

Using software diversity to prevent cache sub-channel attack cross-vm



Outline

- Diversity
- Side channel attack
- Side channel attack cross-vm
- Current countermeasures

• Diversity



Side channel attack cross-vm

Diversity

- Biodiversity
- Culture Diversity
- Similarly to natural systems, software systems including diverse functions and elements are able to cope with many kinds of anticipatable problems and failure



Common ways to implement software diversity

- Manual (timeline of software production)
 - \circ Requirements
 - Different design versions of program
 - Development diversity(example)
- Automatic
 - \circ Randomization
 - Domain-specific Diversity
 - $\circ~$ Integrated Diversity

```
program Adhoc;
function Add( x, y : Integer ) : Integer;
begin
    Add := x + y
end:
function Add( s, t : String ) : String;
begin
    Add := Concat( s, t )
end;
begin
                                           (* Prints "3"
    Writeln(Add(1, 2));
*)
    Writeln(Add('Hello, ', 'World!'));
                                           (* Prints
"Hello, World!" *)
end.
```

Different types of side-channel attacks cross-vm in the cloud

- In the cloud,
 Several different virtual machines can be in the same host.
 Networking side channel attack.
 - e.g co-resident watermarking. Try to insert malicious code into other co-resident virtual machines.



•Cache-based side channel attacks.

Timing attack and Cache timing attack

"Timing attack based on the leakage of information of secret parameters through variations in the running times of a cryptographic device."

Cache timing, timing attack on cache☺



Remote Access host	
REMOTE ACCESS	HOST
Login Information	Host Identification
Username or email miohn	Host Description work
Password IIIII	Host Key
S New Liser Signup	Confirm Host Key



Copyright, Pro Softnet Corporation, All rights reserve

Why preventing cache timing attack from different virtual machines matters?



Cloud

VM Isolation





Vulnerabilities of encryption libraries

Deduplication



Types of cache timing side channel attacks

According to procedures:

- 1) PRIME + PROBE
- 2) EVICT + TIME
- 3) FLUSH + RELOAD(cloud)
- According to which round of encryption:
- 1) First round + second round attack
- 2) Final round

Cache Architecture

- When any memory byte is needed, its place in cache is calculated;
- CPU asks the cache;
- If there, the cache returns the data;
- If not, the data is pulled in from memory;
- If the calculated cache line is occupied by data with a different tag, that data is evicted.
- If the line is *dirty* (modified) it is written back to memory first.



Cache miss Cache hit

PRIME + PROBE

- Step 1 Fill cache with attacker's own data(will be change for sure)
- Step 2 Trigger encryption
- Step 3 Load the target information want to check and timing(e.g AES will load Ttable, check which entries in the T-table is accessed during encryption.)

Evict + Time

- Step 1 Triger encrypt, calculate average encryption time t1
- Step 2 Evict specific cache set
- Step 3 Encrypt again, timing the encryption time t2, if t2 > t1, means that cache set is used during encryption.
- In cloud: require targeting VM and attacker VM are using the same core, so they are sharing L1, L2 cache, since it requires attack can modify the cache content

Memory Deduplication

- Improve the memory utilization by
 recognizing processes(or VMs) that
 place the same data in memory
 e.g. Two virtual machine share libraries
 Implemented: VMware (Transparent
 - Page Sharing- TPS)

KVM(Kenel same-page

KSM)

		Virtual machines	
Private pages	Private pages	Private pages	
Shared (common) pages Hypervisor			
	Physical machine		

AES Encryption Processes



https://www.youtube.com/watch?v=mlzxpkdXP58

T-table



FLUSH + RELOAD

• Premise: attacker can monitor of a single memory line

Step 1 Attack uses the clflush command to flush desired memory lines from the cache to make sure that they have to be retrieved from the main memory next time

Step 2 Encryption

Step 3 Timing and tell whether had access that monitored memory line by reloading the cache line again.

$$c_i = k_i \oplus T\left[s_{[i]}\right]$$

Countermeasures for what?Reason? Tell the problem and explain the solution

- Pre-fetching
- Disable memory duplication
- Software diversity: diversify
- Hardware mechanism: separation cache used in L3

