An Efficient Indexing Technique for AES Lookup Table to Prevent Side-Channel Cache-Timing Attack

Refazul Islam Refat
Dept. of Computer Science and Engineering, University of Dhaka, Dhaka, Bangladesh
E-mail: refazul.refat@gmail.com

Euna Islam and Md. Mosaddek Khan
Dept. of Computer Science and Engineering, University of Dhaka, Dhaka, Bangladesh
E-mail: euna.islam@gmail.com; mosaddek@ese.univdhaka.edu

Received: 15 May 2018; Accepted: 26 June 2018; Published: 08 September 2018

Abstract—In the era of virtualization, co-residency with unknown neighbours is a necessary evil and leakage of information through side-channels is an inevitable fact. Preventing this leakage of information through side-channels, while maintaining high efficiency, has become one of the most challenging parts of any implementation of the Advanced Encryption Standard (AES) that is based on the Rijndael Cipher. Exploiting the associative nature of the cache and susceptible memory access pattern, AES is proved to be vulnerable to side-channel cache-timing attacks. The reason of this vulnerability is primarily ascribed to the existence of correlation between the index Bytes of the State matrix and corresponding accessed memory blocks. In this paper, we idealized the model of cache-timing attack and proposed a way of breaking this correlation through the implementation of a Random Address Translator (RAT). The simplicity of the design architecture of RAT can make itself a good choice as a way of indexing the lookup tables for the implementers of the AES seeking resistance against side-channel cache-timing attacks.


I. INTRODUCTION

Out of the five candidate algorithms (MARS [1], RC6, Rijndael [2], Serpent [3], and Twofish [4]) for the AES model, Rijndael was pronounced as a new standard on November 26, 2001 as FIPS PUB 197 [5]. The security barrier of Rijndael cipher is so strong that a cryptographic break is infeasible with current technology. Even Bruce Schneier, a developer of the competing algorithm Twofish, admired Rijndael cipher in his writing, “I do not believe that anyone will ever discover an attack that will allow someone to read Rijndael traffic” [6].

A brute force method would require $2^{128}$ operations for the full recovery of an AES-128 key. A machine that can perform 8.2 quadrillion calculations per second will take 1.3 quadrillion years to recover this key. Nevertheless, partial information leaked by side-channels can tear down this complexity to a very reasonable level. Side-channel attacks do not attack the underlying cipher; instead, they attack implementations of the cipher on systems that inadvertently leak data. Ongoing research in the last decade has shown that the information transmitted via side-channels, such as execution time [7], computational faults [8], power consumption [9] and electromagnetic emissions [10, 11, 12], can be detrimental to the security of Rijndael [13] and other popular ciphers like RSA [14].

In this paper, we will be primarily focusing on the vulnerability of Rijndael AES to side-channel cache-timing attack and see how simple cache misses in a multiprogramming environment can lead to disastrous consequences. If the plaintext is known, recovery of half of the key by any privileged process is a matter of seconds. The prevention is not trivial; there exists trade-off between performance and degree of multiprogramming. We will present a concept that balance this trade-off by introducing a little memory overhead. It is worth mentioning that even though the process is arbitrary, the cipher text will always remain the same.

II. RELATED LITERATURE

In October 2005, Dag Arne Osvik, Adi Shamir and Eran Tromer presented a paper demonstrating several cache-timing attacks against AES [15]. One attack was able to obtain an entire AES key after only 800 operations triggering encryptions, in a total of 65 milliseconds. The attack required the attacker to be able to run programs on the same system or platform that is performing AES. That paper presented some decent defense against side-channel cache-timing attacks such as avoiding memory access, disabling cache sharing, 30
dynamic table storage etc. They revised their paper on 2010 [16] and focused on modern microprocessor implementations and presented few more preventions. Most of them lacked conceptual consolidation and provided as implicit implication. Some of them even requires alteration of cipher text that might cause significant overhead in distributed systems to maintain synchronization. Other works included introduction of asynchronous circuitry in AES implementation [7].

Before understanding the weakest part of the cipher, we are going to clarify the concept of cache memory and how they are accessed during real time AES implementation in the remainder of this section.

A. Memory Access in AES Implementations

The memory access patterns of AES are particularly susceptible to cryptanalysis. The cipher is abstractly defined by algebraic operations and could, in principle, be directly implemented using just logical and arithmetic operations. However, performance-oriented software implementations on 32-bit (or higher) processors typically use an alternative formulation based on lookup tables as prescribed in the Rijndael Specification [13]. A number of lookup tables are precomputed once by the programmer or during system initialization. Here, there are 8 such tables, $T_0, T_1, T_2, T_3$ and $T_4, T_5, T_6, T_7$, each containing 256 4-byte words. The contents of the tables, defined in [13], are inconsequential to most of the cache timing attacks because of memory protection.

During key setup, a given 16-byte secret key $k = (k_0, k_1, ..., k_{15})$ is expanded into 10 round keys, $K^{(r)}$ or $r = 1, 2, ..., 10$. Each of these rounds is divided into 4 words of 4 bytes each: $K^{(r)} = (K_0^{(r)}, K_1^{(r)}, K_2^{(r)}, K_3^{(r)})$. The 0th round key is just the raw key: $K_0^{(0)} = (k_{4j}, k_{4j+1}, k_{4j+2}, k_{4j+3})$ for $j = 0, 1, 2, 3$. The details of the rest of the key expansions are mostly inconsequential [16].

Given a 16-byte plaintext $p = (p_0, p_1, ..., p_{15})$, encryption proceeds by computing a 16-byte intermediate state $x^{(r)} = (x^{(r)}_0, x^{(r)}_1, ..., x^{(r)}_{15})$ at each round $r$. The initial state $x^{(0)}$ is computed by $x^{(0)}_i = p_i \oplus k_i$ for $i = 0, 1, ..., 15$. Thus, the first 9 rounds are computed by updating the intermediate state as Equation 1, for $r = 0, 1, ..., 8$.

![Fig. 1. The Figure Demonstrates how $x^{(r)}_0, x^{(r)}_1, x^{(r)}_2, x^{(r)}_3$ are Computed. Note that the Bytes of the State Matrix are being used as index of the Lookup Tables. Since the Lookup Table Remains the same for a Particular S-Box, An Index byte will always Cause Lookup from the same block of Cache Memory.](image-url)
Finally, to compute the last round, the above process of table lookup is repeated with $r = 9$, except that $T_0, T_1, T_2, T_3$ are replaced by $T_0^{(10)}, T_1^{(10)}, T_2^{(10)}, T_3^{(10)}$. The resulting $x^{(10)}$ is the cipher text. Compared to the algebraic formulation of AES, here the lookup tables represent the combination of ShiftRows, MixColumns and SubBytes operations; the change of lookup tables in the last round is due to the absence of MixColumns [16]. It is clear from the above discussion that computation of a state matrix requires 16 Table lookups and 16 XOR operations.

We know, the first round state matrix is computed by XOR operation between the 16-byte plain text and the 16-byte plain key. One important point to notice from the previous discussion is that the state bytes are being used as index to the lookup tables. Thus any information of the accessed index will directly translate to Encryption Key Byte. For example, assume we know $x_i^{(0)}$. Since plaintext is known for triggered encryptions, we know $x_i^{(0)}$ too (Equation 2).

$$x_i^{(0)} ← x_i^{(0)} ⊕ k_i^{(0)}$$ (2)

This implies -

$$k_i^{(0)} ← x_i^{(0)} ⊕ x_i^{(1)}$$ (3)

Thus knowing $x_i^{(1)}$ for $i = 0, 1, 2, \ldots, 15$ is sufficient to recover the full encryption key under the assumption that the plaintext is known. This is the weakest part of Round 1; in fact, the weakest part of the Cipher.

**B. How Lookup Tables fit in Memory and Cache**

Having discussed the issue of memory access in AES implementation, we now turn our focus on how AES Lookup tables fit in Memory and in Cache. We know, each Lookup table has 256 4-Byte entry. That gives it a total size of 256 * 4 = 1024 Bytes. Now, typical size of memory block is 64 Bytes even on modern microprocessors. Hence, for simplicity, we will assume the memory block size B to be 64 Bytes throughout this paper. Therefore, each table $T_i$ will occupy 16 memory blocks. Also assume that, the initial address of $T_i$ congruent to the cache. That is, 1st block of the table will be cached in one of the W cache lines in the 1st cache set.

**III. THE IDEALIZED PRIME+PROBE TECHNIQUE**

The Prime+Probe technique helps to visualize the vulnerability of AES Cipher during the 1st Round. To keep things simple, we consider an idealized environment in which only the attacking process and the victim encrypting process exists. Also assume that the attacking process will be able to invoke or interrupt the encrypting process at any time. Now since we are talking about triggered encryption, the plaintext is also known. The Prime+Probe technique has the following 3 phases: Prime, Trigger and Probe.
An Efficient Indexing Technique for AES Lookup Table to Prevent Side-Channel Cache Timing Attack

Prime: During the Prime phase, the cache is filled completely by the attacking process. This can be done by allocating a contiguous amount of memory as big as the size of the cache and performing a read on it. If \( S = 32 \), \( W = 8 \) and \( B = 64 \) Bytes, then

\[
\text{Total size of the cache} = S \cdot W \cdot B = 16 \quad (4)
\]

As implied by Fig. 3, performing a serial read on a contiguous memory as big as the cache results in inception of filling the cache. Continuing this way, we will end up filling the cache completely after the reading is finished.

Trigger: With the Cache filled with our data, we are in a position to invoke the encrypting process and trigger an encryption of some known plain text. As the encrypting process reaches the First Round and performs the first lookup, it is interrupted. It is worth noting here that we are idealizing the model to illustrate the susceptibility of the First Round, and recovering the AES Encryption Key is not the actual intention of the contribution presented in this paper.

Probe: After we interrupt the encrypting process just after it performed its first lookup during the First Round of AES, we perform the Prime again; that is, we fill the cache again. However, in this time, we maintain a clock to take the measurement of how long it takes to read a memory block. Before the Trigger phase, the cache was completely filled by our data. But after that phase, one block of the Lookup table \( T_0 \) is now in the cache due to the lookup operation. Therefore, all but that block will suffer from a cache miss.

Observing the position of the cache miss, we can easily identify which block of \( T_0 \) was accessed since we assumed Lookup Tables \( T_i \) to be congruent with the cache. Now, if we can identify which block of \( T_0 \) was accessed during the First Round, we can extract some valuable piece of information. Since size of memory block is 64 Bytes, each memory block holds 16 entry of the Lookup Table. So for example, if 7th block of \( T_0 \) was accessed, we can conclude the index \( x_0^{(1)} \) was somewhat between 0x60 and 0x6F; that is, high nibble is 0x6. We know -

\[
x_0^{(1)} = x_0^{(0)} \oplus k_0^{(0)}
\]

Here, \( x_0^{(0)} \) is the 1st Byte of the plaintext and \( k_0^{(0)} \) is the 1st Byte of the Encryption Key. Now, since addition and subtraction in Galois Field are the same XOR operation, rearranging the equation yields -

\[
k_0^{(0)} = x_0^{(0)} \oplus x_0^{(1)}
\]

Because we know \( x_0^{(0)} \) from the plain text and the high nibble of \( x_0^{(1)} \) from the Prime+Probe technique, we can compute the high nibble of the first Byte of the Encryption Key by a simple XOR operation. Proceeding this way for every Bytes \( x_i^{(1)} \) for \( i = 0, 1, 2, \ldots, 15 \), we can calculate \( k_0^{(0)} \) for \( i = 0, 1, 2, \ldots, 15 \) and dig up half of the Encryption Key. The idealized procedure is summarized in Algorithm 1.
An Efficient Indexing Technique for AES Lookup Table to Prevent Side-Channel Cache Timing Attack

IV. PROPOSED RANDOM ADDRESS TRANSLATOR (RAT)

From the analysis in the previous sections, it is prominent that the existence of linear correlation between index Bytes of the state matrix and the location of corresponding accessed block in memory is the primary reason of vulnerability of AES in the first round. Given size of memory block (64 Bytes in most of the cases), any information about the memory access during the encryption process would directly reveal high nibble of the index Bytes. Thus if the size of the key is 128 Bit, pre-knowledge of the high nibbles of the index Bytes would reduce the brute force complexity to $2^{64}$. A machine that can perform 1 quadrillion operations per seconds would take less than 5 seconds to break the rest of the key in brute force attempt.

Now, if the correlation can be broken, so that the index Byte will no longer be used directly to locate the desired block of the Lookup table, most of the cache-timing threats will be eliminated. The key concept to the solution is to perform memory access to obtain the desired content but from an arbitrary, unpredictable location. This can be accomplished by introducing a Translator table between the state matrix and the lookup table. The index Bytes of the state matrix will then be used as index to the Translator table to get the location of the desired block.

Example Scenario,
$S=32; W=8; B=64$ Bytes
Total Cache Size = $32 \times 8 \times 64$ Byte = 16 KB

Assumptions:
1. Starting Address of $T_{i}$ is known.
2. Cache Line Size = 64 Bytes
The Translator table is initially aligned with the Lookup table as illustrated in Fig. 7. This implies that if the address of \(i\)-th row in the lookup table is \(X_i\), the content of the \(i\)-th row of the Translator table is also \(X_i\). Initially, it has no effect on the accessed memory block of the lookup table because it is fully aligned with it. After that, let’s take two rows of the Translator table \(T_i\) and \(T_j\) and corresponding two rows of the original lookup table \(L_i\) and \(L_j\) and perform a swap between them. This is illustrated in Fig. 8.

Continuing this way, there can be at most 128 swaps between the entries of the table. Note that an entry can be swapped only once, so keeping trace of swapped entries is important. After scrambling, the correlation is completely broken as illustrated in Fig. 9.

With the RAT activated, there is no such phrase such as, index 0x05 will access the 1st block of the Lookup table. Because the content of the address 0x05 is now at 0x7B address (Fig. 9) and that tracing is recorded by the RAT. The procedure for constructing the RAT is formulated in Algorithm 2 and 3.

After that, a scrambling is done on both tables simultaneously and arbitrarily. Note that no entry is allowed to be scrambled or swapped with other entry more than once. Also note the presence of \(\text{rand}()\) methods. It symbolizes the randomness of the swapping process.

---

**Algorithm 1** Abstract procedure for recovering half of the Key during the First Round

1. TriggerTheEncryption();
2. for \(i = 0\) to 15 do
3. FillCache();
4. HaltEncryption();
5. \(j = \text{MissedBlockIndex}();\)
6. \(j = j \ll 4;\)
7. \(k_i^{(0)} = j \oplus x_i^{(0)};\)
8. end for

**Algorithm 2** Initializing the RAT

1. for \(i \leftarrow 0x00\) to 0xFF do
2. \(\text{RAT}[i] \leftarrow i;\)
3. \(\text{Swapped[i]} \leftarrow \text{false};\)
4. end for
An Efficient Indexing Technique for AES Lookup Table to Prevent Side-Channel Cache Timing Attack

Fig. 7. Random Address Translator Initialization.

Fig. 8. Random Address Translator after First Swap.
Fig. 9. Random Address Translator Completely Scrambled. It can be noticed that some of the Entries can be left “unswapped” as the Process Is Arbitrary. Here 2nd Entry Is Left Intact.

V. PERFORMANCE EVALUATION

There are some other proposed solutions in this field [16, 17, 18, 19, 20, 21, 22]. We take some of the most prominent ones and compare it with our proposed solution.

Avoiding Memory Access. Since the cache-timing attacks exploit the effect of memory access on the cache, any implementation that does not perform any table lookup will not suffer from this sort of attacks. Nevertheless, this solution has a major drawback. That is the performance is degraded by an order of magnitude. The Sub-Bytes Transformation phase requires 16 table lookups for the 16 Bytes of a State matrix. The ShiftRows Transformation phase requires 16 shift operations. The MixColumns Transformation phase requires 8 shift operations, 8 conditional XOR operations and 12 mandatory XOR operations. At the end, the AddRoundKey Transformation phase requires 16 XOR operations. To summarize the fact, we need a total of 24 shift operations, 16 table lookups, 8 conditional XOR operations, 28 mandatory XOR operations for a single round, as demonstrated in Fig.10.

Algorithm 3 Constructing the RAT

1: for \(i \leftarrow 0x00 \) to \(0x7F \) do
2: \( m \leftarrow \text{rand()} \text{mod} 0xFF; \)
3: \( \text{while} \ Swapped[m] = \text{true} \) do
4: \( m \leftarrow \text{rand()} \text{mod} 0xFF; \)
5: end while
6: \( n \leftarrow \text{rand()} \text{mod} 0xFF; \)
7: \( \text{while} \ Swapped[n] = \text{true} \) do
8: \( n \leftarrow \text{rand()} \text{mod} 0xFF; \)
9: end while
10: swap(RAT[m],RAT[n]);
11: for \(j \leftarrow 0 \) to \(3 \) do
12: swap(\(T_j[m],T_j[n]\));
13: swap(\(T_j^{(10)}[m],T_j^{(10)}[n]\));
14: end for
15: end for
An Efficient Indexing Technique for AES Lookup Table to Prevent Side-Channel Cache Timing Attack

Fig. 10. Comparison of RAT with Straightforward AES Implementations in a Single Round. An Implementation that Completely Avoids Memory Access Requires 16 table Lookups, 24 Shift Operations and 36 XOR Operations (8 of which them are Conditional), Whereas Introduction of RAT Incurs 2 times more table Lookups and 16 XOR Operations in a Single Round.

Fig. 11. Performance Evaluation of RAT in terms of Memory Usage.

In fact, the conditional XOR operation is also vulnerable to timing attack [23]. Other than lagging behind in terms of security, a straightforward implementation requires more computation time. So avoiding memory access does not solve the problem. Having said that, the positive side is that the straightforward implementation does not need to maintain any sort of lookup table other than the S-Box. On the other hand, table lookup with RAT has an overhead of maintaining two tables. The Lookup table requires 4096 Bytes of memory space and RAT requires another 1024 Bytes of space.

Disable Cache Sharing. To protect against software-based attacks, it would suffice to prevent cache state effects from spanning process boundaries. However, this is very expensive to achieve in practical settings. On a single threaded processor, it would require flushing all caches during every context switch. On a processor with simultaneous multi-threading, it would also require the logical processors to use separate logical caches, statically allocated within the physical cache; some modern processors do not support such a mode [16]. Using RAT will not impose any restriction about cache sharing which is perfectly suitable for cloud environment and virtual machines (without causing evictions and filling).
Static or Disabled Cache. One brutal countermeasure against the cache-base attacks is to completely disable the CPU’s caching mechanism. Of course, the effect on performance would be devastating. A more attractive alternative is to activate a “no-fill” mode where the memory accesses are serviced from the cache when they hit it, but accesses that miss the cache are serviced directly from the memory:

1. Preload the AES tables into the cache
2. Activate the “no-fill” mode
3. Perform encryption
4. Deactivating “no-fill” mode

The section spanning (1) and (2) is critical, and attacking process must not be allowed to run during this time [16]. However, the major drawback of this approach is that during the encryption process the overall system will be slowed down by a factor since cache access is limited during the encryption process. However, using RAT, the “no-fill” mode is not required. All the other processes can be allowed to access the cache even during the encryption proceeds.

Dynamic Table Storage. The cache-based attacks observe memory access patterns to learn about the table lookups. Instead of eliminating these, we may try to de-correlate them. For example, one can use many copies of each table, placed at various offsets in memory, and have each table lookup use a pseudo randomly chosen table. By utilizing more compactly, one can use a single table, but pseudo randomly move it around memory several times during each encryption [16]. But the major drawback of this approach is that it will incur cache misses more than ever.

This implies that the encryption process will be slowed down by a factor. More importantly, the performance and security of this approach are largely architecture dependent. Using RAT helps us alleviate this situation by obviating the need of moving around the Lookup table in the memory.

A. Other Major Advantages

Some important benefits of using the RAT are as follows:

- The Cipher Text remains the same and the AES implementation becomes a client dependent process. We do not need to worry about how other ends of the terminals are encrypting their data. As long as the key does not change, Cipher Text remains the same.
- No need of NOP operation. NOP operations are introduced to scramble uniform patterns in various fields. In case of cache-timing analysis, NOP operation would mean accessing Lookup table entry, but doing no operation with it.
- No need to introduce disturbance in plaintext. Sometimes plaintext is modified before it is passed to the encrypting process. This sort of approach is not suitable for large scale application where the interface must be kept as simple as possible.

VI. Future Works

The findings we have presented in this paper have thrown up few new questions that need further investigation. In the remainder of this section, we discuss future work to extend the scope and applicability of this work.

- There are some limitations in the process of constructing the RAT table. One major limitation is that swapping is not allowed more than once for an entry. Thus, there can be a maximum of 128
swaps for a single Lookup table. To allow more number of swaps per entry, more sophisticated data structure is required.

- The big picture that summarizes the performance difference with the straightforward Table lookup approach is given below: As we can see, we need 2 times more table lookups than the ordinary approach. The area of future development is to reduce this particular factor.
- Another big concern is that RAT tables add a fair amount of memory overhead to the encrypting process. Whereas the Lookup tables altogether require 4KB of space, RAT tables requires 1KB of space. Although this is not a big amount, space overhead can be reduced with the assistance of some more tables of smaller sizes.

VII. CONCLUSIONS

It is apparent that Side channel attacks do not divulge everything at once; instead they ease the calculations and assumptions that helps rest of the operations accomplished in a bounded time. Although the existing AES encryption model has sufficient complexity to blow away any straightforward attempt to break the security, side-channels still poses a real threat and a bad implementation of the cipher might lead to dire consequences. The most challenging part of security is now making the implementation of the existing ciphers resistant to side-channel attacks. Preventing leakage of information through side-channel is not a trivial task at all. The empirical observation presented in this paper has shown that even the straightforward Table lookup approach can be made secure without the assistance of any piece of special software or hardware, which is widely acknowledged requirement in the real settings.

REFERENCES


Authors’ Profiles

Refazul Islam Refat received Bachelor of Science (BSc) in Computer Science and Engineering from the University of Dhaka, Bangladesh in 2015. His research interest includes Information Security, Internet of Things and Crypto Currency.
An Efficient Indexing Technique for AES Lookup Table to Prevent Side-Channel Cache Timing Attack

Euna Islam received Bachelor of Science (BSc) in Computer Science and Engineering from the University of Dhaka, Bangladesh in 2015. Her research interest includes Information Security and Software Engineering.

Md. Mosaddek Khan completed his PhD in Computer Science from the University of Southampton, UK in 2018. He also received MS and BSc in Computer Science and Engineering from the University of Dhaka, Bangladesh in 2012 and 2010, respectively. His research interest includes Artificial Intelligence, Multi-Agent Systems and Information Security.