Compact and Provably Secure Lattice-Based Signatures in Hardware


Document Version:
Other version

Queen’s University Belfast - Research Portal:
Link to publication record in Queen’s University Belfast Research Portal

Publisher rights
© 2017 The Authors.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Compact and Provably Secure Lattice-Based Signatures in Hardware

James Howe, Ciara Rafferty, Ayesha Khalid, and Máire O’Neill
Centre for Secure Information Technologies (CSIT), Queen’s University Belfast, Northern Ireland,
{jhowe02, c.m.rafferty, a.khalid, maire.oneill}@qub.ac.uk

Abstract—Lattice-based cryptography is a quantum-safe alternative to existing classical asymmetric cryptography, such as RSA and ECC, which may be vulnerable to future attacks in the event of the creation of a viable quantum computer. The efficiency of lattice-based cryptography has improved over recent years, but there has been relatively little investigation into hardware designs of digital signature schemes. In this paper, the first hardware design of the provably secure Ring-LWE digital signature scheme, Ring-TESLA, is presented, targeting a Xilinx Spartan-6 FPGA. The results better compactness of all previous lattice-based digital signature schemes in hardware, and can achieve between 104-785 signatures and 102-776 verifications per second.

Keywords—lattice-based cryptography, digital signatures, post-quantum cryptography, hardware security, FPGA.

I. INTRODUCTION

A digital signature scheme (DSS) is important for building secure systems and is widely used in most real world security protocols. Almost all currently used DSSs are based on the hardness of the factoring problem (RSA) or the discrete logarithm problem (DSA/ECDSA). With the potential advent of a quantum computer in the not too distant future, current asymmetric cryptography would be rendered insecure. Indeed, quantum computers are expected to break all currently secure instances of RSA and ECDSA in polynomial time. One potential solution is to adopt the use of DSSs based on the hardness of certain lattice problems which are assumed to be resistant to quantum attacks.

Due to significant research advancements in recent years, lattice-based schemes have now become viable alternatives to existing asymmetric cryptography. There are several practical proposals based on varying hard lattice-based problems, such as the NTRU cryptosystem and and Ring-LWE [1], [2]. For further background on practical lattice-based DSSs, the reader is referred to a recent survey on the state-of-the-art [3].

Previous hardware designs of lattice-based DSSs, the GLP scheme [4] and the BLISS scheme [5], demonstrate good performance and outperform ECDSA and RSA. However, significant hardware resources are consumed by the costly discrete Gaussian sampler in the BLISS scheme. Additionally, the schemes rely on extra security assumptions, which do not offer the very appealing average-case to worst-case hardness property offered by Ring-LWE. This quality renders all cryptographic constructions based on it secure, under the assumption that worst-case lattice problems are hard. Furthermore, for both GLP and BLISS, their parameters are not chosen directly from their security reduction, meaning their instantiations are not provably secure.

An alternative lattice-based signature scheme, Ring-TESLA [2], has been proposed which does not require discrete Gaussian sampling during sign or verify, and offers average-case to worst-case hardness with a tight security reduction and a provably secure instantiation. A tight security reduction implies the cryptoscheme is no easier to solve than its hardness problem. Ring-TESLA competes well with GLP and BLISS in software [2], but as yet no hardware designs exist.

This paper presents the first hardware designs of the Ring-TESLA signature scheme. The proposed designs are compact, targeting long-term security and low-area applications. The paper is structured as follows: the Ring-TESLA scheme is detailed in Section II. Section III outlines the proposed hardware designs of the signature scheme and results are given in Section IV.

II. IDEAL LATTICE-BASED SIGNATURES

Lattice-based cryptography is emerging as a promising quantum-resistant alternative to ECC or RSA, and offers efficient performance for both encryption and signatures. For significant efficiency gains, ideal lattices are usually used which allow for smaller key sizes and faster computations by computing over a specific algebraic structure. The Ring-LWE problem [6], commonly used in lattice-based cryptography, is well studied and demonstrates strong computational hardness.

The most practical lattice-based DSSs are based upon the Fiat-Shamir paradigm [3], such as the state-of-the-art BLISS by Ducas et al. [1], which is based upon ideal lattices with NTRU assumptions. NTRU cryptoschemes have existed for a significant period of time, with the only current serious break in NTRU-based schemes targeted NTRUSign [7]. However, the hardness assumptions of NTRU is not related to the hardness of worst-case lattice problems, a useful property of Ring-LWE [8]. Accordingly, a lattice-based DSS based on the Ring-LWE problem has been proposed by Akleylek et al. [2], named Ring-TESLA. Ring-TESLA provides three appealing properties.

Firstly, Ring-TESLA provides a tight security reduction, a provably secure instantiation, and worst-case hardness, which is not provided by GLP or BLISS. Cryptoschemes that have provably secure instantiations are considered stronger in the sense of security [9], especially when non-tight cryptoschemes have been shown to provide weaker security assurances [10].

Secondly, the Ring-TESLA Sign and Verify algorithms do not require discrete Gaussian sampling, which is instead
TABLE I: The 128-bit parameter set for Ring-TESLA.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattice dimension ( n )</td>
<td>512</td>
</tr>
<tr>
<td>Modulus ( q \cdot \log_2(q) )</td>
<td>51750913, (26)</td>
</tr>
<tr>
<td>Weight of the challenge ( \omega )</td>
<td>19</td>
</tr>
<tr>
<td>Gaussian std. dev. ( \sigma )</td>
<td>52</td>
</tr>
<tr>
<td>Drops input ( d )</td>
<td>23</td>
</tr>
<tr>
<td>Error threshold ( L )</td>
<td>2766</td>
</tr>
<tr>
<td>Sign/Verify thresholds ( B, U )</td>
<td>( 2^{22} - 1, 3173 )</td>
</tr>
<tr>
<td>Repetition rate</td>
<td>2.9</td>
</tr>
</tbody>
</table>

To sign a message \( \mu \), a uniform polynomial \( y \overset{\$}{\leftarrow} \mathbb{R}_q \) is generated for use in the calculation of the signature and for validity checks. Firstly, it is used to calculate intermediate polynomials \( v_1 \equiv a_1 y \mod q \) and \( v_2 \equiv a_2 y \mod q \), which are input into the hash function \( H(\cdot) \) along with the message data \( \mu \) to output the bit-string \( c \). An encoding function \( F : \{0,1\}^n \rightarrow \mathbb{B}_{n,\omega} \) (as described in [13]) then maps this bit-string to a LHW polynomial \( c \), with \( \omega \) 1 values and \( n - \omega \) 0 values. The LHW polynomial \( c \) is then used to calculate the signature \( z \equiv y + sc \) as well as the polynomials \( w_1 \equiv v_1 - e_1 c \mod q \) and \( w_2 \equiv v_2 - e_2 c \mod q \), which are used to check the validity of a signature. The verification algorithm is essentially equivalent to the signing algorithm without uniform polynomial generation and the calculation of \( z \).

III. HARDWARE IMPLEMENTATION

Two separate hardware designs for the Sign and Verify algorithms in the Ring-TESLA DSS are proposed. The required modules for both Sign and Verify are a polynomial multiplier, a hash function and a low Hamming weight (LHW) multiplier. Figure 1 illustrates the proposed hardware design of Ring-TESLA Sign. The Verify design is adapted from the proposed Sign hardware design.

A. Hardware components

Polynomial multiplication is the most expensive module required in the proposed designs in terms of latency consumption. The most commonly chosen method for modular polynomial multiplication is the number theoretic transform (NTT), which offers fast performance and incorporates the reduction modulo \( q \). However, it is costly in terms of hardware usage and there are significant restrictions on the parameter selection when a NTT multiplier is used. Alternatively, traditional multiplication techniques can be employed to carry out the polynomial multiplication operations, incurring an additional latency cost. An additional modular reduction module is therefore required.

In this design, the polynomial multiplication is carried out using a variant of schoolbook multiplication, known as Comba multiplication [14], which improves the performance by combining carry handling and reducing write access to memory. This multiplier type fits within the overall compact design goal, unlike NTT. The Comba multiplier is particularly suitable for the FPGA platform and exploits the fast arithmetic within the DSP units [15]. The multiply-and-accumulate (MAC) operations are computed within each DSP slice until the inner products of the schoolbook multiplier are complete.

The modular reduction component uses the Barrett method [16]. Any generic modulus can be used within Barrett reduction, with only one pre-computation is required. Two multiplication units with a maximum of two subtractions are needed in Barrett reduction [17]. In the proposed designs, an individual Comba multiplier is reused to carry out the modular reduction. The combination of Comba multiplication and Barrett reduction create an overall modular multiplication component, which multiplies polynomials over the ring \( \mathbb{R}_q \).

The SHA3 hash function, Keccak, is chosen as the random oracle in the proposed Ring-TESLA designs, due to its speed in hardware as well as its post-quantum security [18]. A LHW
polynomial multiplication module is also designed to compute on the LHW output of the hash function. The LHW multiplier, used in both Sign and Verify, uses the column-wise schoolbook technique, since the LHW calculations only require a small amount of shift and adds. Column-wise is preferred over row-wise as it requires less storage.

B. Signing and Verifying Hardware Designs

Figure 1 shows the Sign hardware architecture for Ring-TESLA. Two dual-port 18 Kb block RAMs (BRAM18) are used to store the global constants \(a_1, a_2\). An unrolled x32 Trivium component is used to generate uniform random bits for the polynomial \(y \in R_{2^{32}}\), whose input, as well as the input of the polynomial \(a_1\) and \(a_2\) is controlled by two counters, which increment on the ready signal of the modular multiplication component. The global constant polynomial selection (\(a_1\) or \(a_2\)) is controlled by the finite state machine. Once each element of \(v_1\) and \(v_2\) is output from the modular multiplication module, the \(d\) least significant bits of each are stored in RAM for use in the hash function, where the full values are also stored for use in the rejection stage.

The binary string output from the hash-function \((c)\) is input to the encoding function \(F(c)\), which outputs a LHW polynomial \(c\). This prior knowledge of a LHW polynomial allows the use of a LHW polynomial multiplier, which only computes values for non-zero elements. The discrete Gaussian distributed \((D_\sigma)\) secret-keys \(s, e_1, e_2\) are also LHW since, for instance, the probability a sample \(x \leftarrow D_\sigma\) also satisfies \(x \in \{-100, 100\}\) (thus 6-7 bits in length) is around 95%.

The LHW polynomial \(c\) is an input into every LHW computation. This includes the calculation of the ciphertext \(z\), and the variables used to accept or reject the signature, \(w_1\) and \(w_2\). The LHW polynomial multiplier exploits the fact that the hash output polynomial \(c\) has only \(\omega = 19\) elements equal to a one, and \(n-\omega = 493\) zero elements. The same LHW multiplier component is reused to calculate the polynomials \(z, w_1,\) and \(w_2\), sequentially. Once a coefficient is output from the LHW multiplier, it is processed by the rejector, which checks whether its size is valid for output. Rejecting a signature element-wise is preferable to minimise run-time.

The column-wise LHW module stores the secret-key values \((s, e_1, \text{ and } e_2)\) in a single-port distributed RAM, and takes the input \(c\) from the hash function. Then, for each element in \(s, e_1, \text{ and } e_2,\) the multiplier accumulates the inner products using a MAC unit. Once each column-wise element is calculated, it is added/subtracted to/from the corresponding values in \(v_1, v_2, \text{ or } y',\) and the final values are checked with their rejection conditions (Line 8 in Algorithm 1), where only the signature \(z\) and hash-string \(c\) are stored. This approach is advantageous since the rejection validity is calculated instantly and in parallel to the next LHW element calculation.

The Sign and Verify FSM operates in two-stages and in a pipelined fashion, due to the latency of the calculations of \(v_1\) and \(v_2\). For the first signature, \(y\) is generated during initialisation when the global constants are read in, with an additional \(y'\) polynomial (see Figure 1) generated during the calculation of \(v_1\) and \(v_2\). Once these calculations have finished, \(y\) is swapped with \(y'\), so the calculations of \(v_1\) and \(v_2\) for the next signature can begin again. This removes the hashing and LHW calculations from the critical path and cycle count, as well as savings for generating a new polynomial \(y\).

Verify operates in a similar fashion to Sign. The Comba/Barrett polynomial multiplier calculates \(a_1z\) and \(a_2\), with \(z\) being stored in BRAM. The results are also stored in RAM and are updated after subtraction with the LHW calculations of \(t_1c\) and \(t_2c\). The final results are input into the hash function, where the signature is validated if the hash value matches the hash-string input from the signature.

C. Optimised design for accelerated performance

There is a trade-off between the latency achievable by the proposed Ring-TESLA hardware designs of Sign and Verify and the associated hardware resource usage. The proposed designs, named SB-I Sign and SB-I Verify, offer reduced area consumption at the cost of additional latency. In these designs, a standard Comba multiplier with one Barrett modular reduction unit is employed. The Barrett modular reduction unit is carried out in parallel while the Comba multiplier is used to minimise latency of the modular multiplication unit. These designs target low area applications, where there is a need for a provably secure instantiation and/or reduced area consumption, and where a slower performance may be acceptable.

The bottleneck in Ring-TESLA is polynomial multiplication. Three additional designs SB-II, SB-IV, and SB-VIII, are undertaken having two, four, and eight parallel Comba multipliers, respectively. These parallelised designs utilise extra BRAM for \(a_1\) and \(a_2\) access. To minimise latency, in all proposed designs only one modular reduction is employed, running in parallel with the multiple Comba multiplier units.

The modulus in this scheme has a length of \(\log_2(q) = 26\)-bits. For each 26-bit multiplication, the Comba multiplier occupies 2 DSP slices on the target Spartan-6 FPGA.

IV. RESULTS AND CONCLUSIONS

The proposed architectures are implemented using the Xilinx ISE Design Suite 14.7 synthesis tool. The target device is a Xilinx Spartan-6 FPGA (S6 LX25). Table II shows the post-place and route results for the proposed hardware designs of Ring-TESLA Sign and Verify. These designs fit comfortably on the low-end FPGA. As expected, the optimised designs of SB-II, SB-IV, and SB-VIII have reduced latency in comparison to SB-I, at the cost of additional area consumption. Results indicate that up to 785 operations per second can be achieved by the proposed designs, with an associated low area cost.
The proposed designs are compared to existing classical DSSs currently used in practice (RSA/ECDSA), and similar lattice-based cryptosystems (GLP/BLISS), as seen in Table II. The Ring-TESLA results compare well with or better than the classical hardware designs of RSA and ECDSA, in terms of lower area consumption and operations per second.

The proposed designs significantly better existing lattice-based DSSs in terms of area consumption. Ring-TESLA Sign SB-I results show ≈45% FPGA slice reduction compared to GLP and BLISS, with results for SB-II and SB-IV reducing FPGA slice consumption to ≈34%. However, the results have an increase in latency to GLP and BLISS. The Ring-TESLA algorithms contribute to this comparative increase in latency. More specifically, Ring-TESLA Sign requires two full polynomial multiplications (for \(v_1\) and \(v_2\)), whereas at the same stage GLP and BLISS only require one. The designs are also compact when considering the larger operand size for Ring-TESLA (26-bits), compared to GLP (23-bits) and BLISS (14-bits).

To conclude, the first hardware designs of the provably secure Ring-TESLA signature scheme are proposed in this research, targeting a low-cost Spartan-6 FPGA device. In the context of design space exploration, using multiple parallel Comb multipliers (plus a Barret modular reduction) instead of an NTT was chosen for design compactness. The designs fit comfortably on a low-end Spartan-6 FPGA, despite having large memory requirements and computationally expensive algorithms. It understandably lacks in throughput performance compared to other lattice-based DSSs, but this is the trade-off for the stronger security it provides. The proposed hardware designs are practical for low-area or high security applications, where tighter security reductions are desirable, at the cost of a slower performance.

REFERENCES


