# CacheBleed: A Timing Attack on OpenSSL Constant Time RSA

Yuval YaromThe University of Adelaide and Data61Daniel GenkinTechnion and Tel-Aviv UniversityNadia HeningerUniversity of Pennsylvania

### Attack Life Cycle



## Round 1



## Cache Structure



- Stores fixed-size *lines*
- Arranged as multiple sets, each consisting of multiple ways.
- Each memory line maps to a single cache set
  - Can be cached in any of the ways in the set

## The Prime+Probe attack [Per'05,OST'05]



Memory

Cache



- Choose a cache-sized
   memory buffer
- Access all the lines in the buffer, filling the cache
- Victim executes, evicting some of the buffer lines from the cache
- Measure the time to access the buffer
  - Accesses to cached lines is faster than to evicted lines

## **Fixed Window Exponentiation**

Algorithm 1: Fixed-window exponentiation

**input** : window size w, base a, modulus k, n-bit exponent  $b = \sum_{i=0}^{\lfloor n/w \rfloor} b_i \cdot 2^{wi}$ **output**:  $a^b \mod k$ 

*I*/*Precomputation*  $a_0 \leftarrow 1$ **for**  $j = 1, \dots, 2^w - 1$  **do**  $\mid a_j \leftarrow a_{j-1} \cdot a \mod k$ **end** 

```
\begin{array}{l} \textit{I/Exponentiation} \\ r \leftarrow 1 \\ \textbf{for } i = \lceil n/w \rceil - 1, \dots, 0 \ \textbf{do} \\ \mid & \textbf{for } j = 1, \dots, w \ \textbf{do} \\ \mid & r \leftarrow r^2 \ \text{mod} \ k \\ \textbf{end} \\ \quad & r \leftarrow r \cdot a_{b_i} \ \text{mod} \ k \\ \textbf{end} \\ \textbf{return } r \end{array}
```

#### Prime+Probe and Windowed Exponentiation





#### Scatter-Gather

- Mitigate Prime+Probe
  - Sequence of accesses to cache lines does not depend on secret data

offset	0	1	2		63	offset	0	1	2		63
Line 0	M <sub>0</sub> [0]	<b>M</b> <sub>0</sub> [1]	<b>M</b> <sub>0</sub> [2]	•••	<b>M</b> <sub>0</sub> [63]	Line 0	$\mathbf{M}_{0}[0]$	<b>M</b> <sub>1</sub> [0]	<b>M</b> <sub>2</sub> [0]	•••	M <sub>63</sub> [0]
Line 1	M <sub>0</sub> [64]	M <sub>0</sub> [65]	<b>M</b> <sub>0</sub> [66]	•••	<b>M</b> <sub>0</sub> [127]	Line 1	<b>M</b> <sub>0</sub> [1]	M <sub>1</sub> [1]	<b>M</b> <sub>2</sub> [1]	•••	<b>M</b> <sub>63</sub> [1]
Line 2	<b>M</b> <sub>0</sub> [128]	M <sub>0</sub> [129]	M <sub>0</sub> [130]	•••	<b>M</b> <sub>0</sub> [191]	Line 2	M <sub>0</sub> [2]	<b>M</b> <sub>1</sub> [2]	M <sub>2</sub> [2]	•••	M <sub>63</sub> [2]
Line 3	<b>M</b> <sub>1</sub> [0]	<b>M</b> <sub>1</sub> [1]	<b>M</b> <sub>1</sub> [2]	•••	M <sub>1</sub> [63]	Line 3	M <sub>0</sub> [3]	M <sub>1</sub> [3]	M <sub>2</sub> [3]	•••	M <sub>63</sub> [3]
Line 4	M <sub>1</sub> [64]	M <sub>1</sub> [65]	M <sub>1</sub> [66]	•••	M <sub>1</sub> [127]	Line 4	<b>M</b> <sub>0</sub> [4]	<b>M</b> <sub>1</sub> [4]	M <sub>2</sub> [4]	•••	M <sub>63</sub> [4]
Line 5	<b>M</b> <sub>1</sub> [128]	M <sub>1</sub> [129]	M <sub>1</sub> [130]	•••	<b>M<sub>1</sub>[191]</b>	Line 5	M <sub>0</sub> [5]	<b>M</b> <sub>1</sub> [5]	M <sub>2</sub> [5]	•••	M <sub>63</sub> [5]
Line 6	M <sub>2</sub> [0]	M <sub>2</sub> [1]	M <sub>2</sub> [2]	•••	M <sub>2</sub> [63]	Line 6	<b>M</b> <sub>0</sub> [6]	<b>M</b> <sub>1</sub> [6]	M <sub>2</sub> [6]	•••	M <sub>63</sub> [6]
Line 7	M <sub>2</sub> [64]	M <sub>2</sub> [65]	M <sub>2</sub> [66]	•••	M <sub>2</sub> [127]	Line 7	M <sub>0</sub> [7]	<b>M</b> <sub>1</sub> [7]	M <sub>2</sub> [7]	•••	M <sub>63</sub> [7]
Line 8	M <sub>2</sub> [128]	M <sub>2</sub> [129]	M <sub>2</sub> [130]	•••	M <sub>2</sub> [191]	Line 8	<b>M</b> <sub>0</sub> [8]	<b>M</b> <sub>1</sub> [8]	M <sub>2</sub> [8]	•••	M <sub>63</sub> [8]
	•	•	•		•		•	٠	•		•
	٠	•	•		•		•	٠	•		•
	•	•	•		•		•	•	•		•
Line 191	M <sub>63</sub> [128]	M <sub>63</sub> [129]	M <sub>63</sub> [130]	• • •	M <sub>63</sub> [191]	Line 191	M <sub>0</sub> [191]	M <sub>1</sub> [191]	M <sub>2</sub> [191]	• • •	M <sub>63</sub> [191] 9

#### **OpenSSL's Layout**





## Cache banks

- To support superscalar processing the cache is divided into cache banks
  - Bits 2-5 of the address determine the bank
- In Sandy Bridge, each bank can serve only one request per cycle.
  - Concurrent access to the same bank causes delays
  - Concurrent access to different banks is always possible





### CacheBleed

1	rdtscp		
2	movq	%rax, %r10	
4			0
4	addl	0x000(%r9),	%eax
5	addl	0x040(%r9),	%ecx
6	addl	0x080(%r9),	%edx
7	addl	0x0c0(%r9),	%edi
8	addl	0x100(%r9),	%eax
9	addl	0x140(%r9),	%ecx
10	addl	0x180(%r9),	%edx
11	addl	0x1c0(%r9),	%edi
•			
•			
•			
256	addl	0xf00(%r9),	%eax
257	addl	0xf40(%r9),	%ecx
258	addl	0xf80(%r9),	%edx
259	addl	0xfc0(%r9),	%edi
$\mathcal{O} \subset \mathcal{I}$			
ZØI	ratscp		
262	subq	%r10, %rax	

## CacheBleed timing



 Need multiple samples to determine cachebank conflicts

## CacheBleed on OpenSSL



- Average of 1,000 sequences on each bin
- Odd and even bins have different timing characteristics

### CacheBleed on OpenSSL - Details



## **Clock Drift**



#### Low-pass filter



#### Normalised + resampled



## Results

16,000 decryptions (1,000 sequences per bin per exponentiation)

– Less than 5 minutes online attack

- Recover three bits of each multiplier
   Miss the first and last one or two multipliers
- Use the Heninger-Shacham algorithm to reconstruct the private key
  - Two CPU hours less than 3 minutes on a highend server.





## OpenSSL "Fix"

- Use 128-bit reads with masking
   Only leaks 2 bits per multiplier
- Read at a different offset in each of the four cache lines

- Order depends on the multiplier



