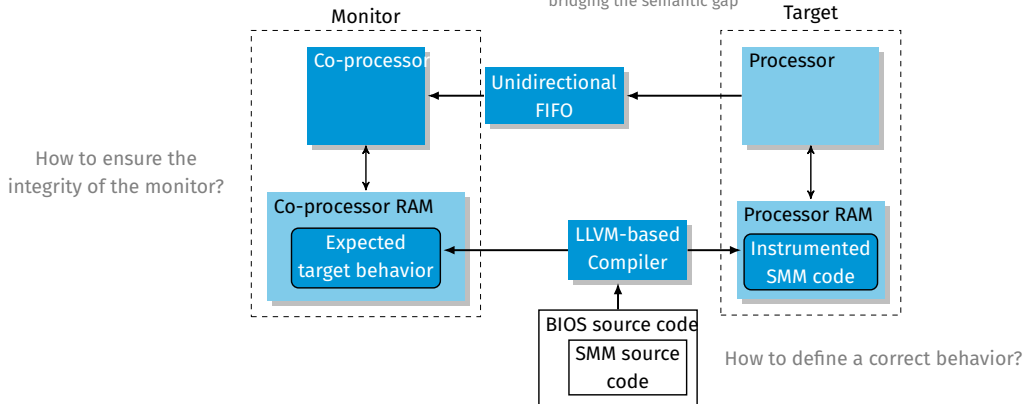


Semantic gap?





How to monitor?
bridging the semantic gap



Our use case: SMM code

- ▶ Written in unsafe languages (i.e., C & assembly)
 - Such languages are often targeted by attacks hijacking the control flow
- ▶ Tightly coupled to hardware
 - Such software modifies hardware configuration registers

Control Flow Graph (CFG)

Define the control flow that the software is expected to follow

→ Control Flow Integrity (CFI)

Invariants on CPU registers

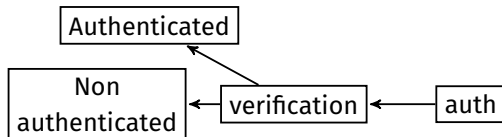
Define rules that registers are expected to satisfy

→ CPU registers integrity

Example

```
void auth(int a, int b) {  
    char buffer[512];  
  
    [...vuln...]  
  
    verification(buffer);  
}  
void verification(char *input) {  
    if (strcmp(input, "secret") == 0)  
        authenticated();  
    else  
        non_authenticated();  
}
```

Simplified graph



Example

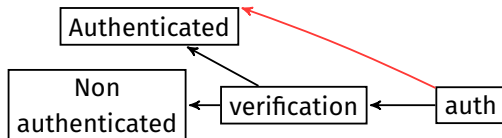
```
void auth(int a, int b) {
    char buffer[512];

    [...vuln...]

    verification(buffer);
}

void verification(char *input) {
    if (strcmp(input, "secret") == 0)
        authenticated();
    else
        non_authenticated();
}
```

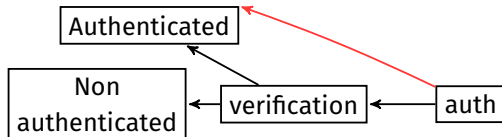
Simplified graph



Example

```
void auth(int a, int b) {  
    char buffer[512];  
  
    [...vuln...]  
  
    verification(buffer);  
}  
void verification(char *input) {  
    if (strcmp(input, "secret") == 0)  
        authenticated();  
    else  
        non_authenticated();  
}
```

Simplified graph



Goal: constrain the execution path to follow a control-flow graph (CFG)

SMM code is tightly coupled to hardware

- ▶ Generic detection methods (e.g., CFI) are not aware of hardware specificities
- ▶ Adhoc detection methods are needed

Some interesting registers for an attacker

- ▶ **SMBASE**: Defines the SMM entry point
- ▶ **CR3**: Physical address of the page directory

→ Their value is saved in memory and is not supposed to change at runtime

How to protect such registers?

- ▶ Send the expected values at boot time
- ▶ Send current values at runtime to detect any discrepancy

Security constraints

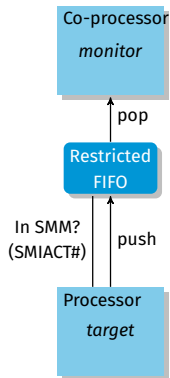
- ▶ Message integrity
- ▶ Chronological order
- ▶ Exclusive access

Performance constraints

- ▶ Acceptable latency of an SMI as defined by Intel BIOS Test Suite: 150 μ s
- ▶ More than 150 μ s per SMI handler leads to degradation of performance or user experience

Additional hardware component

- ▶ Chronological order
→ FIFO
- ▶ Message integrity
→ Restricted FIFO
- ▶ Exclusive access
→ Check if CPU is in SMM (SMIACT# signal)
- ▶ Performance
→ Use a low latency interconnect



Our prototype is implemented in a simulated and emulated environment

SMM code implementations used

- ▶ EDK2: foundation of many BIOSes (Apple, HP, Intel,...)
→ UEFI Variables SMI handlers
- ▶ coreboot: perform hardware initialization (used on some Chromebooks)
→ Hardware-specific SMI handlers

We want to emulate SMM environment and features

QEMU emulator for security evaluation

We want to simulate accurately the performance impact

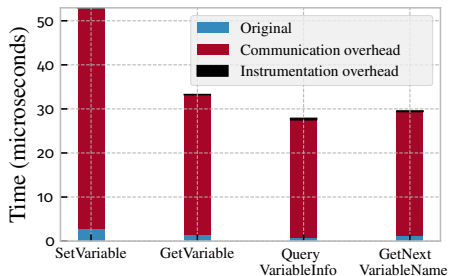
gem5 simulator for performance evaluation

We simulated attacks that exploited vulnerabilities similar to those found in real-world BIOSes

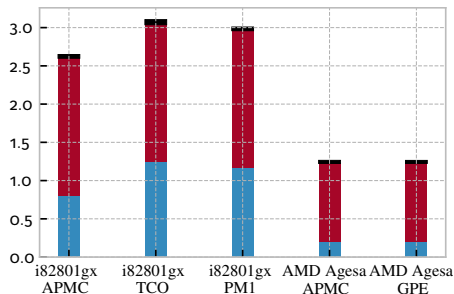
Vulnerability	Attack Target	Security Advisories	Detected
Buffer overflow	Return address	CVE-2013-3582	Yes
Arbitrary write	Function pointer	CVE-2016-8103	Yes
Arbitrary write	SMBASE	LEN-4710	Yes
Insecure call	Function pointer	LEN-8324	Yes

- ▶ Under the 150 microseconds limit defined by Intel
- ▶ Most of the communication overhead is due to the shadow call stack

EDK2



coreboot





- ▶ **SMM code** can be vulnerable to memory errors
- ▶ Attackers exploiting such vulnerabilities can gain full control of the platform



- ▶ **SMM code** can be vulnerable to memory errors
- ▶ Attackers exploiting such vulnerabilities can gain full control of the platform
- ▶ We can detect such attacks using a behavior and event-based intrusion detection system implemented on an **isolated** co-processor
- ▶ Acceptable performance ($< 150 \mu\text{s}$ Intel threshold)

- ▶ **SMM code** can be vulnerable to memory errors
- ▶ Attackers exploiting such vulnerabilities can gain full control of the platform
- ▶ We can detect such attacks using a behavior and event-based intrusion detection system implemented on an **isolated** co-processor
- ▶ Acceptable performance ($< 150 \mu\text{s}$ Intel threshold)
- ▶ What can we do next?
 - React and restore the platform in a sane state
 - Implement other detection approaches to detect other classes of attacks (e.g non-control data attacks)



SILM

Objectives

- ▶ Promote the scientific, teaching and industrial transfer activities on a specific subject
- ▶ Identify scientific and technological challenges in that field
- ▶ Propose a strategic action plan

Organization and funding

- ▶ Funded by the DGA
- ▶ Managed by Inria, on behalf of all the partners of the PEC research centre (Pôle d'excellence cyber)
- ▶ Led by one or several researchers from PEC partners

<https://semestres-cyber.inria.fr/en/>



Inria



CentraleSupélec



Actions

- ▶ Organization of different events: a summer school, dedicated workshops and a regular seminar
- ▶ Animation of a working group and publication of a white-paper
- ▶ Invitation of researchers



- 1. Analysing the behavior and state of hardware components**
 - Fuzzing
 - Reverse-engineering
 - Trace mechanisms
 - Automated specification analysis
- 2. Assessing the security of these hardware components**
 - Side-channel attacks
 - Fault injection
 - Exploiting unspecified behaviors
- 3. Detecting or preventing software attacks**
 - Using dedicated hardware components
 - Software countermeasures against hardware vulnerabilities



G. Hiet
(CS/Inria, CIDRE)

Project coordinator, Workshops, Working Group & White Book



J.-L. Lanet
(Inria LHS/CIDRE)
Workshops



C. Maurice
(CNRS, EMSEC)
Summer school



F. Tronel
(CS/Inria, CIDRE)
Workshops, Summer school



R. Lashermes
(Inria)
Seminar



- ▶ 47 participants from 7 countries: students, young researchers and engineers
- ▶ Lectures, labs and CTF, over 4 days and a half
- ▶ Organizing committee: **Clémentine Maurice**, Frédéric Tronel
- ▶ Slides and videos of the presentations available on the web site

<https://silm-school.inria.fr/>



- ▶ November 20-21 2019 in Rennes, France during the European Cyber Week
- ▶ First session in common with C&ESAR conference
- ▶ 11 invited speakers + 1 paper accepted by C&ESAR conference
- ▶ Organizing committee: **Guillaume Hiet**, **Frédéric Tronel**, Jean-Louis Lanet
- ▶ Slides and videos of the presentations available on the web site

<https://silm-workshop.inria.fr>



- ▶ June 19, 2020 in Genova, Italy
- ▶ Workshop collocated with the 5th IEEE European Symposium on Security and Privacy
- ▶ Invited speakers + CFP
- ▶ Organizing committee: **Guillaume Hiet, Frédéric Tronel**, Jean-Louis Lanet
- ▶ **Submit articles!**

<https://silm-workshop-2020.inria.fr>

SILM Seminar

- ▶ One Friday/month, 2 presentations, at Inria in Rennes, France
- ▶ Organizing committee: **Ronan Lashermes**
- ▶ Slides and videos of the presentations available on the web site

<https://semestres-cyber.inria.fr/en/silm-seminar/>

Working group

- ▶ Objectives: review the state-of-the-art, identify scientific challenges, technical obstacles, industrial transfer perspectives
- ▶ Managed by **Guillaume Hiet**
- ▶ Deliverables and outcomes: white paper + list of project proposals that should be funded in priority



Conclusion

Vulnerabilities on CPU micro-architectures and firmware are a serious issue

- ▶ Be careful of the considered threat model
- ▶ Often combined vulnerabilities in hardware (weak isolation) and software (use of vulnerable algorithms)

What can we do?

- ▶ Remove hardware optimization (HT, speculative execution, cache, etc.)?
 - Implies a huge impact on performances!
- ▶ Patch vulnerable code
 - How can we patch **all** the vulnerabilities? What about third parties?
- ▶ **Propose generic solutions combining hardware and software**
 - We need hardware support (same story as memory errors)
 - Such mechanism has to be configured by software (OS, application or compiler support)

- Bazhaniuk, Oleksandr et al. (2015). “A new class of vulnerabilities in SMI handlers”. CanSecWest, Vancouver, Canada.
- Bulygin, Yuriy et al. (2017). “BARing the System: New vulnerabilities in Coreboot & UEFI based systems”. REcon Brussels.
- Chevalier, Ronny et al. (Dec. 2017). “Co-processor-based Behavior Monitoring: Application to the Detection of Attacks Against the System Management Mode”. In: *ACSAC 2017 - 33rd Annual Computer Security Applications Conference*. Vol. 2017. Proceedings of the 33rd Annual Computer Security Applications Conference. Orlando, United States: ACM, pp. 399–411. DOI: 10.1145/3134600.3134622. URL: <https://hal.inria.fr/hal-01634566>.
- Kallenberg, Corey et al. (2013). “Defeating Signed BIOS Enforcement”. EkoParty, Buenos Aires.
- Pujos, Bruno (2016). *SMM unchecked pointer vulnerability*. URL: <http://esec-lab.sogeti.com/posts/2016/05/30/smm-unchecked-pointer-vulnerability.html>.
- Schwarz, Michael et al. (2017). “Malware Guard Extension: Using SGX to Conceal Cache Attacks”. In: *DIMVA*.



We focus on indirect branches integrity

Type-based verification

Ensures the integrity of indirect calls

```
typedef struct SomeStruct {
    [...]
    char (*foo)(int);
} SomeStruct;
int bar(SomeStruct *s) {
    char c;
    [...]

    c = s->foo(31);
    [...]
}
```



We focus on indirect branches integrity

Type-based verification

Ensures the integrity of indirect calls

```
typedef struct SomeStruct {  
    [...]  
    char (*foo)(int);  
} SomeStruct;  
int bar(SomeStruct *s) {  
    char c;  
    [...]  
  
    c = s->foo(31);  
    [...]  
}
```



We focus on indirect branches integrity

Type-based verification

Ensures the integrity of indirect calls

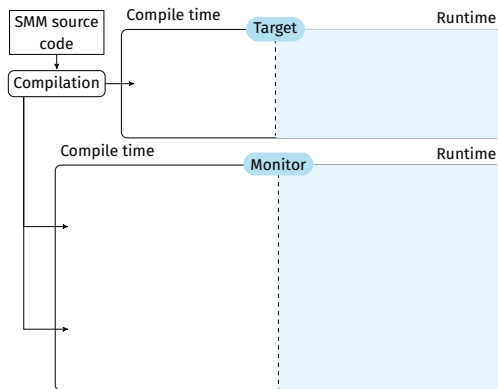
```
typedef struct SomeStruct {  
    [...]  
    char (*foo)(int);  
} SomeStruct;  
  
int bar(SomeStruct *s) {  
    char c;  
    [...]  
  
    c = s->foo(31);  
    [...]  
}
```

We focus on indirect branches integrity

Type-based verification

Ensures the integrity of indirect calls

```
typedef struct SomeStruct {  
    [...]  
    char (*foo)(int);  
} SomeStruct;  
int bar(SomeStruct *s) {  
    char c;  
    [...]  
  
    c = s->foo(31);  
    [...]  
}
```

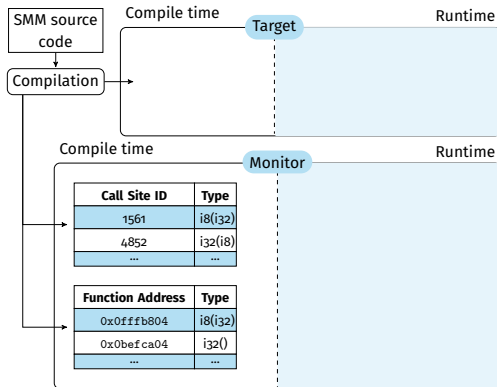


We focus on indirect branches integrity

Type-based verification

Ensures the integrity of indirect calls

```
typedef struct SomeStruct {  
    [...]  
    char (*foo)(int);  
} SomeStruct;  
int bar(SomeStruct *s) {  
    char c;  
    [...]  
  
    c = s->foo(31); /* Call Site ID = 1561 */  
    [...]  
}
```

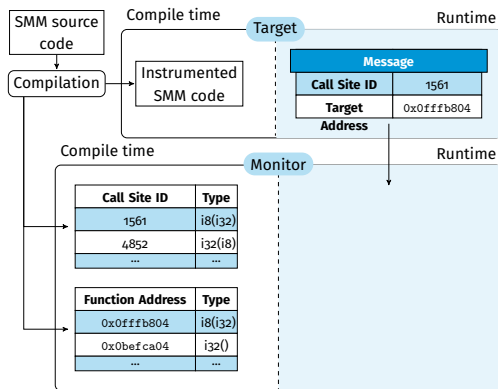


We focus on indirect branches integrity

Type-based verification

Ensures the integrity of indirect calls

```
typedef struct SomeStruct {  
    [...]  
    char (*foo)(int);  
} SomeStruct;  
int bar(SomeStruct *s) {  
    char c;  
    [...]  
    [SendMessage(1561, s->foo)]  
    c = s->foo(31); /* Call Site ID = 1561 */  
    [...]  
}
```

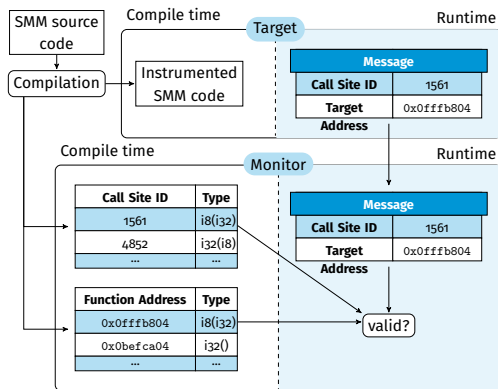


We focus on indirect branches integrity

Type-based verification

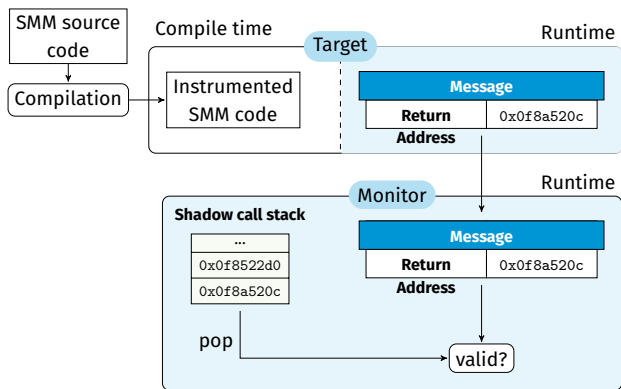
Ensures the integrity of indirect calls

```
typedef struct SomeStruct {  
    [...]  
    char (*foo)(int);  
} SomeStruct;  
int bar(SomeStruct *s) {  
    char c;  
    [...]  
  
    [SendMessage(1561, s->foo)]  
    c = s->foo(31); /* Call Site ID = 1561 */  
    [...]  
}
```



Shadow call stack

Ensures integrity of the return address on the stack



Our analysis with EDK II gave:

- ▶ 158 equivalence classes of size 1,
- ▶ 24 of size 2,
- ▶ 42 of size 3,
- ▶ 2 of size 5,
- ▶ 1 of size 9,
- ▶ and 1 of size 13.

- ▶ Slide 3, Google
- ▶ Slide 8, WikiChip <https://en.wikichip.org/>
- ▶ Slide 9, Instructor materials for "Computer Organization and Design, RISC-V Edition", Patterson & Hennessy, 2018
- ▶ Slide 10, Rian J. Lang
- ▶ Slides 16, 20 icons made by Smashicons from www.flaticon.com
- ▶ Slides 17, 19, 15, 26 icons made by Freepik from www.flaticon.com
- ▶ Slide 15, icon made by monkik from www.flaticon.com

Name	Author	License
Application	Christopher	CC BY 3.0 US
Chip Settings	Luis Rodrigues	CC BY 3.0 US
Gear	Jonathan Higley	CCo 1.0 Universal
Harddrive	Creaticca Creative Agency	CC BY 3.0 US
Microchip	Creative Stall	CC BY 3.0 US
Research	Gregor Cresnar	CC BY 3.0 US