Noname manuscript No.
(will be inserted by the editor)

Recovering OpenSSL ECDSA Nonces Using the FLUSH+RELOAD Cache Side-channel Attack

Yuval Yarom · Naomi Benger

the date of receipt and acceptance should be inserted later

Abstract We illustrate a vulnerability introduced to elliptic curve cryptographic protocols when implemented using a function of the OpenSSL cryptographic library. For the given implementation using an elliptic curve E over a binary field with a point $G \in E$, our attack can recover the majority of the bits of a scalar k when kG is computed using the OpenSSL implementation of the Montgomery ladder. For the Elliptic Curve Digital Signature Algorithm (ECDSA) the scalar k is intended to remain secret. Our attack recovers the scalar k and thus the secret key of the signer and would therefore allow unlimited forgeries. This is possible from snooping on only one signing process and requires computation of less than one second on a quad core desktop when the scalar k (and secret key) is around 571 bits.

1 Introduction

Elliptic curve cryptography (ECC) [23, 25] includes a number of public-key cryptographic protocols whose security relies on the computational intractability of the Elliptic Curve Discrete Logarithm Problem (ECDLP): Given an elliptic curve over a finite field and two points G and H on the curve, find the scalar k such that $H = kG$.

ECC offers a higher encryption strength per key-bit than related methods whose security is reliant on the hardness of computing discrete logarithms in a finite field or factoring the product of large primes. Consequently, ECC uses significantly shorter keys and offers faster operations than other methods, contributing to its rising popularity.

The Elliptic Curve Digital Signature Algorithm (ECDSA) [6, 21, 27] is a standard digital signature algorithm implemented using elliptic curves. One core operation of the ECDSA algorithm, as in many ECC protocols, is the scalar multiplication of a point on the elliptic curve by a pseudo-randomly generated secret nonce. The confidentiality of the nonce is paramount for the security of the algorithm. Past research indicates that partial exposure of nonce bits can be exploited for efficient attacks on the secret key [9, 28].

OpenSSL [30] is a cryptographic software package that implements ECDSA. When using elliptic curves over a binary field $\mathbb{F}_{2^m}$, OpenSSL uses the Montgomery ladder [22, 26] algorithm to compute kG, the scalar multiplication of a publically known point G by the secret nonce k. One of the advantages of the Montgomery ladder is that it has a regular behaviour, performing the same sequence of operations for each nonce bit, irrespective of the value of the bit. This regular behaviour makes it more resilient to side-channel attacks [22, 29].

While the operations performed by the algorithm are regular, their targets depend on the value of the bits of the nonce. To apply the operations to the respective targets, the OpenSSL implementation uses a conditional branch based on the value of the bit. By tracing this branch an attacker can recover the values of the nonce bits and, consequently, break the cryptosystem. In this paper we present our use of the FLUSH+RELOAD cache side-channel attack [39] to trace the branch in the OpenSSL implementation.

The FLUSH+RELOAD attack exploits a security weakness in the IA-32 and X86-64 architectures that allows processes to monitor other processes read and execute access to shared memory pages. Our spy program monitors access to both arms of the conditional branch and uses the information collected from these probes to reconstruct the nonce. This attack is a threat to the security of any cryptographic protocol implemented using the OpenSSL scalar multiplication method when the scalar is intended to remain secret.

In this paper we illustrate the efficiency of the attack by analysing ECDSA and recovering the secret key using only

Yuval Yarom · Naomi Benger
The University of Adelaide, E-mail: yval@cs.adelaide.edu.au, naomi.benger@adelaide.edu.au

one signature at very little computational cost (in both time and memory). This attack is applicable when the malicious party has access to the memory of the targeted device. This is a reasonable assumption as could be the case when using, for example, a multi-user operating system, co-hosted virtual machines in a cloud computing environment or a computer victim to malware.

The paper also presents new information on the limitation of the FLUSH+RELOAD attack. We discuss spatial limitations, affecting the distance between multiple probes, and temporal limitations, affecting the probe resolution. The results of this paper support the findings of Walter [37] that longer keys render a cryptographic algorithm more vulnerable to side-channel analysis.

The rest of this paper is organised as follows: in the following subsection we discuss related research. The next section presents background information on ECDSA, the Montgomery ladder and the FLUSH+RELOAD attack. Section 3 describes our attack on the OpenSSL implementation of ECDSA. The results of the attack are analysed in Section 4. We discuss the implications of the attack and suggest techniques for mitigation in Section 5.

1.1 Related Work

There have been a number of publications addressing the security issues of digital signatures when partial information is leaked [15, 18, 28].

Gopalakrishnan et al. [15] presents algorithms for solving the ECDLP using the additional information of some consecutive bits of the private key. These algorithms outperform the currently best known methods of solving the ECDLP without the extra information or using exhaustive search on the remaining key space. In this work we do not focus on the ECDLP. Instead, we use leaked information about the nonces.

The attacks in Howgrave-Graham and Smart [18] and in Nguyen and Shparlinski [28] rely on having obtained a relatively small number of bit of the nonces used for many signatures and then using the LLL method [24] for solving the related hidden number problem to find the secret key. The attack of Nguyen and Shparlinski [28], for example, given a group of order around 160 bits the probabilistic algorithm would obtain the secret key using 23 signatures (assuming independent and uniformly at random selected messages) in polynomial time, using only seven consecutive least significant leaked bits of each nonce. (Relying on some reasonable assumptions.) Each of these assumes only a small fraction of k is recovered. The main contribution of this work is to illustrate a method, to recover a large majority of the bits, using only one signature. From these, the full nonce is obtained using less than one second of additional computation time. Once the nonce has been fully determined the secret key can be obtained. Though the goal and approach of the works are similar, the methods are very different.

The attack of Brumley and Tuveri [9] uses the above methods to highlight a specific vulnerability in earlier versions of OpenSSL's Montgomery ladder implementation for curves over binary fields. Though the attacks differ, they both illustrate that the OpenSSL implementation of the Montgomery ladder is vulnerable to both remote attacks and attacks launched from virtual machines with access to the memory of the target computer. The countermeasure suggested in Brumley and Tuveri [9] will not thwart our attack.

Cache side-channel attacks have been used against cryptosystems [1–3, 10, 11, 34, 40]. These attacks use the PRIME+PROBE technique [34] to target the L1 cache level. Consequently, the spy program and the victim must execute on the same execution core of the processor. This is in contrast to our attack, which targets the last-level-cache, and can, therefore, be mounted between different cores.

The FLUSH+RELOAD attack has been used by Gullasch et al. [17] and by Yarom and Falkner [39]. Gullasch et al. [17] uses this attack as a replacement to the PRIME+PROBE attack, relying on a scheduler bug to interrupt the victim and gain control of the core it executes on. Yarom and Falkner [39] uses FLUSH+RELOAD to attack the square-and-multiply exponentiation in the GnuPG implementation of RSA. Unlike the Montgomery ladder, square-and-multiply is known to be vulnerable to side-channel attacks.

2 Preliminaries

In this section we present the relevant general background information about the attack and also the specific information required to understand the context of the example attack.

2.1 ECDSA

The ElGamal Signature Scheme [14] is the basis of the US 1994 NIST standard, Digital Signature Algorithm (DSA). The ECDSA is the adaptation of one step of the algorithm from the multiplicative group of a finite field to the group of points on an elliptic curve. The main benefit of using this group as opposed to the multiplicative group of a finite field is that smaller parameters can be used to achieve the same security level [23, 25] due to the fact that the current best known algorithms to solve the discrete logarithm problem in the finite field are sub-exponential and those used to solve the ECDLP are exponential. See Balasubramanian and Koblitz [7], Adleman and Demarrais [4] and developments thereof for more details.

Parameters: An elliptic curve E defined over a finite field $\mathbb{F}_q$; a point $G \in E$ of a large prime order n (generator of the group of points of order n). Parameters chosen as such are generally believed to offer a security level of $\sqrt{n}$ given current knowledge and technologies. Parameters are recommended to be generated following the Digital Signature Standard [27]. The field size q is usually taken to be a large odd prime or a power of 2. The implementation of OpenSSL uses both prime fields and $q = 2^m$; the results in this paper relate to the binary field case.

Public-Private Key pairs: The private key is an integer d, $1 < d < n-1$ and the public key is the point $Q = dG$. Calculating the private key from the public key requires solving the ECDLP, which is known to be hard in practice for the correctly chosen parameters. The most efficient currently known algorithms for solving the ECDLP have a square root run time in the size of the group [13, 38], hence the aforementioned security level.

Suppose Bob, with private-public key pair $\{d_B, Q_B\}$, wishes to send a signed message m to Alice, he follows the following steps:

1. Using an approved hash algorithm, compute $e = Hash(m)$, take $\bar{e}$ to be the leftmost ℓ bits of e (where $\ell = \min(\log_2(q),$ bitlength of the hash)).
2. Randomly select $k \leftarrow_R \mathbb{Z}_n$.
3. Compute the point $(x,y) = kG \in E$.
4. Take $r = x \bmod n$; if $r = 0$ then return to step 2.
5. Compute $s = k^{-1}(\bar{e} + rd_B) \bmod n$; if $s = 0$ then return to step 2.
6. Bob sends (m,r,s) to Alice.

The message m is not necessarily encrypted, the contents may not be secret, but a valid signature gives Alice strong evidence that the message was indeed sent by Bob. She verifies that the message came from Bob by

1. checking that all received parameters are correct, that $r,s \in \mathbb{Z}_n$ and that Bob's public key is valid, that is $Q_B \neq \mathcal{O}$ and $Q_B \in E$ is of order n.
2. Using the same hash function and method as above, compute $\bar{e}$.
3. Compute $\bar{s} = s^{-1} \bmod n$.
4. Find the point $(x,y) = \bar{e}\bar{s}G + r\bar{s}Q_B$.
5. Verify that $r = x \bmod n$ otherwise reject the signature.

Step 2 of the signing algorithm is of vital importance, inappropriate reuse of the random integer led to the highly publicised breaking of Sony PS3 implementation of ECDSA. Knowledge of the random value k, a.k.a. the *ephemeral key* or the *nonce*, leads to knowledge of the secret key. All values (m,r,s) can be observed by an eavesdropper, $\bar{e}$ can be found from m, $r^{-1} \bmod n$ can be easily computed from n and r, and if k is discovered then an adversary can find Bob's secret key through the simple calculation

$$d_B = (sk - \bar{e})r^{-1}.$$

Our attack targets Step 3 of the OpenSSL implementation of ECDSA.

2.2 The Montgomery Ladder

Scalar multiplication is a common operation in cryptography and in a number of incidences (such as step 3 of ECDSA) the scalar is intended to remain secret. This scalar multiplication is most efficiently performed using a double-and-add method (or the related right-to-left method) as outlined in Algorithm 1.

Input: Point P, scalar n, k bits
Output: Point nP
$Q \leftarrow \mathcal{O}$
for *i from k to* 0 **do**
  $Q \leftarrow 2Q$
  if $n_i = 0$ **then**
    $Q \leftarrow Q + P$
  end
end

Algorithm 1: Double-and-add point scalar multiplication

Double-and-add methods, though efficient, are vulnerable to side-channel attacks. The addition law for points on commonly used elliptic curves is not complete. That is, the computation of $P+Q$ differs between the cases $P = Q$ and $P \neq Q$. Consequently, it is possible to distinguish when the `if` in the loop is executed and hence when a bit of n_i is 0.

The Montgomery ladder, described in Montgomery [26], is presented in Algorithm 2. It differs from Algorithm 1 in that both a doubling and an addition of points occur at each step, regardless of the value of the bit. Thus, the Montgomery ladder thwarts side channel attacks which measure the computation at each bit to determine if an addition operation was executed. The branching in Algorithm 2 controls which point is doubled and where the addition of points is stored.

Input: Point P, scalar n, k bits
Output: Point nP
$R_0 \leftarrow \mathcal{O}$
$R_1 \leftarrow P$
for *i from k to* 0 **do**
  if $n_i = 0$ **then**
    $R_1 \leftarrow R_0 + R_1$
    $R_0 \leftarrow 2R_0$
  else
    $R_0 \leftarrow R_0 + R_1$
    $R_1 \leftarrow 2R_1$
  end
end

Algorithm 2: Montgomery ladder point scalar multiplication

Instead of distinguishing additions from doublings, our attack identifies which arm of the `if` statement is taken. The technique we use is described in the next section.

2.3 The FLUSH+RELOAD attack

FLUSH+RELOAD is a recently developed cache side-channel attack [39]. The attack exploits a weakness in the IA-32 and X86-64 processor architectures, which allows processes to manipulate the cache of other processes.

Using the attack, a spy program can trace or monitor memory read and execute access of a victim program to shared memory pages. The spy program only requires read access to the shared memory pages, hence pages containing binary code in executable files and in shared libraries are susceptible to the attack. By monitoring the victim access to specific locations in these pages, the spy program learns when the victim executes the code in the monitored memory locations. From this information the spy program can infer information on the data processed by the victim.

The spy program described in Yarom and Falkner [39] uses the FLUSH+RELOAD attack to retrieve the secret key from the GnuPG RSA decryption. The spy program monitors the phases of the square-and-multiply exponentiation [16] used by GnuPG. As these phases depend on the values of the bits of the exponent, monitoring them allows the spy program to recover the secret exponent.

The attack operates by dividing time into slots. At the beginning of a time slot, the spy program flushes the monitored memory line from the cache of the processor. At the end of the slot, the spy program loads data from the memory line. Loading data from cached memory lines is significantly faster than loading them from memory. Hence, by measuring the time it takes to load the data, the spy program can know whether the line is cached or not. As the line is flushed at the beginning of the slot, having it cached at the end indicates that the processor accessed the line during the time slot.

When the victim memory access overlaps the spy measurement, the spy will miss the access [39]. Consequently, increasing the time slot length reduces the portion of time the spy spends in measurement and with it the probability of missing access. On the other hand, the spy is unable to distinguish between multiple accesses to the same memory line in a single time slot. It also cannot determine the order of memory accesses to different memory lines occurring in the same time slot. Consequently, increasing the time slot reduces the attack's resolution. Hence choosing the length of the time slot presents a tradeoff between the attack resolution and the probability of missing a memory access.

3 Attacking OpenSSL ECDSA

OpenSSL is one of the most commonly used open-source cryptographic libraries. It provides a set of cryptographic services, including both public key and symmetric encryption algorithms, and public key signature algorithms.

OpenSSL's implementation of ECDSA uses the Montgomery ladder algorithm for scalar multiplication on the elliptic curve. We use this implementation to demonstrate that naïve implementations of the Montgomery ladder are susceptible to the FLUSH+RELOAD attack.

Listing 1 shows the relevant section of the implementation of the Montgomery ladder in OpenSSL version 1.0.1e. The bits of the multiplication scalar are stored in the word array `scalar->d`, where the word size is defined by the architecture, e.g. 32 bits for the IA-32 architecture and 64 bits for the X86-64 architecture. The outer loop, at lines 268 to 286 traverses over the words representing the scalar. The inner loop, at lines 271 to 284 traverses the bits in each word. Line 273 tests the bit. For each bit the implementation executes a group add followed by a group double. If the bit is set, the implementation uses lines 275 and 276. For clear bits it uses lines 280 and 281.

Listing 1 OpenSSL implementation of the Montgomery ladder

```
268 for (; i >= 0; i--)
269 {
270     word = scalar->d[i];
271     while (mask)
272     {
273         if (word & mask)
274         {
275             if (!gf2m_Madd(group, &point->X, x1, z1,
                        x2, z2, ctx)) goto err;
276             if (!gf2m_Mdouble(group, x2, z2, ctx))
                        goto err;
277         }
278         else
279         {
280             if (!gf2m_Madd(group, &point->X, x2, z2,
                        x1, z1, ctx)) goto err;
281             if (!gf2m_Mdouble(group, x1, z1, ctx))
                        goto err;
282         }
283         mask >>= 1;
284     }
285     mask = BN_TBIT;
286 }
```

As the listing demonstrates, the implementation is regular: For each bit, the implementation executes exactly the same sequence of operations. The only differences between set and clear bit are the lines that invoke these operations and the targets of these operations. While this is a small difference, it is sufficient for mounting an attack that recovers the values of the bits.

Our spy program uses the FLUSH+RELOAD technique to monitor the execution of the `if` statement in line 273. We distinguish between executing the `then` and the `else` blocks

of the `if` statement. This information reveals the value of the bit tested by the `if` statement.

FLUSH+RELOAD monitors execution by placing probes on shared memory lines. For the attack to recover the bit values, it must distinguish between memory lines access sequences resulting from set bits and those resulting from clear bits. Achieving this depends on several factors: the mapping of source code to memory lines, the sequence of accesses to these memory lines when executing the code and FLUSH+RELOAD's ability to accurately capture the sequences.

The mapping of source lines to cache lines in our build of OpenSSL is depicted in Fig. 1. The machine code created from source lines 273 to 282 covers the virtual memory address range 0x0812130C to 0x081213e8. This range spans four cache lines, marked *A*, *B*, *C* and *D*.

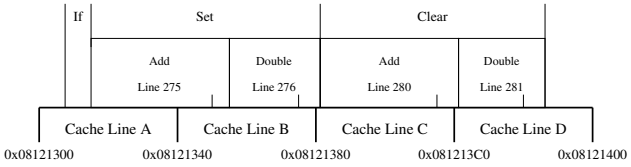

Fig. 1 Mapping from source code to memory

The minimum sequence of memory line accesses required for executing this code can now be constructed. The `if` statement at line 273 is executed for each bit. The code of this statement is in memory line *A*, hence this memory line is accessed when processing of a bit starts. For a set bit, the processing continues with source line 275, which maps to memory lines *A* and *B*. The actual call to the group add function occurs at address 0x08121347. (See mark in Fig. 1.) After a delay for computing the group add, execution continues in memory line *B* to process the return value and to invoke the group doubling function. The group doubling function returns to memory line *B* and execution leaves the `if` body at memory line *D*.

Hence, the sequence of memory line accesses required for a set bit is: *A*, *B*, *add*, *B*, *double*, *B*, *D*. Similarly, for a clear bit, the sequence is: *A*, *C*, *add*, *C*, *D*, *double*, *D*.

Due to the limited temporal resolution of FLUSH+RELOAD, the attack can observe the order of memory accesses only if they are sufficiently separated in time. Hence, in the case of OpenSSL, the attack can only observe the order of memory accesses if they are separated by a call to a group operation. For example, when the bit is set, the attack cannot decide whether the access to memory line *A* precedes or follows the access to memory line *B*. Similarly, when observed by FLUSH+RELOAD, memory accesses issued after the group double are merged with those issued at the start of processing the following bit. Figure 2 shows the observable memory accesses when processing a set bit followed by a clear bit.

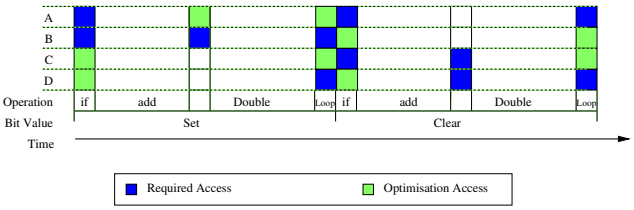

Fig. 2 Observable memory access over time (processing a set then a clear bit)

The diagram also shows memory accesses issued by processor optimisations. These optimisations pre-load memory lines into the cache to reduce the time the program waits for these lines. For example, when the processor uses speculative execution [35], it follows both arms of a conditional branch before evaluating the condition. When the condition is evaluated, the processor commits to the pre-processed computation of the correct arm, disposing of the computation done for the other arm. In the case of OpenSSL this means that even before evaluating the bit, the processor may start processing both line 275 and line 280, triggering memory loads from memory lines *A*, *B* and *C*.

Another optimisation that can cause additional memory line access is spatial prefetching [19]. The processor pairs adjacent memory lines and tries to bring both memory lines into the cache when there is a miss on one of the pair's lines. For example, when there is a cache miss on memory line *A*, the spatial prefetcher may attempt to prefetch memory line *B* and vice versa.

Consequently, as demonstrated in Fig. 2, the memory lines accessed between computing the group add and the group double can be used for recovering the value of the bit. Probing any of lines *A* and *B* gives a positive indication of set bits. Probing either line *C* or *D* gives a positive indication of clear bits. For our attack we probe memory lines *B* and *D*.

Three limitations of the FLUSH+RELOAD attack affect its ability to capture the sequence of memory accesses. The first is the attack temporal resolution which affects its ability to determine the order of accesses performed within a short time from each other. The second limitation is the possibility of an overlap between the memory access and the probe which may result in the attack missing the access. The third limitation is the result of the interaction between the FLUSH+RELOAD attack and the processor optimisation of cache use. In particular, the spatial prefetching optimisation implies that the attack cannot be used to probe two cache lines that form a pair, because probing one of the lines in a pair triggers a prefetch of the other.

For OpenSSL, the attack resolution should be sufficiently high for the attack to be able to distinguish between memory accesses done before and after each bit and those done between the group add and group double operations of each

bit. This can be achieved by setting the time slot size to be less than the time it takes the victim to calculate the group double. As group double calculations are faster than group add calculations, this ensures that the probed memory lines are flushed when the victim computes the group add to be probed when the victim computes the group double.

The probability of an overlap, like the attack resolution, depends on the length of the time slot. Longer time slots mean that the portion of time during which the spy probes is smaller and, therefore, the probability of an overlap is lower.

As predicted by Walter [37], smaller keys are more resilient to the attack. With smaller keys, group operations are shorter, forcing shorter time slots. The shorter time slots lead to an increased probability of an overlap and with it of missing bits.

Missing memory accesses not only prevents the spy program from recovering the value of bits. It may also result in the spy program losing the bit position in the scalar multiplication process. To protect against this possibility, our attack also probes the first and last memory lines of the `gf2m_Mdouble` function. Probing these lines provides the spy program with additional information on the operation of the victim and facilitates recovering the position of captured scalar bits.

The next section describes the details of our experimentation with the attack and its results.

4 Experimental Setup and Results

To test the attack on OpenSSL we used an HP Elite 8300 running Fedora 18. As the OpenSSL package shipped with Fedora does not support ECC, we used our own build of OpenSSL 1.0.1e. To facilitate the mapping from source lines to memory addresses we built OpenSSL with debugging symbols. In real attack settings, the attacker will need to reverse engineer [12] the OpenSSL library. For the experiment we used the OpenSSL `sect1571r1` curve. (NIST Binary-Curve B-571 [27].)

With the selected curve, group add operations take 23,612 cycles on average. The first group double operation takes 6,552 cycles on average, whereas further group double operations take 11,962 cycles. Based on the discussion in Section 3, we picked a slot length of 10,240 cycles.

Figure 3 shows the results of the probes during 50 time slots. Probes taking less than the threshold of 120 clock cycles indicate a victim access to the probed line. The shaded areas in the diagram indicate the computation of the group double operation, identified by probing for group double start and end. For clarity, we have omitted these probes from the diagram.

For example, in the first time slot, the spy program captured access to memory lines *B* and *D*, as well as an access to the last memory line of the group double. The end of the group double is also the end of processing a bit, hence at time slot 1, the victim finished processing a bit and started the next one. The next captured probe is at time slot 5. In this time slot, the victim accessed memory line *B* and started executing a group double. Access to line *B* between the group add and the group double operations indicates that the bit is set. Processing the next bit starts at time slot 8 and ends in time slot 16. The access to memory line *D* in time slot 13 indicates that the second bit is clear.

In the absence of access to memory lines *B* or *D* in time slot 42, the start of a group double indicates that the spy program missed a value of a bit. However, the fact that a bit was processed at that time slot is not missed, demonstrating the value of probing the start of the group double operation.

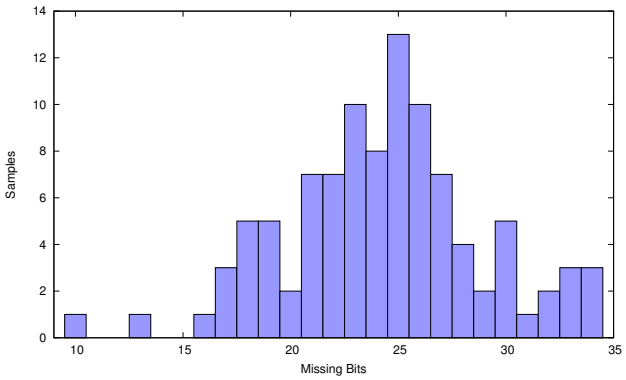

Fig. 4 Missing bits per signature

To measure the number of missing bits we traced the computation of 100 signatures. On average, the attack misses only 4.26% of the bits or 25.28 bits per signature. The distribution of number of bits missing per signature is in Fig. 4. The number of missing bits ranges from 10 to 34, with the median at 25 missing bits per signature.

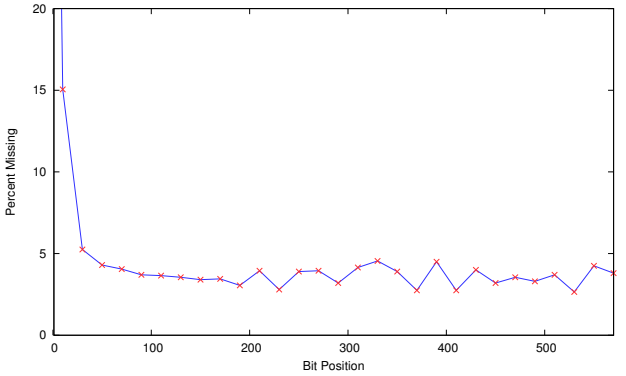

Fig. 5 Missing bits per bit position

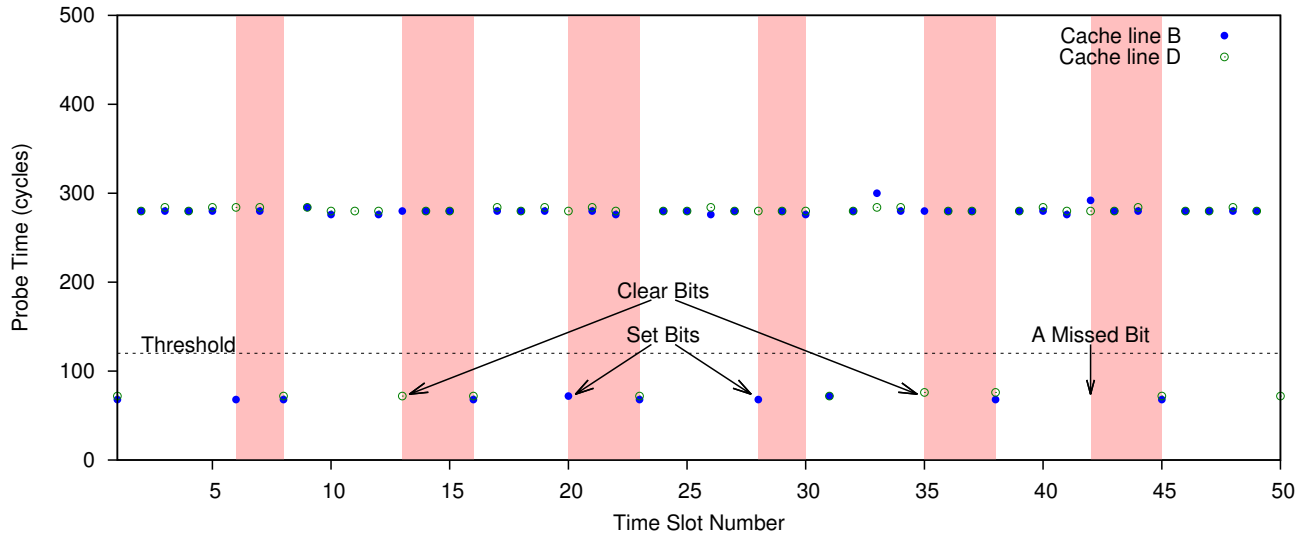

Fig. 3 Probe timing during signing

As Fig. 5 demonstrates, the distribution of missing bits position is not uniform. The first bit (bit 0) is always missed. This is mostly the result of the short time it takes OpenSSL to compute the group double operation for the first bit. Even ignoring the first bit, it is evident that missing bits tend to cluster towards the most significant bits of the scalar. Around 15% of the bits in positions 1 to 20 are missed, compared with 3.6% of the bits from position 50 onward, where the distribution of missing bits is approximately uniform.

5 Discussion

Full Recovery of the Nonce

Given the high proportion of nonce bits recovered by the FLUSH+RELOAD attack, using the LLL based techniques [18, 28] described in Sect. 1.1 seems computationally excessive. With a worst case of 34 bits missing, the baby step giant step algorithm [32] would require less then 10 Megabytes of memory and less than one second of computation to complete the nonce.

Implications

As the ECDLP is not targeted by this attack, the signature protocol is made no more vulnerable by our results. This attack targets the scalar multiplication implementation of OpenSSL and is therefore particular to implementations using this and similar implementations of the Montgomery ladder. The vulnerability introduced by this implementation is due to the bits of the secret nonce determining which conditional branch is taken. Our spy program is able to determine how the algorithm executes by having access to the victim's memory and used this knowledge to reconstruct values used by the software.

As demonstrated in this paper, the FLUSH+RELOAD attack has a higher resolution and better accuracy then previously known attacks. FLUSH+RELOAD applies across multiple cryptographic schemes and we show that it applies to implementation hitherto not considered vulnerable to side-channel attacks. It, therefore, presents new threats to confidentiality of data.

This threat is not limited to cryptographic software. The attack can be applied to other software and may be able to extract sensitive information from other software, including keystroke timing information [31,33], statistical data on network traffic and disk use and business logic.

Mitigation

Preventing data flow from secrets to branch conditions is one way for mitigating the attack. The Networking and Cryptography library (NaCl), implemented by Bernstein, Lange and Schwabe [8], provides an implementation of the Montgomery ladder that is not vulnerable to our attack. Instead of using branches, NaCl uses arithmetic operations to select the arguments and targets of the group operation. Consequently, NaCl's use of the cache is independent of the values of the bits in the nonce. Fixing OpenSSL by using methods similar to those used in NaCl will provide protection against the attack.

While fixing OpenSSL would prevent the attack we describe, it is not a panacea for the FLUSH+RELOAD attack. As discussed above, the attack is generic and can apply to other software. It would be advisable for implementors of cryptographic software to avoid using secret information to

determine which operations or values are accessed when the memory locations can be distinguished by an eavesdropper.

The FLUSH+RELOAD technique exploits the lack of restrictions on the ability to flush specific memory lines from the cache, which enables processes to interact using read-only pages. This is a security weakness of the IA32 and the X86-64 architectures. Addressing this weakness requires a hardware fix. A possible fix is to restrict the ability to flush memory to memory pages to which the process has write access and to memory pages to which the system allows such access. This access control could be implemented by adding memory types that restrict flush access to the PAT (Page Attribute Table) [20, chap. 11]

6 Conclusions and future work

The results of this work imply that the OpenSSL Montgomery ladder implementation should be avoided in all implementations of elliptic curve protocols when a scalar multiplication step involves a secret parameter. This attack is applicable when the malicious party has access to the memory of the targeted device, a completely reasonable assumption, possible when using a multi-user operating system, a virtual machine or a computer victim to malware.

The results of this work also support the theory of Walter [37] that smaller keys are more resilient to side-channel analysis. In this attack a higher proportion of the nonce can be obtained for larger key sizes. This implies that as we, naturally, transition to larger parameters in response to increasing computing capabilities, prevention of side-channel attacks should be incorporated into the implementation design, as is the methodology adopted by the authors of the NaCl cryptographic library [8].

Further work in this line of research is to apply the FLUSH+RELOAD attack outside the realm of scalar multiplication. A possible use is for implementing the Vaudenay padding oracle attack [5, 36]. It would also be interesting to test the extent to which the attack can be applied to non-cryptographic software. The nature of the threat the attack presents to business logic and to customer privacy should be evaluated.

Acknowledgements The authors wish to thank Dr Katrina Falkner for the advice and support.

This research was performed under contract to the Defence Science and Technology Organisation (DSTO) Maritime Division, Australia.

References

1. Acıiçmez, O.: Yet another microarchitectural attack: exploiting I-Cache. In Ning, P., Atluri, V., eds.: Proceedings of the ACM Workshop on Computer Security Architecture, Fairfax, Virginia, United States (November 2007) 11–18
2. Acıiçmez, O., Brumley, B.B., Grabher, P.: New results on instruction cache attacks. In Mangard, S., Standaert, F.X., eds.: Proceedings of the Workshop on Cryptographic Hardware and Embedded Systems, Santa Barbara, California, United States (August 2010) 110–124
3. Acıiçmez, O., Schindler, W.: A vulnerability in RSA implementations due to instruction cache analysis and its demonstration on OpenSSL. In Malkin, T., ed.: Proceedings of the Cryptographers' Track at the RSA Conference, San Francisco, California, United States (April 2008) 256–273
4. Adleman, L.M., Demarrais, J.: A Subexponential Algorithm for Discrete Logarithms over all Finite Fields. Mathematics of Computation **61**(203) (1993) 1–15
5. AlFardan, N.J., Paterson, K.G.: Plaintext-recovery attacks against datagram TLS. In: Proceedings of the 19th Annual Network & Distributed Systems Security Symposium, San Diego, California, United States (February 2012)
6. American National Standards Institute: ANSI X9.62, Public Key Cryptography for the Financial Services Industry: The Elliptic Curve Digital Signature Algorithm. (1999)
7. Balasubramanian, R., Koblitz, N.: The Improbability that an Elliptic Curve has Subexponential Discrete Log Problem under the Menezes - Okamoto - Vanstone Algorithm. Journal of Cryptology **11**(2) (1998) 141–145
8. Bernstein, D.J., Lange, T., Schwabe, P.: The security impact of a new cryptographic library. In: Proceedings of the 2nd international conference on Cryptology and Information Security in Latin America. LATINCRYPT'12, Berlin, Heidelberg, Springer-Verlag (2012) 159–176
9. Brumley, B.B., Tuveri, N.: Remote timing attacks are still practical. In Atluri, V., Diaz, C., eds.: Computer Security - ESORICS 2011. Volume 6879 of Lecture Notes in Computer Science., Springer-Verlag (2011) 355–371
10. Canteaut, A., Lauradoux, C., Seznec, A.: Understanding cache attacks. Technical Report 5881, INRIA (April 2006)
11. Chen, C., Wang, T., Kou, Y., Chen, X., Li, X.: Improvement of trace-driven I-Cache timing attack on the RSA algorithm. The Journal of Systems and Software **86**(1) (2013) 100–107
12. Cipresso, T., Stamp, M.: Software reverse engineering. In Stavroulakis, P., Stamp, M., eds.: Handbook of Information and Communication Security. Springer (2010) 659–696
13. Gallant, R.P., Lambert, R.J., Vanstone, S.A.: Improving the parallelized pollard lambda search on anomalous binary curves. Mathematics of Computation **69**(232) (2000) 1699–1705
14. Gamal, T.E.: A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory **31**(4) (1985) 469–472
15. Gopalakrishnan, K., Thériault, N., Yao, C.Z.: Solving discrete logarithms from partial knowledge of the key. In Srinathan, K., Pandu Rangan, C., Yung, M., eds.: Progress in Cryptology – INDOCRYPT 2007. Volume 4859 of Lecture Notes in Computer Science., Springer (2007) 224–237
16. Gordon, D.M.: A survey of fast exponentiation methods. Journal of Algorithms **27**(1) (April 1998) 129–146
17. Gullasch, D., Bangerter, E., Krenn, S.: Cache games — bringing access-based cache attacks on AES to practice. In: Proceedings of the IEEE Symposium on Security and Privacy, Oakland, California, United States (May 2011) 490–595
18. Howgrave-Graham, N., Smart, N.P.: Lattice attacks on digital signature schemes. Designs, Codes and Cryptography **23**(3) (2001) 283–290
19. Intel Corporation: Intel 64 and IA-32 Architecture Optimization Reference Manual. (April 2012)
20. Intel Corporation: Intel 64 and IA-32 Architectures Software Developer's Manual Volume 3A: System Programming Guide, Part 1. (March 2013)

21. Johnson, D., Menezes, A., Vanstone, S.: The elliptic curve digital signature algorithm (ECDSA). International Journal of Information Security **1**(1) (August 2001) 36–63
22. Joye, M., Yen, S.M.: The Montgomery powering ladder. In Kaliski, Jr., B.S., Koç, Ç.K., Paar, C., eds.: Cryptographic Hardware and Embedded Systems—CHES 2002. Volume 2523 of Lecture Notes in Computer Science., Springer-Verlag (2003) 291–302
23. Koblitz, N.: Elliptic curve cryptosystems. Mathematics of Computation **48**(177) (January 1987) 203–209
24. Lenstra, A., Lenstra, H., Lovász, L.: Factoring polynomials with rational coefficients. Mathematische Annalen **261**(4) (1982) 515 –534
25. Miller, V.S.: Use of Elliptic Curves in Cryptography. In Williams, H.C., ed.: Advances in Cryptology - Crypto '85. Volume 218 of Lecture Notes in Computer Science., Springer (1985) 417–426
26. Montgomery, P.L.: Speeding the Pollard and elliptic curve methods of factorization. Mathematics of Computation **48**(177) (January 1987) 243–264
27. National Institute of Standards and Technology: FIPS PUB 186-4 Digital Signature Standard (DSS). (2013)
28. Nguyen, P.Q., Shparlinski, I.E.: The insecurity of the elliptic curve digital signature algorithm with partially known nonces. Designs, Codes and Cryptography **30**(2) (September 2003) 201–217
29. Okeya, K., Kurumatani, H., Sakurai, K.: Elliptic curves with the Montgomery-form and their cryptographic applications. In: Public Key Cryptography. Volume 1751 of Lecture Notes in Computer Science., Springer-Verlag (2000) 238–257
30. OpenSSL. `http://www.openssl.org`
31. Ristenpart, T., Tromer, E., Shacham, H., Savage, S.: Hey, you, get off my cloud: Exploring information leakage in third-party compute clouds. In Al-Shaer, E., Jha, S., Keromytis, A.D., eds.: Proceedings of the 16th ACM Conference on Computer and Communication Security, Chicago, Illinois, United States (November 2009) 199–212
32. Shanks, D.: Class number, a theory of factorization, and genera. In: Proceedings of the Symposium of Pure Mathematics 20, American Mathematical Society (1971) 415–440
33. Song, D.X., Wagner, D., Tian, X.: Timing analysis of keystrokes and timing attacks on SSH. In Wallach, D.S., ed.: Proceedings of the 10th USENIX Security Symposium, Washington, DC, United States (August 2001) 25
34. Tromer, E., Osvik, D.A., Shamir, A.: Efficient cache attacks in AES, and countermeasures. Journal of Cryptology **23**(2) (January 2010) 37–71
35. Uht, A.K., Sindagi, V.: Disjoint eager execution: An optimal form of speculative execution. In: Proceedings of the 28th International Symposium on Microarchitecture, Ann Arbor, Michigan, United States (November 1995) 313–325
36. Vaudenay, S.: Security flaws induced by CBC padding. applications to SSL, IPSEC, WTLS. . . . In: EUROCRYPT 2002, Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques, Amsterdam, The Netherlands (April 2002) 534–546
37. Walter, C.D.: Longer keys may facilitate side channel attacks. In Matsui, M., Zuccherato, R.J., eds.: Selected Areas in Cryptography. Volume 3006 of Lecture Notes in Computer Science., Springer-Verlag (2004) 42–57
38. Wiener, M.J., Zuccherato, R.J.: Faster attacks on elliptic curve cryptosystems. In Tavares, S.E., Meijer, H., eds.: Selected Areas in Cryptography. Volume 1556 of Lecture Notes in Computer Science., Springer (1998) 190–200
39. Yarom, Y., Falkner, K.: Flush+reload: a high resolution, low noise, l3 cache side-channel attack. Cryptology ePrint Archive, Report 2013/448 (2013) `http://eprint.iacr.org/`.
40. Zhang, Y., Jules, A., Reiter, M.K., Ristenpart, T.: Cross-VM side channels and their use to extract private keys. In Yu, T., Danezis, G., Gligor, V.D., eds.: Proceedings of the 19th ACM Conference on Computer and Communication Security, Raleigh, North Carolina, United States (October 2012) 305–316