Anonymous Attestation for IoT

Santosh Ghosh, Andrew H. Reinders, Rafael Misoczki, and Manoj R. Sastry

Intel Labs, Intel Corporation, USA
{santhosh.ghosh, andrew.h.reinders, rafael.misoczki, manoj.r.sastry}@intel.com

Abstract. Internet of Things (IoT) have seen tremendous growth and are being deployed pervasively in areas such as home, surveillance, healthcare and transportation. These devices collect and process sensitive data with respect to user’s privacy. Protecting the privacy of the user is an essential aspect of security, and anonymous attestation of IoT devices are critical to enable privacy-preserving mechanisms. Enhanced Privacy ID (EPID) is an industry-standard cryptographic scheme that offers anonymous attestation. It is based on group signature scheme constructed from bilinear pairings, and provides anonymity and sophisticated revocation capabilities (private-key based revocation and signature-based revocation). Despite the interesting privacy-preserving features, EPID operations are very computational and memory intensive. In this paper, we present a small footprint anonymous attestation solution based on EPID that can meet the stringent resource requirements of IoT devices. A specific modular-reduction technique targeting the EPID prime number has been developed resulting in 50% latency reduction compared to conventional reduction techniques. Furthermore, we developed a multi-exponentiation technique that significantly reduces the runtime memory requirements. Our proposed design can be implemented as SW-only, or it can utilize an integrated Elliptic Curve and Galois Field HW accelerator. The EPID SW stack has a small object code footprint of 22kB. We developed a prototype on a 32-bit microcontroller that computes EPID signature generation in 17.9s at 32MHz.

Keywords: Enhanced Privacy ID, EPID, Authentication, Attestation, Bilinear Pairing, Optimal-ate Pairing

1 Introduction

Enhanced Privacy ID (EPID) was introduced by Brickell and Li [11] in 2007 to address the need for anonymous attestation. The authentication technique in EPID is based on Boneh, Boyen, and Shacham’s group signature scheme [6] and Furukawa and Imai’s group signature scheme [20]. One of the main features of EPID consists of its sophisticated revocation techniques which are constructed from pairing-based cryptography [10][13][14].
The EPID scheme follows the privacy-preserving technique in digital signatures that was introduced by Camenisch and Lysyanskaya in 2001 [16] as the generalization of the anonymous credential system due to [15] and the generalization of the group signature scheme proposed by Ateniese et. al [2]. With Strong-RSA assumption, [16] provides the provably secure signature scheme where a member can participate in a group and can generate a signature without revealing its identity. Later in 2004, this scheme was extended to Direct Anonymous Attestation (DAA) [9], which was adopted by the Trusted Computing Group (TCG) as the method for remote attestation in Trusted Platform Module (TPM). The DAA scheme has evolved in the past several years to achieve higher security, improved efficiency and fine-grained revocation capabilities [10–14]. This resulted in EPID being adopted as an ISO standard in 2013 [18].

The EPID scheme as described in [12] is based on bilinear pairings on Barreto-Naehrig (BN) elliptic curves [4], which are pairing-friendly elliptic curves. It uses Optimal-ate pairing [31] on BN curves comprised of three bilinear groups: $G_1$, $G_2$, and $G_T$. In EPID scheme, each of these groups are defined with a 256-bit prime ordered sub-groups where, $G_1$ is defined over a 256-bit BN curve on the base field ($F_q$), $G_2$ is defined on BN curve with sextic twist on $F_{q^2}$ so that a member of $G_2$ is defined as the tuple $(x, y)$ where $x, y \in F_{q^2}$ and the group $G_T$ is defined on $F_{q^{12}}$ as an integer group. The EPID Sign and Verify operations involve multiple Optimal-ate pairing operations as well as multiple exponentiations on $G_1$ and $G_T$ which are essentially computed on 256-bit and 3072-bit numbers. Therefore, a naive implementation of the EPID protocol incurs significantly high latency, compute energy and area/memory footprint overhead which are the bottlenecks for enabling EPID on resource constrained IoT devices.

1.1 Contributions

To address the above problem, we introduce an optimized SW implementation of EPID scheme for resource-constrained embedded IoT platforms. A set of novel techniques has been explored, implemented and carefully integrated resulting in the proposed EPID software stack with minimal code and memory footprint, as well as it keeps up an acceptable latency on 32-bit microcontrollers. Additionally, there is a compile-time knob that replaces the specific SW modules by a driver module of an Elliptic Curve and Galois Field HW Engine. If the target platform consists of such a dedicated HW accelerator then specific operations can be offloaded for additional latency improvements of the EPID operations. The design of such HW Engine is out of the scope of this work. It can be based on any of the existing HW Engines that supports short Weierstrass form of elliptic curves over 256-bit prime field ($F_p$) [8,19,22,26]. The most significant optimization techniques used in our EPID software implementation are described as follows:

– EPID Prime Specific Reduction. We explored a reduction technique targeting the EPID prime modulus which reduces the reduction latency by 50% when compared to traditional Barrett reduction.
Latency-Footprint Tread-off Exponentiations. EPID Sign/Verify operation consists of several exponentiation operations in $G_1$ and $G_T$ which take most of the associated computation time. The lowest latency of those operations can be achieved by increasing the window size with several pre-computations. However, this increases the memory footprint for execution. We developed a footprint-and-latency balanced technique for computing Single as well as Multi-exponentiation operations in $G_1$ and $G_T$. The respective software functions are implemented in constant-time and with unified executions to prevent timing and simple power/EM vulnerabilities.

Pre-computation of members pairing value and basic signature. There are four pairing values involved in EPID Sign which are computed on public and private-key components. Similar four pairings are involved in EPID Verify which are computed only on public-key components. In the proposed EPID software stack, we compute them in an offline fashion and store them within the same public/private key data structure, thus reducing the significant latency overhead of Sign and Verify operations.

Reduction of Verification Latency through revocation test followed by signature verification. In the EPID protocol, there are multiple revocation lists. For a signature verification, the Verifier typically establishes the proof for the signature and then validates that the signature/private-key have not been revoked. The later one is achieved by validating the revocation-proof embedded in the signature. In our investigation, we concluded that the revocation check requires a relatively lower latency when compared to the actual proof generation for the signature validation. Therefore, in our work, we first check that a given signature is valid across all revocation lists then we perform the actual signature validation thus saving a significant latency and compute energy for the verifier.

The paper is organized as follows. A brief introduction of EPID scheme is provided in Section 2. Our proposed software stack is described in Section 3. Section 4 provides our latency optimization techniques in the base-field and its elliptic curve group operations. Section 5 provides detailed descriptions of the extension fields and their execution in our EPID software. Optimal-ate pairing computation has been described in Section 6. EPID Join, Sign and Verify execution techniques are described in Section 7. The implementation results and comparison with state of the art implementations of the underlying functions are captured in Section 8. We conclude the paper in Section 9.

2 Background

As described in [14], the EPID scheme is comprised of four types of entities: an issuer, a revocation manager, users, and verifiers. The issuer can be the same entity as the revocation manager. An EPID scheme has the following four major functions:

- Setup: The issuer creates a group public key and a group issuing private key. The issuer publishes the group public key.
- Join: The issuer and the user cooperate to add the user to the group. The user obtains a membership private key unknown to the issuer.

- Proof of Membership: A prover interacts with a verifier to convince the verifier that he is a member of the group in good standing (i.e., not revoked). It has the following steps: (1) the prover sends a request to the verifier, (2) the verifier responds with a message $m$, (3) the prover generates a signature on $m$ based on his membership private key, and (4) the verifier verifies the signature using the group public key.

- Revocation: The revocation manager puts a group member into one of the revocation lists. There are three types of revocations:
  1. Private-Key-Based Revocation: the revocation manager revokes a user’s membership private key if it is discovered.
  2. Signature-Based Revocation: the revocation manager revokes a user based on a signature created by the user.
  3. Name-Base Revocation: a verifier forces users to use pseudonyms in name-base mode, and revokes individual pseudonyms.

The EPID scheme generates a proof-of-membership signature whose properties match common group signature properties: (1) unforgeability, i.e., only non-revoked group members are able to generate valid signatures, (2) anonymity, i.e., the verifier cannot identify the actual signer given a valid signature, and (3) unlinkability, i.e., it is computationally infeasible to determine whether two different signatures were computed by the same group member. It was developed from the Direct Anonymous Attestation (DAA) scheme [9] and applied Camenisch-Lysyanskaya (CL) signature scheme [16]. For private-key based revocation, it uses verifier local revocation [7], i.e., the revocation check is done only at the verifier’s side. For the other two types of revocation, EPID provides proof of knowledge protocols under the strong RSA assumption and the decisional Diffie-Hellman assumption to prove that a user’s membership has not been revoked by the revocation list [14].

Many group signature schemes handle revocation by adding traceability where the revocation manager has the ability to open a signature and identify the actual signer. To revoke a user based on his signature, the revocation manager first finds out the user’s private key or his identity, then put the user into the revocation list. However, EPID provides much higher privacy than traceability-based signature revocation. Traceability provides that a revocation manager can determine which user generated which signatures without any acknowledgment from the user that is being traced. This is not desirable from a privacy perspective. With EPID, if a user’s membership private key has been revoked, i.e., placed in a revocation list, the user will know that he is revoked or being traced. If the user finds that he is not in the revocation list, then he is assured that nobody can trace him, including the issuer and the revocation manager, which is a major advantage of the EPID scheme.
2.1 EPID Parameters

The order of \( G_1, G_2, G_T \) is a 256-bit prime \( p \). The characteristic of the finite field used in the elliptic curve is chosen as a 256-bit prime \( q \). The parameter \( b \) defines the elliptic curve \( y^2 = x^3 + b \). The parameter \( t \) is used to generate the prime \( q = 36t^4 + 36t^3 + 24t^2 + 6t + 1 \) and \( p = 36t^4 + 36t^3 + 18t^2 + 6t + 1 \). The variable \( \text{neg} \) is used to identify the sign of \( t \) and it is queried in the computation of the Optimal-ate pairing. The parameter \( \beta \) defines the extension field \( F_q^2 \) over \( F_q \). Similarly, the parameter \( \xi \) defines the extension field \( F_q^6 \) over \( F_q^2 \).

\[
\begin{align*}
    p &= FFFFFFFF \text{ FF} \text{CF0CD 4EE5F25E EE71A49E} \\
    &= 0\text{C0C65FB 1299921A F62D536C D10B500D} \\
    q &= FFFFFFFF \text{ FF} \text{CF0CD 4EE5F25E EE71A49F} \\
    &= 0\text{C0C65FB 12980A82 D3292DBB AED33013} \\
    b &= \text{00000000 00000000 00000000 00000000} \\
    &= \text{00000000 00000000 00000000 00000000 00000003} \\
    t &= 6882\text{F5C0 30B0A801} \\
    \text{neg} &= 1 \text{ (where neg is true if t is a negative value)} \\
    \beta &= -1 \\
    &= FFFFFFFF \text{ FF} \text{CF0CD 4EE5F25E EE71A49F} \\
    &= 0\text{C0C65FB 12980A82 D3292DBB AED33012} \\
    \xi[0] &= \text{00000000 00000000 00000000 00000000} \\
    &= \text{00000000 00000000 00000000 00000000 00000002} \\
    \xi[1] &= \text{00000000 00000000 00000000 00000000} \\
    &= \text{00000000 00000000 00000000 00000000 00000001}
\end{align*}
\]

3 Optimized SW Architecture

Figure 1 depicts the hierarchical architecture and underlying functional layers of our embedded EPID Software-Hardware stack. Operations in the lower levels in shaded boxes can be performed by either hardware or software. There is a global flag in the embedded SW stack which indicates whether a hardware accelerator is available on the target platform. If no hardware accelerator is available then the entire hierarchy is implemented and executed in software. However, if the target platform contains a HW accelerator, then the software offloads those operations into the HW. Note that the underlying HW design is out of the scope of this work. However, in the following sub-section we provide how the current SW stack is configured and how it utilizes the underlying HW Engine.

In the proposed EPID SW stack, \( F_q \) is the finite field of order \( q \). \( F_q^2 \) is a degree-two finite field extension of \( F_q \), of order \( q^2 \). \( F_q^6 \) is a degree-three extension
of $F_{q^2}$, of order $q^6$. $F_{q,12}$ is a degree-two extension of $F_{q^6}$, of order $q^{12}$. $F_{q}^*\text{Elem}$ are data structures representing elements of their respective finite fields. EccPointFq holds affine (x,y pair) representations of points on the curve over $F_q$. EccPointJacobiFq holds Jacobi $(X : Y : Z$, corresponding to $X/Z^2$, $Y/Z^3$) representations of points on the curve over $F_q$. EccPointFq2 holds affine (x,y pair) representations of points on EPID’s twisted curve $y^2 = x^3 + b/\xi$ over $F_{q^2}$. EccPointJacobiFq2 holds Jacobi representations of points on EPID’s twisted curve over $F_{q^2}$. In general, $F_{q}^*\text{Add}$ computes the sum of two values, in its respective finite field; $F_{q}^*\text{Sub}$ computes the difference of two such values; $F_{q}^*\text{Mult}$ computes the product; $F_{q}^*\text{Inv}$ computes the multiplicative inverse of the input value; and $F_{q}^*\text{Neg}$ computes the additive inverse of the input value, in the finite field.

### 3.1 Utilization of HW Accelerator

An IoT device may optionally have an inbuilt HW accelerator for performing 256-bit Prime Field and Elliptic Curve operations. Figure 2 depicts a trivial architecture and interface of such a HW engine. It is comprised of an Elliptic Curve Cryptography (ECC) & $F_p$ HW, the Interface and Bus Protocol Logic and a set of Addressable Registers. In the SoC environment, it is connected with an MCU which can perform read and write transactions in the ECC & $F_p$ HW Engine’s address space by Memory Mapped I/O (MMIO) instructions. Through a few writes into some of these addressable registers, the MCU can configure the HW accelerator for a specific operation. Similarly, the MCU can receive
the result of an operation from the HW accelerator by executing MMIO read instructions. We are not going into any further details about the HW accelerator in this paper and keeping major focus on the embedded EPID SW stack for IoT platforms. For further information about how to design and build such a HW engine we refer \[8,19,22,26\] to the readers.

![Fig. 2: The 256-bit Galois Field and Elliptic Curve HW Engine and Interface.](image)

4 Latency Optimization in the Base Field

Galois Field over the 256-bit prime \(q\) is used as base field in the EPID scheme. We use BN curve group \(E(F_q)\) over this base field with prime order \(p\), which is used as the subgroup order in each of the three groups \((G_1, G_2, G_T)\) used in EPID. Integer multiplication and reductions by \(p\) and \(q\) are the basic mathematical operations for computing each functional layers of the EPIS SW stack. The following subsection provides the optimized reduction technique for \(p\) and \(q\) that we explored and implemented in the current EPID SW stack.

4.1 EPID Specific Modulus Reduction

Our new EPID-prime reduction technique has been designed for the prime numbers used in EPID aiming at better efficiency on 32-bit processors. Algorithm 1 provides the detailed steps for computing the proposed reduction which is primarily based on Barrett reduction. It computes the reciprocal of the most significant 32 bits of the prime. Then it multiplies the Barrett multiplier (32-bit reciprocal) with the most significant 32 bits of the intermediate reduced result. It multiplies the most significant 32 bits of the last result with prime modulus. Then it subtracts this result from the most significant \(k+32\) bits of the intermediate reduced result. Finally it left-shifts the intermediate result by 31 bits which is subsequently considered as the new intermediate reduced result. The above steps are repeated until the result is reduced to a \(k\)-bit number.

EPID has a prime \(q\) that is very close to \(2^{256}\), greater than \(2^{256} - 2^{224}\). This lends itself to an efficient reduction method, targeted for 32-bit platforms, that reduces 31 bits at a time. We treat the highest-order 288 bits of a 512-bit number to be reduced like a 288-bit number \(n\), and reduce it to a 257-bit number. Then we left shift the result by 31 bits and repeat the procedure until we ended up with the final 256-bit reduced result. For \(k\) the highest-order 32 bits of \(n\), we know
Algorithm 1 EPID Modulus Reduction

Require: The 256-bit Modulus $P$, 512-bit integer $M$
Ensure: $N = M \mod P$

$i \leftarrow 8$

while $i > 0$ do
    $R_{31:0} \leftarrow 2^{288} / P$
    $S_{31:0} \leftarrow (R \cdot M_{511:224})_{543:512}$
    $T_{287:0} \leftarrow S_{31:0} \cdot P_{255:0}$
    $U_{287:0} \leftarrow M_{511:224} - T_{287:0}$
    $M_{511:0} \leftarrow M_{511:0} \ll 31$
    $i \leftarrow i - 1$
    $M_{511:0} \leftarrow M_{511:0} >> 24$
    $S_{31:0} \leftarrow (R \cdot M_{511:224})_{543:512}$
    $T_{287:0} \leftarrow S_{31:0} \cdot P_{255:0}$
    $U_{287:0} \leftarrow M_{511:224} - T_{287:0}$
    $M_{511:0} \leftarrow M_{511:0} \ll 31$

if $M_{511:256} < P_{255:0}$ then
    Return $M_{511:256}$
else
    Return $M_{511:256} - P_{255:0}$

that $k \cdot q \leq n - k \cdot q + 2^{257}$ because $0 < n - k \cdot q = n - k \cdot 2^{256} + (2^{256} - q) \cdot k < (n \mod 2^{256}) + 2^{224} \cdot k \leq 2^{256} + 2^{256}$. This process can be repeated eight times, and then once more on the remaining eight bits, to reduce a 512-bit number to a 257-bit number, it computes $n' < 2^{256} + 2^{252}$. $n' - q$, and chooses either $n'$ or $n' - q$ as the value of $n \mod q$ based on whether $n' - q$ is at least, or less than, zero.

This method resembles Barrett reduction, but the multiplier is a power of 2, so the step of multiplying by the Barrett multiplier is eliminated; this improves the efficiency of this method without adding more memory cost.

5 Memory and Latency Trade-off in Extension Field

The EPID setup, join, sign verify operations require several extension field arithmetic operations. The optimized formula and implementation techniques of those mathematical operations are highlighted in the following subsections.

5.1 $F_{q^2}$, $F_{q^6}$, and $F_{q^{12}}$ Arithmetic

$F_{q^2}$ arithmetic relies on $F_q[x]/(x^2 + 1)$, a polynomial irreducible over $F_q$. It is stored as a tuple, $(a, b)$ of $F_q$ elements representing a reduced $F_q$ polynomial $a + bx$, where $x^2 = -1$. Addition and subtraction are pairwise addition in $F_q$, and multiplication is done in the same way as the fast complex multiplication: $(a, b)(a', b') = (a \cdot a' - b \cdot b', (a + b) \cdot (a' + b') - a \cdot a' - b \cdot b')$. Multiplicative inverse uses $(a', b') = (a, -b)/[(a, b)(a, -b)]$, where $(a, b)(a, -b)$ is a scalar in $F_q$. 
$F_{q^6}$ is $F_{q^2} [y]/(y^3 - \xi)$, where $\xi$ is a non-square, non-cube element in $F_{q^2}$. Arithmetic is hard-coded to have as few multiplications as possible, as follows:

\[
(a, b, c) \cdot (a', b', c') = \\
\left( (a + b)(a' + b') - a \cdot a' - b \cdot b' \right) \xi + a \cdot a', \\
\xi (c \cdot c' - c \cdot c' + b \cdot b') + (a + c)(a' + c') - a \cdot a' - c \cdot c' + b \cdot b'.
\]

The modular inverse is obtained by similar computation; we construct the conjugate that solves $(a, b, c) \cdot (a, b', c') = (k, 0, 0)$ for some $k$, and then divide by $k$:

\[
(a', b', c') = (a^2 - \xi b^2 c^2, c^2 \xi - a^2 b r, b^2 - a^2 c^2) \\
(a', b', c') \cdot (a, b, c) = (a'^2 + \xi b c + \xi b c', 0, 0) \\
k = a'a + \xi b^2 c + \xi b c' \\
(a, b, c)^{-1} = (a'/k, b^2/k, c'/k)
\]

$F_{q^{12}}$ is $F_{q^2} [z]/(z^2 - y)$, which is an extension field because $\xi$ is not a square or a cube. Arithmetic is done in the same way as in $F_{q^2}$ arithmetic is, with the exception that $(a, b)(a', b') = (aa' - ybb', (a + b)(a' + b') - aa' - bb')$, where $y$ is the cube root of $\xi$, represented as the $F_{q^6}$ element $(0, 1, 0)$.

### 5.2 $E(F_q)$ and $E'(F_{q^2})$ Group Operations

$E(F_q)$ is the elliptic curve over $F_q$ subject to the equation $y^2 = x^3 + 3$, with order $p$. $E'(F_{q^2})$ is the sextic twist of this curve, over $F_{q^2}$, subject to the equation $y^2 = x^3 + 3/\xi$. $\xi$ is picked so that the order of $E'(F_{q^2})$ is $p(2q - p)$. The formula for point addition on the curves is the same, because the Weierstrass curve addition formula does not depend on the constant.

Among the curves of form $y^2 = x^3 + ax + b$, setting $a = 0$ improves the performance of the curve because the addition formula includes an $a$ term, but when $a = 0$, this term disappears. In the EPID software, we implemented the group operations on both the elliptic curves on Jacobian Projective Coordinates [17].

### 5.3 $F_{q^{12}}$ Multi-exponentiation

Multiexponentiation computes $A^{a} B^{b} C^{c} D^{d}$ for $A, B, C, D \in F_{q^{12}}, a, b, c, d \in F_p$. Typical approaches include computing four exponentiations by traditional binary exponentiation, running one loop in which an accumulator is conditionally multiplied by A, B, C, and D as necessary then squared, or computing a lookup table of $[1, a, b, ab, c, ac, be, abc, d, ad, bd, abd, cd, acd, bcd, abcd]$ and then multiplying the accumulator by the appropriate member of this table, as desired. The first approach avoids reusing code, but takes four times as long as the third approach; the second approach saves three squarings per iteration of the loop,
but squaring costs less than multiplication so it is desirable to also reduce the number of multiplications. Exponentiation with table lookup requires a lookup table with at least 11 entries, taking up 4 kB of RAM during runtime even if we assume a, b, c, and d are present as inputs; for simplicity, a table might be compiled of all 16 values, using 6 kB. Side-channel-resistant memory access requires polling all of these values every time a memory lookup is done, which makes its advantages less attractive. Instead, we compromised on a hybrid approach using two tables, one effectively containing 1, a, b, ab and one containing 1, c, d, cd. This approach has the advantage of requiring only that we compute and store only ab and cd, with 2 entries instead of 11. We select from the table a, b, ab, and multiply by the accumulator, then select either the original or the new product; we again select from the table c, d, cd and multiply, then again select the accumulator or the new product as the new value of the accumulator.

Algorithm 2 EPID $F_{q^{12}}$ Multiexponentiation

Require: The $F_{q^{12}}$ elements A, B, C, and D, and $F_p$ exponents a, b, c, and d

Ensure: $R = A^a B^b C^c D^d$

i ← 255
R ← 1 ∈ $F_{q^{12}}$
S ← 1 ∈ $F_{q^{12}}$
ab ← a × b
cd ← c × d

while i ≥ 0 do
    S ← a if $a_i = 1$ else S
    S ← b if $b_i = 1$ else S
    S ← ab if $a_i = b_i = 1$ else S
    S ← R × S
    R ← S if $a_i \lor b_i = 1$ else R
    S ← c if $c_i = 1$ else S
    S ← d if $d_i = 1$ else S
    S ← cd if $c_i = d_i = 1$ else S
    S ← R × S
    R ← S if $c_i \lor d_i = 1$ else R
    i ← i − 1

Return $R = A^a B^b C^c D^d$

6 Optimal-ate pairing over BN Curve

Algorithm 3 computes the Optimal-ate pairing [31]. We chose a pairing friendly Barreto-Naehrig (BN) elliptic curve $E : y^2 = x^3 + 2; z = -(2^{62} + 2^{55} + 1)^i < 0$ for efficiency, with the Optimal-ate pairing. The algorithm consists of two major parts - namely, Miller’s loop, and final exponentiation. Several software and hardware implementations for the Optimal-ate pairing over BN
curves are reported in \cite{5,23,25,27,29}. The EPID uses a negative $t$, because of which, after the Miller loop, we set accumulator $T = [−|s|]Q$, and the value of $f_{(s,Q)}(P)$ is raised to the power $q^6$ which is equivalent to $f^{-1}$ as shown in \cite{1}. The latter operation is computed by conjugation in $F_{q^{12}}$. Thereafter, the algorithm computes $g_{sQ,π⋅Q}^{π⋅Q}(P)$ and $g_{sQ+π⋅Q,−π⋅Q}^{π⋅Q}(P)$, which are multiplied into $f$.

**Algorithm 3** Optimal-ate pairing on BN curve ($t < 0$)

**Require:** $P = (x_P, y_P) ∈ E(\mathbb{F}_q)$, $Q = (x_Q, y_Q) ∈ E(\mathbb{F}_{q^2})$, $s = 6t − 2 = \Sigma_{i=0}^{n-1} s_i2^i$

**Ensure:** $\alpha_{opt}(Q,P) ∈ F_{p^{12}}$

$T ← (x_Q, y_Q, 1)$, $f ← 1$

for $i = n − 2$ downto 0 do

$g ← l(T,T)(P)$, $T ← 2T$, $f ← f^2$, $f ← f ⋅ g$

if $s_i = 1$ then

$g ← l(T,Q)(P)$, $T ← T + Q$, $f ← f ⋅ g$

$T ← −T$, $f ← f^q$

$Q_1 ← π⋅Q(Q)$, $Q_2 ← −π⋅Q(Q)$

$g ← l(T,Q_1)(P)$, $T ← T + Q_1$, $f ← f ⋅ g$

$g ← l(T,Q_2)(P)$, $T ← T + Q_2$, $f ← f ⋅ g$

$f ← \left(f^{q^6−1}\right)^{q^2+1}$

$f ← f^{(q^4−q^2+1)/p}$

return $f$

Algorithm 3 employs arithmetic in $F_{q^{12}}$. High-performance arithmetic over extension fields is achieved through a tower of extensions using irreducible binomials as defined in Section 5. In the EPID protocol, we follow the towering $F_{q^2} → F_{q^6} → F_{p^{12}}$ which makes the arithmetic (mainly squaring) required for this final exponentiation cheaper. The choice $q ≡ 3 \pmod{4}$ accelerates arithmetic in $F_{q^2}$, since $β = −1$ is a square-free element, and multiplications by $β$ can be computed as simple subtractions \cite{21}. The final exponentiation is computed in three major steps: $f ← f^{q^6−1}$, $f ← f^{q^2+1}$, and $f ← f^{(q^4−q^2+1)/p}$. The first two exponentiations are easily computed by using conjugate, inverse, Frobenius and multiplication operations in $F_{q^{12}}$. The last one is the most computationally intensive operation in the final exponentiation step. However, after raising $f$ to the power $q^6 − 1$, it becomes a member of the cyclotomic subgroup in $F_{q^{12}}$ which allows underlying square and exponentiation to be easier than their naive computation in $F_{q^{12}}$, which helps to speed up the last part of the final exponentiation.

7 Optimized EPID Functions

In order to make EPID feasible for IoT embedded devices with scarce computational capability, we developed techniques to minimize the necessary working memory and code size, besides targeting the implementation for 32-bits architec-
tures. Since it is unlikely that edge devices will play the issuer role, we focused on optimizing the operations required by members and verifiers.

7.1 EPID Setup and Join

In EPID, the issuer is responsible for setting up the EPID group for the purpose of device authentication. The EPID group corresponds to a group of trusted devices, as decided by the issuer. The group is created by generating a public/private keypair: the private key is a random $F_p$ element that is not 0 nor 1, and the public key is a tuple $(\text{gid}, h_1, h_2, w)$, which consists of a group ID $\text{gid}$, random $E(F_q)$ elements $h_1$ and $h_2$, and $w = g_2^w$, where $g_2$ is a parameter of EPID and the generator of the group $G_2 \subset E'(F_q^2)$.

Once this group is created, the issuer must communicate with each member via a trusted channel, to assert that the member is the entity it claims to be. It is over this channel that the Join protocol occurs.

EPID Join occurs in several steps; first, the member obtains the group public key to which it is joining, $(\text{gid}, h_1, h_2, w)$. The member generates a private key $f$ which remains a secret, and generates a membership request $F = h_1^f$ with proof of knowledge $c, s$, where $s = r + cf$ and $c = \text{HASH}(h_1^r, F)$ for some random $r \in F_p$, which is sent to the issuer. This membership request $(F, c, s)$ is received by the issuer, who verifies that it is a valid membership request by checking that $H(h_1^c F^{-c}, F) = c = H(h_1^r, F)$. The issuer then computes a random $x$, and sends the member the membership credential $(\text{gid}, A, x)$ where $x \in F_p$ is a random nonzero element and $A = (g_1 F)^x$.

This credential is enough for the member to prove they have the private key $f$ matching the membership credential $A, x$; to prove this, all that the member is required to do is show they possess $A, x$, and $f$ such that $P(A, wg_2^x) = P(g_1 h_1^f, g_2) = P(g_1, g_2) + f \cdot P(h_1, g_2)$, which proves that $A = (g_1 F)^x$. The pairing function gives EPID the flexibility to randomize and therefore obscure the identity of the member, by obscuring the identity-associated or secret key values of $f, A$ and $g_2^x$, and by randomizing any values computed from them, while still proving that it possesses values which match in this way. These values can be anonymized, by randomly obscuring them, or pseudonymized, by basing one of the obscuring values on the hash of a known string. Pseudonymizing the strings links signatures signed with that pseudonym together, at the choice of the member, usually at the request of a verifier.

7.2 EPID Sign

The member prepares to sign messages with its private key by precomputing several pairing values; the signature itself is computed without running the expensive pairing function, but it still must computes several exponentiations in the field $F_{q^{12}}$. This requires us to optimize for $F_{q^{12}}$ multiexponentiation as described before, as well as elliptic curve operations.
Member precomputation, which occurs before any signature is signed, consists of the four Optimal-ate pairing computations $pm_1 = P(A, g_2)$, $pm_2 = P(h_1, g_2)$, $pm_3 = P(h_2, g_2)$, $pm_4 = P(h_2, w)$, where $A$ belongs to the member’s private key, $h_1, h_2, w$ are part of the public key and $g_2$ is the generator of $G_2$.

EPID Sign must both prove possession of a private key and hide the identity of that private key. Signing a message $m$ with private key $A, x, f$ and group public key $(gid, h_1, h_2, w)$ goes as follows:

A random name is produced, in the form of an elliptic curve point $B \in E(F_q)$.

Random parameters $a, r_x, r_f, r_a, r_b$ are generated to obscure the values of $x, f, a,$ and $b$. Then the signature is computed as:

$$b = ax$$

$$K = B^f$$

$$T = Ah^a$$

$$R_1 = B'^f$$

$$R_2 = pm_1^{−r_x} \times pm_2^f \times pm_3^{r_a−r_x} \times pm_4^a$$

$$c = H(H(p)|g_1||g_2||h_1||h_2||w||B||K||T||R_1||R_2)||m$$

$$s_x = r_x + c \cdot x$$

$$s_f = r_f + c \cdot f$$

$$s_a = r_a + c \cdot a$$

$$s_b = r_b + c \cdot b$$

The final signature is $(B, K, T, c, s_x, s_f, s_a, s_b)$.

7.3 EPID Verify

Verification works backwards to reconstruct the values of $R_1$ and $R_2$ from $c$. The verifier also has four precomputed pairing values: $pv_1 = P(h_1, g_2)$, $pv_2 = P(h_2, g_2)$, $pv_3 = P(h_2, w)$, $pv_4 = P(g_1, g_2)$, where $h_1, h_2, w$ are part of the public key and $g_1, g_2$ are the generator of $G_1, G_2$ respectively.

$$R_1 = K^{−c} \cdot B^{s_f}$$

$$R_2 = P(T, g_2^{−s_xw^{−c}}) \times pv_1^{s_f} \times pv_1^{s_a} \times pv_1^{s_b}$$

The verifier verifies that $R_1$ and $R_2$ are correct by recomputing the hash that the signer produced, proving that $R1$ and $R2$ are correct, which therefore proves that the relationship between $P(B, g_2)$ and $P(K, g_2)$ is the same exponent as exists between $P(h_1, g_2)$ and $P(h_1^{−f\gamma}, wg_2^x)$, which is to say the exponent $f \cdot \gamma$ is hidden from the signer by the value of $x$, which is hidden from the verifier with $r_x$. The other values exist to hide the values of the variables in question.

The verifier has to also ensure that the member is not on any of the revocation lists. The verifier does this in three ways.
Private key revocation: The private key revocation list is a list of private keys $f$ that have been exposed and subsequently revoked. Because $K = B^f$ in the signature provided, the verifier has only to check that $K \neq B^{f'}$ for each $f'$ in the list. If the two were to match, the member key has been revoked.

Group revocation: If the group has been revoked, the verifier rejects all signatures signed under that group. This only requires checking the gid. Verifier-based revocation: The verifier can request that signers sign using $B = H(\text{pseudonym})$. The signer then produces signatures with a unique, identifiable $K$, which is constant only when signing messages under that pseudonym. The verifier has the option to restrict or block certain members by adding this $K$ to the list and rejecting signatures whose $B$ and $K$ match this list.

Signature revocation: This revocation is interactive; the user must submit proofs that it is not on the list of revoked signatures. For each revoked signature $B', K'$, the member chooses random nonzero $F_p$ elements $\mu$ and $\nu = -f \cdot \mu$, and random nonzero $F_p$ elements $r_\mu, r_\nu$. Then the member computes:

$$ T = K'^{\mu} B'^{\nu} $$
$$ R_1 = K'^{\nu} B'^{\nu} $$
$$ R_2 = K'^{\mu} B'^{\nu} $$
$$ c = H(p || g_1 || B || K || B' || K' || T || R_1 || R_2 || m) $$
$$ s_\mu = r_\mu + c \cdot \mu $$
$$ s_\nu = r_\nu + c \cdot \nu $$

$\sigma = (T, c, s_\mu, s_\nu)$; $T$ is essentially $K^{(f' - f) \mu}$, but by the signer proving it knows $\mu$ and $\nu$, and that $\nu = -\mu \cdot f$, it pre-commits to the exponents being the same. If $\nu \neq -\mu \cdot f$, then $R_1 \neq K'^{\mu} B'^{\nu}$; the member cannot force a particular value of $c$, so this amounts to a proof of that relationship. The same is true with respect to the value of $T$ and the relationship between $B'$ and $K'$, which also proves that the signature containing $B'$ and $K'$ was signed by a different member. The verifier computes $R_1 = K'^{\mu} B'^{\nu} T^{-c}, R_2 = K'^{\mu} B'^{\nu}$, and recomputes $c$ to make sure the value is correct.

8 Footprint and Performance

The complete embedded EPID software stack was developed in C and designed targeting 32-bit microcontrollers as [24] and [28] demonstrated runtime data that discourage the use of any smaller (8-bit and 16-bit) microcontrollers for computing public key operations. The object code of the signer/verifier/issuer library is 22k bytes, compiled with gcc using the compilation flag: -Os, targeted for a representative IoT platform based on a low-power 32-bit Intel Quark microcontroller with 80k bytes internal SRAM. Using static code analysis, we calculated that the sign/verify operations consume about 10k bytes in the memory stack. Table 1 provides the latency of underlying operations of our EPID implementation on Intel® Curie® IoT platform.
Table 1: Latency of Galois Field and Elliptic Curve Operations on Intel® Curie®

<table>
<thead>
<tr>
<th>Operation *</th>
<th>(\mathbb{F}_q)-M</th>
<th>(\mathbb{F}_{q^2})-M</th>
<th>(\mathbb{F}_{q^{12}})-M</th>
<th>(\mathbb{E}(\mathbb{F}_q))-ME</th>
<th>(\mathbb{E}(\mathbb{F}_{q^2}))-SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock Cycles</td>
<td>8,900</td>
<td>27,998</td>
<td>553,415</td>
<td>437,623</td>
<td>54,798,513</td>
</tr>
</tbody>
</table>


The Optimal-ate pairing is one of the most expensive operations used by EPID. There are four such pairing values that are involved in EPID signing operation. Each member precomputes these values and uses them for signing multiple messages with the same key. Table 2 provides a performance comparison of our software implementation of the Optimal-ate pairing, running on Intel®’s Curie® platform with some of the existing implementations in different platforms. The Optimal-ate SW running on Intel Core processors are significantly faster than microcontrollers due to two main reasons: 64-bit instructions and the availability of high volume cache memory.

Table 2: Latency (Clock Cycles) of Optimal-ate pairing on 256-bit BN curves

<table>
<thead>
<tr>
<th>Work, Year</th>
<th>Platform</th>
<th>Latency (Clock Cycles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[27], 2010</td>
<td>Intel Core 2 Quad Q6600 (64-bit)</td>
<td>4,470,408</td>
</tr>
<tr>
<td>[5], 2010</td>
<td>Intel Core i7 (64-bit)</td>
<td>2,330,000</td>
</tr>
<tr>
<td>[1], 2011</td>
<td>Intel Core i5 (64-bit)</td>
<td>1,688,000</td>
</tr>
<tr>
<td>[30], 2014</td>
<td>ARM Cortex-M0 + HW module</td>
<td>47,643,000</td>
</tr>
<tr>
<td>Ours, 2018</td>
<td>Intel Curie (32-bit MCU)</td>
<td>188,053,871</td>
</tr>
</tbody>
</table>

Table 3 provides the performance of the full EPID software stack on Intel Curie platform. On Intel Curie platform, one EPID sign takes 572,902,590 clock cycles or 17.9s when it runs at 32MHz clock. Similarly, our experiment shows each Verify takes 1,020,079,132 clock cycles or 31.9s at the same clock frequency. It also provides additional results on Intel Core i5 with 32-bit mode. It is observed that there are significant latency difference between two 32-bit platforms which is mainly due to the availability of high-speed cache memory in the Core i5 processor whereas Curie is based on 32-bit Intel Quark microcontroller with 80kB low-speed SRAM.

Table 3: Performance of EPID SW stack on Intel Curie

<table>
<thead>
<tr>
<th></th>
<th>Intel Curie (32-bit) (Clock Cycles)</th>
<th>Intel Core i5 (32-bit) (Clock Cycles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPID Sign</td>
<td>572,902,590</td>
<td>194,199,881</td>
</tr>
<tr>
<td>EPID Verify</td>
<td>1,020,079,132</td>
<td>347,854,538</td>
</tr>
</tbody>
</table>
As we discussed before, the performance of the EPID scheme can be improved further by utilizing the underlying ECC & $F_p$ HW Engine and implementing/compiling our proposed software with a specific knob. In typical such hardware engines [8,19,22,26], each modulus multiplication can be computed in as low as 100’s of clock cycles compared to 8,900 clock cycles on an Intel Curie platform. Therefore, our proposed EPID software stack in a suitable IoT platform can perform one EPID Sign or anonymous attestation in less than 1s.

9 Conclusion

IoT devices are used in a wide range of applications that deal with privacy sensitive data. Enhanced Privacy ID (EPID) is an industry standard that provides anonymous attestation. EPID Sign/Verify operations involve large number arithmetic, and are computationally and memory intensive. This makes it challenging for deploying EPID in resource constrained IoT devices. In this paper, we have proposed a design for small footprint anonymous attestation based on EPID, and built a prototype to demonstrate its feasibility for IoT. The EPID SW stack has a small object code footprint of 22kB. Results from our prototype on a 32-bit microcontroller show that EPID Sign has latency of 572,902,590 clock cycles or 17.9s at 32MHz.

References

5. Beuchat, J., González-Díaz, J.E., Mitsunari, S., Okamoto, E., Rodríguez-Henríquez, F., Teruya, T.: High-speed software implementation of the optimal


