Frodo: Take off the Ring!
Practical, Quantum-Secure Key Exchange from LWE

Joppe Bos
NXP Semiconductors
joppe.bos@nxp.com

Craig Costello
Microsoft Research
craigco@microsoft.com

Léo Ducas
CWI
l.ducas@cwi.nl

Illya Mironov
Google Inc.
mironov@gmail.com

Michael Naehrig
Microsoft Research
mnaehrig@microsoft.com

Valeria Nikolaenko*
Stanford University
valerini@stanford.edu

Ananth Raghunathan
Google Inc.
pseudorandom@gmail.com

Douglas Stebila
McMaster University
stebilad@mcmaster.ca

ABSTRACT

Lattice-based cryptography offers some of the most attractive primitives believed to be resistant to quantum computers. Following increasing interest from both companies and government agencies in building quantum computers, a number of works have proposed instantiations of practical post-quantum key exchange protocols based on hard problems in ideal lattices, mainly based on the Ring Learning With Errors (R-LWE) problem. While ideal lattices facilitate major efficiency and storage benefits over their non-ideal counterparts, the additional ring structure that enables these advantages also raises concerns about the assumed difficulty of the underlying problems. Thus, a question of significant interest to cryptographers, and especially to those currently placing bets on primitives that will withstand quantum adversaries, is how much of an advantage the additional ring structure actually gives in practice.

Despite conventional wisdom that generic lattices might be too slow and unwieldy, we demonstrate that LWE-based key exchange is quite practical: our constant time implementation requires around 1.3ms computation time for each party: compared to the recent NewHope R-LWE scheme, communication sizes increase by a factor of 4.7×, but remain under 12 KiB in each direction. Our protocol is competitive when used for serving web pages over TLS; when partnered under 12 KiB in each direction. Our protocol is competitive when used for serving web pages over TLS; when partnered with ECDSA signatures, latencies increase by less than a factor of 1.6×, and (even under heavy load) server throughput only decreases by factors of 1.5× and 1.2× when serving typical 1 KiB and 100 KiB pages, respectively. To achieve these practical results, our protocol takes advantage of several innovations. These include techniques to optimize communication bandwidth, dynamic generation of public parameters (which also offers additional security against backdoors), carefully chosen error distributions, and tight security parameters.

Keywords

Post-quantum cryptography; learning with errors; key exchange; OpenSSL; TLS

1. INTRODUCTION

Recent advances in quantum computing (cf. [26, 39]) have triggered widespread interest in developing practical post-quantum cryptographic schemes [48]. Subsequently, standards bodies and government agencies have announced their intentions to transition to cryptographic standards that offer quantum-resistance; this includes the National Institute of Standards and Technology (NIST) [50], the National Security Agency (NSA) [51], and the PQCRYPTO project [9] funded by the European Union.

Traditional number-theoretic cryptographic problems such as the integer factorization problem and the discrete logarithm problem (over both multiplicative groups modulo a prime \( p \) and elliptic curve groups) are vulnerable to polynomial-time quantum attacks [60, 56]. Lattice-based cryptography, beginning with Ajtai’s seminal work [4] (cf. the recent survey [54] for a comprehensive list of relevant references), is an exciting field of research that, in addition to providing a richer diversity of underlying primitives, offers the potential to build practical quantum-resistant cryptography.

1.1 Key Exchange and Forward Secrecy

Although large-scale quantum computers are likely to affect all of the cryptographic building blocks used to secure the internet, the threat of quantum computing to each specific primitive differs significantly. On the symmetric side, the cryptanalytic speedups (using Grover’s algorithm [32]) aided by a quantum computer are well-understood. It is currently believed that it will suffice to double the secret key sizes of symmetric encryption algorithms and MACs to defend against quantum computers.
On the asymmetric side, cryptographers are currently examining the potential of a new range of mathematical problems (like those based on lattices) for post-quantum security. While there are several interesting candidates for post-quantum signature schemes, a more pressing concern to practitioners is the deployment of post-quantum public key encryption and key exchange protocols: the secrecy of today’s communications can be compromised by a quantum computer that is built decades into the future (by an attacker who stores these communications until then); on the other hand, authentication can only be forged at connection time.

In this paper, we focus on the most pressing issue in public key post-quantum cryptography: the development of a secure, practical key exchange protocol that can be used to secure internet traffic in anticipation of a quantum adversary. This focus on (quantum-secure) key exchange aligns with the notion of forward secrecy deployed in TLS and, for example, the notion of long-term security used for classified documents by government agencies (see Table 4 in [11]).

1.2 Generic vs. Ideal Lattices

Public-key cryptography is built on the presumed difficulty of hard computational problems. Whenever such a problem is introduced and has potential as a cryptographic primitive, cryptographers naturally study whether there are special instances of the problem that offer practical benefits, and in particular, whether they do so without compromising the presumed hardness. There are several historic examples of special instances that have proven to be a disastrous choice compared to suitably chosen random instances, e.g., the use of supersingular curves in elliptic curve cryptography (ECC) [46, 28]. In some cases, special instances introduce well-understood security issues that can be weighed against potential benefits, e.g., the use of small public exponents in RSA [22]. In many scenarios, however, the size of such a security gap, or whether there is a gap at all, remains unclear.

One example of an unknown security gap currently exists in lattice-based cryptography. Five years after Regev’s paper introducing the Learning With Errors (LWE) problem was published [57], Lyubashevsky, Peikert and Regev [45] introduced a specialized version of the problem called Ring Learning With Errors (R-LWE), which offers significant storage and efficiency improvements. LWE is a mature and well-studied cryptographic primitive that relies only on the worst-case hardness of a shortest vector problem in generic lattices. R-LWE has additional algebraic structure and relies on the (worst-case) hardness of problems in ideal lattices. Ideal lattices correspond to ideals in certain algebraic structures, such as polynomial rings. NTRU lattices [34], which are a class of lattices with some structure (but different from R-LWE lattices) have also been used to build cryptosystems.

The hardness of lattice problems on regular lattices as well as ideal lattices merits more study. Although the algebraic structure of R-LWE (and NTRU) makes for more practical key-sizes and protocol communication, this algebraic structure might inspire less confidence in the underlying security. Currently, the best algorithms to solve hard problems in ideal lattices [21, 43] are the same as those that target regular lattices (modulo small polynomial speedups), and it is not known whether the R-LWE problem is significantly easier than the LWE problem for the same parameter sizes. Certain sieving algorithms obtain a constant factor time and/or space improvement in the ideal case [59, 15], but (at best) this only shaves a few bits off of the known bit-security. At the very least, however, the additional ring structure might introduce subtleties in the choice of error distribution [55].

Several recent papers [53, 14, 64, 6] have proposed and implemented key exchange variants that rely on the hardness of the R-LWE problem [45]. In contrast, in this paper we develop and evaluate a secure and practical key exchange protocol from the original LWE problem [58], the security of which is based on hard problems in lattices that do not possess any ring structure. While both academic [9] and government [10, 49] bodies are currently exploring candidates for post-quantum cryptography, we believe it prudent to give a concrete evaluation of a lattice-based scheme that is invoked without the additional structure, and to draw comparisons to previous ring-based implementations. The design of our implementation and its accompanying security analysis was performed with a view towards real-world (and in particular, internet) deployment. To our knowledge, there has not yet been a practical implementation of key exchange based on the original LWE problem.

1.3 Our Contributions

We demonstrate the feasibility of LWE-based key exchange with a new key exchange protocol which we call “Frodo,” and we provide a proof of its security based on the original LWE problem. We give parameter sets for LWE at several security levels, including “recommended” parameters that achieve 128-bit security against quantum adversaries (using the best known attacks that incorporate recent advances in lattice cryptanalysis), and “paranoid” parameters that are based on lower bounds for sieving algorithm complexities. Our protocol incorporates several innovations:

- **Efficiently sampleable noise distribution.** We present four discrete noise distributions that are more efficient to sample than previously used discrete Gaussian distributions (without increasing the lattice dimensions); the safety of using these in place of rounded continuous Gaussians follows from analyzing the corresponding Rényi divergence.
- **Efficient and dynamic generation of public parameters.** We generate the LWE public matrix $A$ for every key exchange from a small random seed, which has two benefits. Using a new $A$ with every connection avoids the use of global parameters that can lead to “all-for-the-price-of-one” precomputation attacks, such as the Logjam attack on finite-field Diffie–Hellman [1] (we note that these attacks have the potential to be even more devastating in the context of lattices – see [6]). Furthermore, in LWE this approach gives rise to significant bandwidth savings, since we avoid the transmission of the large matrix $A$ by instead transmitting the small seed. While this approach was already explored in the context of R-LWE in [6], in the context of LWE it becomes more challenging due to the large amount of randomness that is required. A significant step towards achieving high performance was optimizing this pseudorandom generation of $A$, since in our case it still consumes around 40% of the total computation time for each party. In order to target a large number of platforms and applications, we optimized this step for memory-access patterns; pseudorandom construction
of $A$ from a small random seed is done component-wise so that devices with memory constraints can generate blocks of $A$, use them, and discard them.

**Implementation.** We provide an open-source implementation of our protocol in the C language that provides protection against simple timing and cache-timing attacks. We have integrated our implementation, as well as publicly available implementations of several other post-quantum key exchange protocols [14] [6] [63] [24], to allow a direct comparison of post-quantum candidates in a common framework. We have also integrated our framework into OpenSSL, allowing it to be used with Apache for performance testing in real-world settings. In addition, we implemented and evaluated the hybrid version of each ciphersuite; this partners each of the proposed post-quantum key exchange primitives with ECDHE in order to mitigate the potential of (classical) cryptanalytic advances against the newer primitives.

The implementation of our protocol and parameter finding scripts are available at https://github.com/lwe-frodo/. Our framework for comparing post-quantum algorithms and integration into OpenSSL is available at https://github.com/open-quantum-safe/. Performance comparison and evaluation in TLS. We evaluate the performance of our protocol, along with the other post-quantum protocols mentioned above, in two ways:

- **Microbenchmarks.** These measure the performance of standalone cryptographic operations. Using LWE key exchange at the “recommended” security level, each party’s operations run in around 1.3ms. As expected, this is slower but still comparable to ideal lattice-based schemes (they have runtimes between 0.15ms—2.15ms per party), but it is significantly faster than the post-quantum software based on supersingular isogenies [24], and is comparable to traditional elliptic curve Diffie–Hellman using nistp256 (which takes about 0.7ms per party). Table 4 contains detailed comparisons.

- **Performance within HTTPS/TLS.** To measure the real-world impact of using generic lattices, we measure the connection time and server throughput when using post-quantum ciphersuites in TLS. This includes their performance under heavy sustained load while serving hundreds of connections every second. In this realistic scenario, the performance gap between ideal and generic lattices shrinks considerably. When ECDSA certificates are used, the time to establish a connection using our LWE key exchange ciphersuite is just 18.3ms (or 23ms when deployed in a hybrid mode with ECDHE). The handshake size is around 23.1 KiB. Compared to NewHope [6] (which is a very fast R-LWE based scheme), we see that using generic lattices in our case gives rise to just a factor 1.5× increase in latency (12.1ms or 16.5ms in hybrid mode), and a factor 4.7× increase in handshake size (4.9 KiB). Given that modern web pages load with latencies of hundreds of milliseconds and with sizes of a megabyte or more, the overhead of LWE key exchange is manageable in many cases and unnoticeable in many more. We measured connections per second at a variety of page sizes (see Table 5): at 100 KiB, the throughput of LWE is only 1.2× less than that of NewHope, and this drops to just 1.15× when each ciphersuite is deployed in a hybrid mode with ECDHE.

Ultimately, the size and speed of our protocol shows that we need not rely exclusively on ideal lattices to construct practical lattice-based key exchange: LWE-based key exchange is indeed a viable candidate for a practical quantum-resistant scheme.

2. RELATED WORK

Regev [58] introduced the LWE problem accompanied by a (quantum) reduction to certain worst-case lattice problems and demonstrated how to build a public-key encryption scheme based on the hardness of this problem. Peikert [52] gave the first classical reduction from LWE to standard lattice problems, however this reduction required moduli that were exponential in the size of the lattice dimension. Subsequently, Brakerski et al. [16] then showed that for moduli that are polynomial in the size of the lattice dimension, standard, worst-case problems in lattices are classically reducible to the LWE problem. Constructions based on LWE have led to a large variety of public-key cryptosystems [58] [52] [31] [18] [3] [2] [17]. Its ring analogue, R-LWE [45], yields more efficient constructions of public-key cryptosystems [53].

Based on the key encapsulation mechanism described by Peikert [53], Bos et al. [14] presented a Diffie-Hellman-like key exchange protocol whose security is based on the R-LWE problem. They demonstrated the feasibility of the protocol by integrating their implementation into the TLS protocol in OpenSSL, reporting a slight performance penalty over classically secure key exchange (using elliptic curves). Soon after, Alkim et al. [6] presented a range of improvements over the implementation in [14]. In particular, the in-depth security analysis performed in [6] paved the way for a superior set of parameters to those used in [14]: Alkim et al. showed that higher security could be achieved with a smaller modulus and a new, more efficiently sampleable noise distribution. Moreover, their improved error reconciliation and more aggressive implementation culminated in R-LWE key exchange software that was more than an order of magnitude faster than the software presented by Bos et al..

Previously, Ding et al. [37] had proposed an alternative instantiation of lattice-based key exchange that builds on Regev’s public-key cryptosystem [58], giving a Diffie-Hellman-like protocol from both LWE and R-LWE. The deduced shared key in their scheme is not uniform random, and they subsequently suggest to use an extractor; however, this reduces the effective length of the key. Our protocol is based on similar ideas to the LWE protocol in [37], but as in the R-LWE schemes in [14] and [6], we incorporate (and extend) Peikert’s reconciliation technique [53] and further modify the protocol to conserve bandwidth. Moreover, the analysis in [37] was performed for single-bit key exchange, and [37] did not include a concrete proposal for quantum-secure key exchange or an implementation of the scheme.

3. KEY EXCHANGE FROM LWE

**Notation.** If $\chi$ is a probability distribution over a set $S$, then $x \leftarrow \chi$ denotes sampling $x \in S$ according to $\chi$. If $S$ is a set, then $U(S)$ denotes the uniform distribution on $S$, and we denote sampling $x$ uniformly at random from $S$ either with $x \leftarrow U(S)$ or sometimes $x \leftarrow S$. Matrices are denoted by bold face capital letters. If $\chi$ is a distribution over a set $S$, then $X \overset{d}{\leftarrow} \chi(S^{n \times m})$ denotes generating an $n \times m$
Alice

\[ \text{seed}_A \xleftarrow{\$} U(\{0,1\}^n) \]

\[ A \leftarrow \text{Gen(}\text{seed}_A\text{)} \]

\[ S, E \xleftarrow{\$} \chi(\mathbb{Z}_q^{n \times m}) \]

\[ B \leftarrow A S + E \]

Bob

\[ \text{seed}_A, B \in \{0,1\}^n \times \mathbb{Z}_q^{n \times m} \]

\[ A \leftarrow \text{Gen(}\text{seed}_A\text{)} \]

\[ S', E' \xleftarrow{\$} \chi(\mathbb{Z}_q^{n \times m}) \]

\[ B' \leftarrow S' A + E' \]

\[ E'' \xleftarrow{\$} \chi(\mathbb{Z}_q^{n \times m}) \]

\[ V \leftarrow S' B + E'' \]

\[ C \leftarrow \langle V \rangle_{2^B} \]

\[ K \leftarrow \text{rec}(B', S', C) \]

Using the protocol in TLS. To integrate our LWE key exchange protocol into TLS, we have the server play the role of Alice and the client play the role of Bob. Alice’s first flow is sent as the ServerKeyExchange message, and Bob’s response is sent as the ClientKeyExchange message. Alice and Bob use \( K \) as the premaster secret.

Hybrid ciphersuites. As lattice-based cryptography is undergoing a period of intense development and scrutiny, a conservative approach towards deployment of lattice-based key exchange is to pair it with legacy schemes such as elliptic curve Diffie-Hellman (ECDH). Since the message flow of our proposed solution is identical to the existing ECDH(E) key exchange protocol, the two can be run concurrently as part of the same “hybrid” ciphersuite, with both outputs mixed in into the premaster secret via a KDF (as was done in [14]); see [36] for performance results.

3.2 A generalized reconciliation mechanism

In the protocol in Figure 1 Alice and Bob compute the secret matrices \( V \in \mathbb{Z}_q^{\pi \times n} \) and \( B' \in \mathbb{Z}_q^{\pi \times m} \). Each of the \( \pi m \) entries in \( V \) is approximately equal in \( \mathbb{Z}_q \) to the corresponding entry in \( B' \). To modify such an approximate key agreement into an exact key agreement, Bob sends \( C = \langle V \rangle_{2^B} \) to Alice which allows both parties to reconcile the same shared key \( K \). In this section we describe the reconciliation mechanism that allows such exact agreement to be achieved, i.e., we detail the function \( \langle \cdot \rangle_{2^B} \) used in line 6 of Bob’s flow, and the functions \( |\cdot|_{2^B} \) and \( \text{rec} \) used when Alice and Bob compute their respective shared keys. Our reconciliation mechanism is a generalized version of Peikert’s mechanism [53] that, for every approximate agreement in \( \mathbb{Z}_q \), allows the extraction of more (exact) shared bits. This increased extraction comes at the cost of an increased probability of failed reconciliation, but we argue that the resulting probability is still small enough for practical applications. Previous works that used Peikert’s mechanism [14, 45] have not needed to extract more than one bit per approximate agreement in \( \mathbb{Z}_q \), since the number of such agreements far exceeded the number of shared bits required to form a secure session key. As we discussed in [31], we need \( \pi m \cdot B \geq 256 \) for our desired level of quantum security, so a larger \( B \) (the number of bits extracted per approximate agreement) means we can achieve smaller \( \pi m \) and \( \pi \), which in turn means sending smaller LWE matrices.

We focus on the case when the modulus \( q \) is a power of 2; this can be generalized to an arbitrary modulus using techniques described in [53]. Let the number \( B \) of bits that we are aiming to extract from one coefficient in \( \mathbb{Z}_q \) be such that \( B < (\log_2 q) - 1 \). Let \( \tilde{B} = (\log_2 q) - B \). For any \( v \in \mathbb{Z}_q \), represented as a unique integer in \([0, q)\), we define the following functions. The \textit{rounding} function \(|\cdot|_{2^B}\) is defined as

\[ |v|_{2^B} : v \mapsto \left[ 2^{-\tilde{B}} v \right] \mod 2^B, \]

i.e., \(|v|_{2^B} \in [0, 2^B)\). Note that \(|v|_{2^B} \) outputs the \( B \) most significant bits of \((v + 2^{\tilde{B} - 1}) \mod q\). This means that \(|v|_{2^B}\) partitions \( \mathbb{Z}_q \) into \( 2^B \) intervals of integers with the same \( B \) most significant bits, up to a cyclic shift of the values that centers these intervals around 0. The \textit{cross-rounding} function \( \langle \cdot \rangle_{2^B} \) is defined as

\[ \langle v \rangle_{2^B} : v \mapsto \left[ 2^{-\tilde{B}+1} v \right] \mod 2^B, \]

matrix \( X \) by sampling each of its entries independently from \( S \) according to \( \chi \). If \( A \) is a probabilistic algorithm, \( y \xleftarrow{\$} A(x) \) denotes running \( A \) on input \( x \) with randomly chosen coins and assigning the output to \( y \).

The LWE problem is characterized by three parameters: the modulus \( q \), the dimension of the matrix \( n \), and the error distribution \( \chi \). Given an integer modulus \( q \), denote by \( \mathbb{Z}_q \) the ring of integers modulo \( q \). The distribution \( \chi \) is typically taken to be a rounded continuous or discrete Gaussian distribution over \( \mathbb{Z} \) with center zero and standard deviation \( \sigma \). However, as we discuss in detail in [35] alternative choices are possible and have advantages over such distributions. The concrete choice of the LWE parameters \( q, n, \sigma \) determines the security level of the protocol, and their selection is described in [43].

3.1 The new key exchange protocol

In this section we describe an unauthenticated key exchange protocol based on the LWE problem. The protocol is shown in Figure 1.

In the key exchange protocol in Figure 1 both Alice and Bob generate the same large matrix \( A \in \mathbb{Z}_q^{n \times n} \) that is combined with the LWE secrets to compute their public keys as instances of the LWE problem. Alice’s \( \pi \) LWE instances and Bob’s \( m \) LWE instances are combined to compute a secret matrix in \( \mathbb{Z}_q^{\pi \times m} \), where \( B \) uniform bits are extracted from each entry to form the session key \( K \). Thus, the dimensions \( \pi \) and \( m \) should be chosen such that \( K \) has (at least) the number of required bits for the target security level. For example, in targeting 128 bits of post-quantum security, it should be the case that \( \pi m \cdot B \geq 256 \). This condition ensures that we obtain a uniform 256-bit secret for the session key and even an exhaustive key search via Grover’s quantum algorithm would take \( 2^{128} \) operations. The protocol in Figure 1 allows for the ratio between \( \pi \) and \( m \) to be changed, in order to trade-off between Bob’s amount of uploaded data for Alice’s computational load. This could be useful in mobile devices, where energy efficiency of uploads is at most half that of downstream traffic [62, 35].

Figure 1: The LWE-based key exchange protocol with LWE parameters \( n, q, \chi \), and protocol specific parameters \( \pi, \pi, B \in \mathbb{Z} \). The matrix \( A \in \mathbb{Z}_q^{n \times n} \) is generated from \( \text{seed}_A \) via a pseudo-random function \( \text{Gen} \).
which divides \( \mathbb{Z}_q \) into two subsets according to their \((B + 1)\)th most significant bit, splitting up each interval above into two equally sized subintervals. Replacing \([\cdot]_B\) by \([\cdot]_{AB}\) would amount to simply taking the \(B\) most significant bits of \(v\). However, using \([\cdot]_{AB}\) ensures that the size of the error introduced by rounding is unbiased with respect to the sets induced by the cross-rounding function.

We are now in a position to define the reconciliation function, \( \text{rec} \), following \[53\]. On input of \( w \in \mathbb{Z}_q \) and \( b \in [0, 2^B) \), \( \text{rec}(w, b) \) outputs \([v]_{AB}\), where \( v \) is the closest element to \( w \) such that \([v]_{AB} = b\). The two claims below are the generalizations of the claims in \[53\], the proofs of which are analogous. They demonstrate that releasing \( [v]_{AB}\) does not reveal any information about \([v]_{AB}\), but it can serve as a hint for the two parties that are trying to agree on \( B\) bits, based on sufficiently close numbers \( w \) and \( v\). Claim \[3.2\] means that, if corresponding elements of \( B\) \( S\) and \( V\) are within \( 2^B - 2\) of one another, the key agreement in Fig. \[1\] is exact.

**Claim 3.1.** If \( v \in \mathbb{Z}_q \) is uniformly random, then \([v]_{AB}\) is uniformly random given \( [v]_{AB}\).

**Claim 3.2.** If \(|v - w| < 2^B - 2\), then \( \text{rec}(w, [v]_{AB}) = [v]_{AB}\).

*Round-and-truncate.* We observe that the lower-order bits of \( B\) and \( B'\) exchanged by the parties have vanishingly small influence on the negotiated key. To conserve bandwidth, a version of the protocol could be used in which entries of \( B\) and \( B'\) are rounded to multiples of \( 2^C\), and the lower \( C\) bits, which are now zeros, are truncated and never transmitted. Since \( B\) and \( B'\) are now transmitted with lower accuracy, this introduces another source of error in the reconciliation process. Although our implementation does not currently exploit this option, we note that if it were to be used, Bob should sample uniform noise and add it back to the lower order bits. This is to ensure that truncation does not affect the proof of security.

### 3.3 Error distributions

The key exchange protocol in Figure \[4\] is described in terms of an unspecified error distribution \( \chi \) over a set \( S\). We now describe the concrete choice of error distribution used in our implementation, which is an instantiation of *inversion sampling* that uses a precomputed table corresponding to a discrete cumulative density function (CDF) over a small interval. The four distributions we use are defined by the discrete probability density functions (PDFs) in Table \[1\]. We use the first distribution, \( D_1\), as an example to illustrate how the sampling routine works. Modifying the corresponding PDF into a CDF gives the table \( T = [43, 104, 124, 127]\). We then sample 8 bits uniformly at random; the first 7 bits correspond to a uniform random integer \( y \in [0, 127] \) which is used to return the smallest index \( \hat{x} \in [0, 3] \) such that \( y \leq T[\hat{x}]\), and the last (eighth) bit is used to determine the sign \( s \in \{-1, 1\} \) of the sampled value \( x = s \cdot \hat{x}\).

For each distribution in Table \[1\] performing inversion sampling can be done efficiently using at most seven precomputed look-up tables required for sampling any of the above distributions are at most 14 bytes each. Obtaining a single sample amount to accessing an element from one of these small look-up tables; this can be done in a memory and timing sidechannel-resistant manner by always scanning all

<table>
<thead>
<tr>
<th>dist.</th>
<th>bits req.</th>
<th>var.</th>
<th>Probability of</th>
<th>Renyi order</th>
<th>diverg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_1 )</td>
<td>8</td>
<td>1.25</td>
<td>0 ±1, ±2, ±3, ±4, ±5, ±6</td>
<td>20.0</td>
<td>1.002062</td>
</tr>
<tr>
<td>( D_2 )</td>
<td>12</td>
<td>1.00</td>
<td>16709989 983244-24</td>
<td>1</td>
<td>5.0</td>
</tr>
<tr>
<td>( D_3 )</td>
<td>12</td>
<td>1.75</td>
<td>1206919404-104</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>( D_4 )</td>
<td>16</td>
<td>1.75</td>
<td>1903041470164900-1659</td>
<td>24</td>
<td>20</td>
</tr>
</tbody>
</table>

**Table 1:** The discrete PDFs (and the number of bits required to obtain one sample from them) for the four noise distributions used in this work, which are approximations to rounded continuous Gaussian distributions of the given variance \( \sigma^2\); the closeness of these approximations is specified by the given Renyi divergence and its order.

### 4. SECURITY ASSESSMENT AND PARAMETER SELECTION

In this section, we explain our methodology to provide conservative security estimates against both classical and quantum attacks, and subsequently, we propose parameters for the protocol in the previous section. The methodology is similar to the one proposed in \[6\], with slight modifications that take into account the fact that some quasi-linear accelerations \[59\] \[15\] over sieving algorithms \[12\] \[32\] are not available without the ring structure. We restate this analysis for self-containment.

We remark that our methodology is significantly more conservative than what is usually used in the literature \[6\]. Our goal is not just to demonstrate feasibility, but to provide long-term and real-world security. To that end, we acknowledge that lattice cryptanalysis is far less mature than the cryptanalysis against schemes based on the hardness of factoring and computing discrete logarithms, for which the best-known attack can safely be considered as a best-possible attack.

#### 4.1 Methodology: the core-SVP hardness

Due to the (very) limited number \( m \) of LWE samples available to an attacker \((m = n + m)\) or \( m = n + \pi\), we are not concerned with BKW-like attacks \[40\], nor are we concerned with linearization attacks \[8\]. This essentially leaves us with two BKZ-style \[21\] attacks, which are usually referred to as *primal* and *dual* attacks; we review both below.

The BKZ algorithm with block-size \( b\) requires up to polynomially many calls to an SVP oracle in dimension \( b\). However, using some heuristics essentially allows the number of oracle calls to be decreased to a linear number \[20\]. To account for further improvement, we only count the cost of one such call to the SVP oracle, i.e., the core-SVP hardness.
Such precaution seems reasonable, especially in the case of the dual attack that involves running BKZ several times, in which case it is plausible that most of the lattice reduction cost may be amortized.

Even the concrete cost of a single SVP computation in dimension $b$ is problematic to evaluate. This is due to the fact that the numerically optimized pruned enumeration strategy does not yield a closed formula [30, 21]. Even with pruning, enumeration is asymptotically super-exponential, while sieving algorithms have exponential complexity $2^{O(b^c)}$ (where the constant $c$ in the exponent is well-understood). A sound and simple strategy is therefore to determine a lower bound on the concrete bit-security given by a particular set of parameters.

Classically, the best known constant is $c_C = \log_2 \sqrt{3}/2 \approx 0.292$, which is provided by the sieving algorithm of [12], and, in the context of quantum attacks, the best known constant is $c_Q = \log_2 \sqrt{13}/9 \approx 0.265$ (see §14.2.10 in [42]). Since all variants of the sieving algorithm require a list of $2^{b\sqrt{b}}$ vectors to be built, it also seems plausible that $c_P = 0.2075$ can serve as a worst-case lower bound. (Here the subscripts C, Q and P differentiate between the sieving constants used to obtain "classical", "quantum" and "paranoid" estimates on the concrete bit-security given by a particular set of parameters.)

### 4.2 Primal attack

The primal attack constructs a unique-SVP instance from the LWE problem and solves it using BKZ. We examine how large the block dimension $b$ is required to be for BKZ to find the unique solution. Given the matrix LWE instance $(A, b = As + e) \in \mathbb{Z}_q^{m \times n} \times \mathbb{Z}_q^n$, one builds the lattice $\Lambda = \{x \in \mathbb{Z}_q^{m+n+1} : (A^T I_n) x = 0 \bmod q\}$ of dimension $d = m+n+1$ and volume $q^m$. The vector $v = (s, e, 1) \in \Lambda$ is a unique-SVP solution of norm $\lambda \approx \sqrt{m} + m$, where $\zeta$ is the standard deviation of the error distribution used to sample $e$. In our case, the number of samples used, $m$, may be chosen between $0$ and $n + \bar{m}$ (or $n + \bar{m}$), and we numerically optimize this choice.

Using the typical models of BKZ (under the geometric series assumption and the Gaussian heuristic [20, 5]), one concludes that the primal attack is successful if and only if $\zeta \sqrt{\delta} \leq \delta^{2b-d-1} \cdot q^{m/d}$, where $\delta = ((\pi b)^{1/2} - 1/2\pi e)^{1/2^{b-1}}$.

### 4.3 Dual attack

In the dual attack, one searches for a short vector $(v, w)$ in the dual lattice $\Lambda = \{(x, y) \in \mathbb{Z}_q^m \times \mathbb{Z}_q^n : A^T x = y \bmod q\}$, with the aim of using it as a distinguisher for the decision-LWE problem. The BKZ algorithm with block-size $b$ will output such a vector of length $\ell = \delta^{d-1} q^{n/d}$.

Having found $(x, y) \in \Lambda$ of length $\ell$, an attacker computes $z = v \cdot b = v^T As + v^T e = w^T s + v^T e \bmod q$. If $(A, b)$ is indeed an LWE sample, then $z$ is distributed as a Gaussian, centered around 0 and of standard deviation $\zeta$, otherwise $z$ is distributed uniformly modulo $q$. The maximal variation distance between these two distributions is bounded by $\epsilon \approx 4 \exp(-2\pi^2 \tau^2)$, where $\tau = \zeta / q$. Thus, given such a vector of length $\ell$, the attacker may distinguish LWE samples from random with advantage at most $\epsilon$.

It is important to note that a small distinguishing advantage $\epsilon$ does not provide appreciable help to an adversary that is attacking a key exchange protocol: since the agreed key is to be used to derive a symmetric cipher key, any advantage below $1/2$ does not significantly decrease the search space in a brute force search for the symmetric cipher key. (Recall that the reconciled key is processed with a random oracle before it is used for any other purposes.)

We therefore require an attacker to amplify his success probability by finding approximately $1/\epsilon^2$ such short vectors. Since the sieve algorithms provide $2^{0.2075 b}$ vectors, the attack must be repeated at least $R = \max(1, 1/(2^{0.2075 b} \epsilon^2))$ times. We again stress that we are erring on the conservative side, since the other vectors that are output by the sieving algorithm are typically a little larger than the shortest one.

### 4.4 Proposed parameters

Our proposed parameters are summarized in Table 2 and their security detailed in Table 3.

| Challenge parameters. | We provide a set of challenge parameters as a target that should be reasonably accessible within the current cryptanalytic state-of-the-art. Attacking these parameters may even be feasible without subtle optimizations such as pruned enumeration. We do not provide hardness claims for those parameters because the required BKZ block-size for breaking them is far below 200, and since our analysis only considers sieving algorithms, it is not valid in that range.

| Classical parameters. | We also propose a classical parameter set that provides 128-bits of security against classical attacks. We do not recommend these parameters in practice since their two extremes. To determine the overall failure rate of our protocol, we combine this relationship with the probability distribution of the distance. In the continuous case, it is easy to check that the distribution of this distance has variance $\sigma^2 = 2mc^2 + c^2$, where $c^2$ is the variance of the continuous Gaussian distribution. However, using more computationally-intensive but tighter analysis, we can compute the distribution of the distance corresponding to our... |
The best classical attack takes $2^b$ messages.

For the specific parameter set, e.g. the Classical parameter set, which reduces the total amount of communication and computation by approximately a factor of 2.

### 5. PROOF OF SECURITY

The security of our key exchange protocol can be reduced to the learning with errors problem. Our proof uses a variant with short secrets and matrices (instead of vectors) that is equivalent to the original LWE problem.

### 5.1 The LWE problem and variants

**Definition 5.1** (Decision LWE problem). Let $n$ and $q$ be positive integers. Let $\chi$ be a distribution over $\mathbb{Z}$. Let $s \leftarrow \mathcal{U}(\mathbb{Z}_q^n)$. Define the following two oracles:

- $O_{\chi,s}: a \leftarrow \mathcal{U}(\mathbb{Z}_q^n), e \leftarrow \chi(\mathbb{Z}^n_q); \text{return (a, as + e)}$.
- $U: a \leftarrow \mathcal{U}(\mathbb{Z}_q^n), u \leftarrow \mathcal{U}(\mathbb{Z}^n_q); \text{return (a, u)}$.

The decision LWE problem for $(n, q, \chi)$ is to distinguish $O_{\chi,s}$ from $U$. In particular, for algorithm $A$, define the advantage $\text{Adv}_{n,q,\chi}^\text{dlwe}(A) = |\Pr(s \leftarrow \mathbb{Z}_q^n : A_{\chi,s}() = 1) - \Pr(A_U() = 1)|$.

The proof of security of our key exchange protocol relies on a variant of the LWE problem stated below, where secrets $s$ are drawn from $\chi$. It was shown by Applebaum et al. [7] that the short secret variant has a tight reduction to the decision LWE problem.

### 5.2 Security of the key exchange protocol

To prove security of the key exchange protocol, consider an LWE key-exchange adversary that tries to distinguish the session key $K$ from a uniformly random key $K'$ given the transcript of the key exchange protocol. Formally, we define the advantage of such an adversary $A$ as:

$$\text{Adv}_{n,\pi,\bar{m},B,q,\chi}^\text{dlwe}(A) = |\Pr[A(\bar{A}, B, B', C, K') = 1] - \Pr[A(\bar{A}, B, B', C, K) = 1]|,$$

where $A, B, B', C$, and $K$ are as in Figure 4 with LWE parameters $n$, $q$, and $\chi$, additional parameters $\pi$, $\bar{m}$, $B$, and $K'$ is a uniform bit string of length $\pi \cdot m \cdot B$.

The following theorem implies that under the decision LWE assumption (with short secrets) for parameters $n$, $q$, and $\alpha$, efficient adversaries have negligible advantage against the key exchange protocol of [3].

**Theorem 5.1.** Let $n, \pi, \bar{m}, B$, and $q$ be positive integers, and let $\chi$ be a distribution on $\mathbb{Z}_q$. If $G$ is a secure pseudorandom function and the decision LWE problem is hard for $(n, q, \chi)$, then the key exchange protocol in Figure 4 yields keys indistinguishable from random. More precisely, $\text{Adv}_{n,\pi,\bar{m},B,q,\chi}^\text{dlwe}(A) \leq \text{Adv}_{\text{Gen}}^\text{dlwe}(B_0) + \pi \cdot \text{Adv}_{n,q,\chi}^\text{dlwe}(A \circ B_1) + \bar{m} \cdot \text{Adv}_{n,q,\chi}^\text{dlwe}(A \circ B_2)$ where $B_1$ and $B_2$ are the reduction algorithms given in Figure 7 and $B_0$ is implicit in the proof. The runtimes of $B_0, A \circ B_1$, and $A \circ B_2$ are essentially the same as that of $A$.

**Proof.** The proof closely follows Peikert’s proof of IND-CPA security of the R-LWE KEM; see Lemma 4.1 of [53], and Bos et al.’s proof of R-LWE DH key exchange [14]. It proceeds by the sequence of games shown in Figure 8. Let $S_i$ be the event that the adversary guesses the bit $b_i$ in Game $i$.
Game 0. This is the original game, where the messages are generated honestly as in Figure 1. We want to bound $\Pr(S_0) - \Pr(S_1) \leq \Adv_{\Gen,X}^{\textup{ddh-learn}}(B_0)$. 

Difference between Game 0 and Game 1. An adversary that can distinguish these two games immediately leads to a distinguisher $B_0$ for $\Gen$: $|\Pr(S_0) - \Pr(S_1)| \leq \Adv_{\Gen,X}^{\textup{ddh-learn}}(B_0)$. 

Game 2. In this game, Alice’s ephemeral public key $B'$ is generated uniformly at random, rather than being generated from a sampler for the $(n,\tau)$- matrix decision LWE problem. Note that in Game 2, the (remaining) LWE pairs are: $(A, B')$ and $(B, V)$ both with secret $S'$. 

Difference between Game 1 and Game 2. In Game 1, $(A, B)$ is a sample from $O_{K, S}$. In Game 2, $(A, B')$ is a sample from $U(\mathbb{Z}_q^{n \times \tau}) \times U(\mathbb{Z}_q^{n \times \tau})$. Under the decision LWE assumption for $(n, \chi)$, these two distributions are indistinguishable with a factor of $\pi$. 

More explicitly, let $B_1$ be the algorithm shown in Figure 3 which takes as input a pair $(A, B)$. When $(A, B)$ is a sample from $O_{K, S}$ where $S \overset{\$}{\leftarrow} \chi(\mathbb{Z}_q^{n \times \tau})$, then the output of $B_1$ is distributed exactly as in Game 1. When $(A, B)$ is a sample from $U(\mathbb{Z}_q^{n \times \tau}) \times U(\mathbb{Z}_q^{n \times \tau})$, then the output of $B_1$ is distributed exactly as in Game 2. Thus, if $A$ can distinguish Game 1 from Game 2, then $A \circ B_1$ can distinguish samples from $O_{K, S}$ from samples from $U(\mathbb{Z}_q^{n \times \tau}) \times U(\mathbb{Z}_q^{n \times \tau})$. Thus, $|\Pr(S_0) - \Pr(S_1)| \leq \Adv_{\Gen}^{\textup{ddh-learn}}(A \circ B_1)$. 

Game 3. Game 3 is a simple rewrite of Game 2. Bob’s ephemeral public key $B'$ and shared secret $V$ are simultaneously generated from $S'$ rather than sequentially. In Game 3, the single LWE pair $\left(\left[A', B'\right], \left|V\right|\right)$ with secret $S^T$ is an instance of the $(n + \tau, m)$-matrix decision LWE problem. 

Difference between Game 2 and Game 3. Game 3 is just a conceptual rewrite of Game 2, we have that $\Pr(S_0) = \Pr(S_1)$. 

Game 4. In Game 4, there are no LWE pairs. Bob’s ephemeral public key $B'$ and shared secret $V$ are generated uniformly at random, rather than simultaneously from $S'$. 

Difference between Game 3 and Game 4. In Game 3, $\left(\left[A', B'\right], \left|V\right|\right)$ is the transpose of a sample from oracle $O_{K, S'}$ in the $(n + \tau, m)$-matrix decision LWE problem. In Game 4, $\left(\left[A', B'\right], \left|V\right|\right)$ is the transpose of a sample from $U(\mathbb{Z}_q^{n \times \tau}) \times U(\mathbb{Z}_q^{n \times \tau})$. Under the decision LWE assumption for $(n, \chi)$, these two distributions are indistinguishable with a factor of $\pi$. 

More explicitly, let $B_2$ be the algorithm shown in Figure 3 which takes as input $(Y, Z) \in \mathbb{Z}_q^{n \times (n + \tau)} \times \mathbb{Z}_q^{n \times \tau}$. When $(Y, Z)$ is a sample from $O_{K, S'}$ in the $(n + \tau, m)$-matrix decision LWE problem, where $S' \overset{\$}{\leftarrow} \chi(\mathbb{Z}_q^{n \times \tau})$, then the output of $B_2$ is distributed exactly as in Game 3. When $(Y, Z)$ is a sample from $U(\mathbb{Z}_q^{n \times \tau}) \times U(\mathbb{Z}_q^{n \times \tau})$, the output of $B_2$ is distributed exactly as in Game 4. If $A$ can distinguish Game 3 from Game 4, then $A \circ B_2$ can distinguish samples from $O_{K, S'}$ from samples from $U(\mathbb{Z}_q^{n \times \tau}) \times U(\mathbb{Z}_q^{n \times \tau})$. Thus, $|\Pr(S_0) - \Pr(S_1)| \leq \Adv_{\Gen}^{\textup{ddh-learn}}(A \circ B_2)$. 

Analysis of Game 4. In Game 4, the adversary is asked to guess $b'$ and thereby distinguish between $K$ and $K'$. In Game 4, $K'$ is clearly generated uniformly at random from $\{0, 1\}^{\pi \times \tau}$. $K$ is generated from rounding $V$, and $V$ is uniform, so $K$ is too. The adversary is also given $C$, but by Claim 3, we have that, for uniform $V$, $K = |V|_2^\tau$ remains
uniform even given $C = \langle V \rangle_{2^6}$. Thus, the adversary has no information about $b^*$, and hence $Pr(S_l) = 1/2$.

### 5.3 Security when used in TLS

The accepted model for security of secure channel protocols such as TLS is Jager et al.’s authenticated and confidential channel establishment (ACCE) model \[35\]. Bos et al. \[14\] show that a TLS ciphersuite that uses signatures for authentication and a key exchange mechanism that satisfies the ddh-like security property (5.2) achieves the ACCE security notion. Since our Theorem 6.1 shows that the LWE-based protocol has ddh-like-security, we immediately (modulo a small change of notation) inherit ACCE security of the resulting signed-DH-like ciphersuite using LWE key exchange. Bos et al. note that their result requires that a change be made in the TLS protocol: the server’s signature must be moved later in the handshake and must include the full transcript. This is to be able to rely on a plain decisional assumption, rather than an “oracle” assumption (PRF-ODH) that was required in Jager et al.’s proof; we inherit this requirement as well.

Alternatively, one could leave the signature in place and use an IND-CCCA-secure key encapsulation mechanism following Krawczyk et al. \[11\], constructed via standard constructions such as \[29\], albeit at the expense of increasing the number of bits transmitted.

### 6. IMPLEMENTATION

In this section we discuss two aspects of our implementation: representing matrix elements and generating the public matrix $A$. Matrix operations in our implementation use the straightforward approach, i.e., the product of an $n \times m$ matrix and an $m \times p$ matrix incurs $nmp$ multiplications and $O(nmp)$ additions. We make this choice because our matrix dimensions are not large enough for the asymptotically faster methods \[63, 23\] to offer a worthwhile trade-off (see \[25\]).

#### Representing matrix elements.

All parameter sets in Table 2 have $q = 2^e < 2^{16}$. To facilitate an efficient implementation, we use a redundant representation: we represent elements as from $\mathbb{Z}_{q^4} = \mathbb{Z}/(2^{16} - 4q)\mathbb{Z}$ instead of $\mathbb{Z}_q$. This has the advantage that, when elements are stored in 16-bit datatypes, all arithmetic is performed modulo $2^{16}$ for free (implicitly). Converting from this redundant representation to elements in $\mathbb{Z}_q$ is as simple as ignoring the 16 – $x$ most significant bits, which amounts to a single bitwise AND instruction.

The exchanged matrices $B$, $B'$, and $C$ are packed down to their optimal representation. The computational overhead due to packing and unpacking is outweighed by savings in communication complexity.

#### Generating the matrix $A$.

The matrix $A \in \mathbb{Z}_{q^4}^{n \times n}$ is a public parameter used by both parties in the key exchange. It could be taken as a fixed system parameter, saving this communication effort between parties. This approach was taken in \[14\] in the ring setting, where the polynomial $a$ was fixed system wide. However, as discussed in Section 3 of \[6\], such a choice raises questions about possible backdoors and all-for-the-price-of-one attacks. Therefore, the scheme in \[6\] generates a fresh polynomial $a$ from a uniformly random seed for every instantiation of the key exchange.

Here we adopt a similar approach and propose that the server choose a fresh seed for every new key exchange instance, and that $A$ be generated pseudorandomly from this seed. The seed is sent to the client in the ServerKeyExchange message (see Figure 1) and allows the client to pseudorandomly generate the same $A$. The matrix $A$ is generated from a 16-byte seed using AES128 in the ECB mode. Substituting a randomly sampled matrix with the one derived in this manner can be justified by appealing to the random oracle heuristic; exploring its applicability in our context against a quantum adversary is an interesting open question.

Depending on the architecture, particularly for memory-constrained embedded devices, storing the matrix $A$ in its entirety and loading it into memory for matrix multiplication might be too costly. In such scenarios, we propose to generate, use and discard parts of the matrix $A$ on-the-fly.

To facilitate on-the-fly matrix generation by both the client and the server, we pursue the following approach. The matrix $A$ is derived by applying AES128-ECB to pre-filled matrix rows with guaranteed uniqueness.\(^2\) The matrix thus defined can be computed either by rows, one row at a time, or by columns, 8 columns at a time, depending on whether $A$ is multiplied by $S$ on the right (client-side) or on the left (server-side). The cost of this pseudorandom generation can be a significant burden and can be amortized by relaxing the requirement for a fresh $A$ in every new key exchange (e.g., by allowing $A$ to be cached on the server for a limited time).

### 7. EVALUATION

We evaluate the performance of LWE-based key exchange on the following characteristics: 1) speed of standalone cryptographic operations; 2) speed of HTTPS connections; 3) communication costs. Our LWE implementation is written in C and the implementation is as described in \[6\]. In this section, we report the results of our implementation of LWE-based key exchange and compare our results with several other post-quantum primitives.

We selected the following post-quantum algorithms with publicly available implementations:

- BCNS R-LWE key exchange, C implementation \[14\];
- NewHope R-LWE key exchange, C implementation \[8\];
- NTRU public key encryption key transport using parameter set EES473EP1, C implementation \[63\]; and
- SIDH (supersingular isogeny Diffie–Hellman) key exchange, C implementation \[24\].

The implementation of Bernstein et al.’s “McBits” high-speed code-based cryptosystem \[13\] was not publicly available at the time of writing. We also included OpenSSL’s implementation of ECDH (on the NISTp256 curve) and RSA with a 3072-bit modulus for comparisons against widely deployed non-post-quantum key exchange at the 128-bit classical security level. The compiler is gcc version 4.8.4 and software is compiled for the x86_64 architecture.

We integrated our C implementation of our LWE protocol, as well as all the other implementations, into OpenSSL v1.0.1f following Stebila’s OpenSSL v1.0.1f fork for the R-LWE experiments in \[14\], to allow for comparison of all algorithms in the same context and to facilitate the HTTPS performance evaluation.\(^3\) Our implementation includes a com-

---

\(^2\)Our implementation uses the AES-NI instruction set where supported.

\(^3\)While 5.3 notes that moving the server’s signature to later in the handshake is required to achieve provable security of the full LWE-based ciphersuite under the decision LWE assumption, our
mon API for post-quantum key exchange methods, making it easier to add and compare new candidates as they become available. Finally, our implementation and evaluation uses AES256 GCM authenticated encryption and SHA384 to ensure post-quantum security for symmetric key operations.

We use a common hardware platform for all measurements. Standalone cryptographic operations were measured on the same computer, which acted as the server for measuring speed of TLS connections. This is an n1-standard-4 Google Cloud VM instance\(^4\) with 15 GB of memory, which has 4 virtual CPUs; in our instance, each of these was implemented as a single hardware hyper-thread on a 2.6GHz Intel Xeon E5 (Sandy Bridge). Clients for measuring the throughput of TLS connections were run on an n1-standard-32 Google Cloud VM instance with 120 GB of memory and 32 virtual CPUs, which ensured that we could fully load the server. Although some implementations included optimizations using the AVX2 instruction set, the cloud servers we used did not support AVX2; while vectorization of matrix operations using AVX2 will improve performance, our results may be more indicative of performance on widespread non-AVX2 hardware.

### 7.1 Standalone cryptographic operations

Table 4 reports the performance of standalone post-quantum cryptographic operations, as well as standard cryptographic operations for comparison. We obtained these results by integrating their implementations into OpenSSL as discussed above, and in particular using the openssl speed command. In the table, Alice0 denotes Alice’s initial ephemeral key and message generation (e.g., Alice’s operations up to and including her message transmission in Figure 1); Bob denotes Bob’s full execution, including ephemeral key and message generation and key derivation (all of Bob’s operations in Figure 1); and Alice1 denotes Alice’s final key derivation (e.g., Alice’s operations upon receiving Bob’s message in Figure 1).

**Discussion.** Microbenchmarks of LWE-based key exchange are very promising when compared with other protocols. At approximately 1.3ms each, the runtime of Frodo is orders of magnitude faster than SIDH, faster than NTRU (on the server side), about 1.8x slower than ECDH nistp256, and about 9x slower than R-LWE (NewHope). While there is a large gap between the microbenchmark performance of LWE against R-LWE, we recall that LWE has the increased confidence of using generic rather than ideal lattices, and we will observe in the next section that this gap is significantly narrowed when measured in an application that uses TLS. Additionally, we observe that the runtime of our paranoid parameters requires a slight (between 14% and 19%) overhead in bandwidth and compute in return for much higher security.

Our modifications to openssl speed allowed us to benchmark the impact of freshly generating the matrix A for each connection versus using a precomputed A. For our recommended parameters, the time to generate A was 0.54ms ± 0.01, which is 42% of Alice’s total runtime (similarly for Bob). This is substantial, and means that optimization of this step is valuable (see §6 for more details); our original naive implementation was several times slower. This also highlights the benefit that could come from caching A: a busy TLS server might reuse the same A for an hour or a day while still limiting the risk of global all-for-the-price-of-one attacks.

**Embedded system.** We also performed micro-benchmarks on the low-cost, community-supported BeagleBone Black development platform which has a AM335x 1GHz ARM Cortex-A8. We measured the standalone functionality using the GNU/Linux perf infrastructure with gcc version 4.6.3; measurements are averages over thousands of runs. For the Frodo recommended parameters, Alice0 takes 77.5M cycles; Bob takes 80.22M cycles; and Alice1 takes 1.09M cycles.

### 7.2 HTTPS connections

Table 5 reports the performance of a TLS-protected web server using the various key exchange mechanisms, following the methodology of Gupta et al.\(^3\) that was used by Bos et al.\(^1\). The methodology reports performance with HTTP payloads of 4 different sizes (1 byte, 1 KiB, 10 KiB, and 100 KiB), to demonstrate the diminishing importance of key exchange performance on TLS performance as payload size increases in realistic workloads. For each key exchange primitive, we measured performance using RSA signatures (with 3072-bit keys) and ECDSA signatures (with nistp256 keys): as noted in the introduction, future quantum adversaries would be able to retroactively decipher information protected by a non-quantum-safe key exchange, but not be able to retroactively impersonate parties authenticated using non-quantum-safe digital signatures. (We omit TLS-level benchmarks of Frodo-Paranoid since its microbenchmarks are similar to Frodo-Recommended, and similarly omit SIDH since its performance is on the order of a handful of connections per second.)

For the server, we used Apache httpd version 2.4.20, compiled against our customized OpenSSL library above, and using the prefork module for multi-threading. The client and server were connected over the standard data center network with a ping time of 0.62ms ± 0.02ms which did not achieve saturation during any of our experiments.

**Connections/second (throughput).** Multiple client connections were generated by using the http_load too\(^5\) version 09Mar2016, which is a multi-threaded tool that can generate multiple http or https fetches in parallel. We found that we could sometimes achieve better performance by running multiple http_load processes with fewer threads in each process. As we increase loads beyond the point at which we report throughput, there is very little increase in connections/sec but a significantly larger increase in mean connection time. At this point, we posit that the bottleneck in serving TLS requests shifts to the Apache process scheduling and request buffer management rather than ciphersuite performance. Our client machine had sufficient power to ensure our server computer reached at least 90% CPU usage throughout the test. Results reported are the mean of five executions, each of which was run for thirty seconds.

---

4. [https://cloud.google.com/compute/docs/machine-types#standard_machine_types](https://cloud.google.com/compute/docs/machine-types#standard_machine_types)

Table 4: Performance of standalone cryptographic operations, showing mean runtime (± standard deviation) in milliseconds of standalone cryptographic operations, communication sizes (public key / messages) in bytes, and claimed security level in bits. In TLS, the RSA public key is already included in the server’s certificate message, so RSA key transport imposes no additional communication from server to client.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Alice0 (ms)</th>
<th>Bob (ms)</th>
<th>Alice1 (ms)</th>
<th>Communication (bytes)</th>
<th>Claimed security level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A→B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A→B</td>
<td>w/o</td>
<td>w/ load</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSA 3072-bit</td>
<td></td>
<td>0.09</td>
<td>4.49</td>
<td>387 / 0°</td>
<td>128</td>
</tr>
<tr>
<td>ECDH nistp256</td>
<td>0.366 ± 0.006</td>
<td>0.698 ± 0.014</td>
<td>0.331 ± 0.01</td>
<td>32</td>
<td>128</td>
</tr>
<tr>
<td>BCNS</td>
<td>1.01 ± 0.006</td>
<td>1.59 ± 0.007</td>
<td>0.174 ± 0.001</td>
<td>4,096</td>
<td>163</td>
</tr>
<tr>
<td>NewHope</td>
<td>0.112 ± 0.003</td>
<td>0.164 ± 0.001</td>
<td>0.034 ± 0.001</td>
<td>1,824</td>
<td>229</td>
</tr>
<tr>
<td>NTRU EES743EP1</td>
<td>2.00 ± 0.055</td>
<td>0.281 ± 0.047</td>
<td>0.148 ± 0.005</td>
<td>1,027</td>
<td>256</td>
</tr>
<tr>
<td>SIDH</td>
<td>135 ± 1.91</td>
<td>464 ± 6.74</td>
<td>301 ± 0.97</td>
<td>564</td>
<td>192</td>
</tr>
<tr>
<td>Frodo Recomm.</td>
<td>1.13 ± 0.09</td>
<td>1.34 ± 0.02</td>
<td>0.13 ± 0.01</td>
<td>11,377</td>
<td>156</td>
</tr>
<tr>
<td>Frodo Paranoid</td>
<td>1.25 ± 0.02</td>
<td>1.64 ± 0.02</td>
<td>0.15 ± 0.01</td>
<td>13,057</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 KiB</td>
<td>12,976</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 KiB</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 KiB</td>
<td>10 KiB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 KiB</td>
<td>100 KiB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Performance of Apache HTTP server using post-quantum key exchange, measured in connections per second, connection time in milliseconds, and handshake size in bytes. Connection time under load is reported as the 75th percentile value across 20 measurements. Values measured across experiments were within 2× the reported value. All TLS ciphersuites used AES256-GCM authenticated encryption with SHA384 in the MAC and KDF. Note that different key exchange methods are at different security levels; see Table 4 for details.

<table>
<thead>
<tr>
<th>Ciphersuite</th>
<th>Key exchange</th>
<th>Sig.</th>
<th>Connection time (ms)</th>
<th>Handshake size (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>w/o load</td>
<td>w/ load</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1B 1KiB 10KiB 100KiB</td>
<td></td>
</tr>
<tr>
<td>ECDHE</td>
<td>ECDSA</td>
<td>1187</td>
<td>1107 ± 103 1088 ± 103</td>
<td>961 ± 68</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>814</td>
<td>810 ± 5.2     790 ± 5.9</td>
<td>710 ± 12</td>
</tr>
<tr>
<td>BCNS</td>
<td>ECDSA</td>
<td>922</td>
<td>907 ± 83      893 ± 16</td>
<td>819 ± 83</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>722</td>
<td>710 ± 9.7     716 ± 2.7</td>
<td>638 ± 3.7</td>
</tr>
<tr>
<td>NewHope</td>
<td>ECDSA</td>
<td>1616</td>
<td>1413 ± 39     1351 ± 148</td>
<td>985 ± 77</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>983</td>
<td>970 ± 67      949 ± 36</td>
<td>771 ± 41</td>
</tr>
<tr>
<td>NTRU</td>
<td>ECDSA</td>
<td>725</td>
<td>723 ± 8.8     708 ± 17</td>
<td>612 ± 31</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>553</td>
<td>534 ± 7.3     548 ± 1.4</td>
<td>512 ± 4.9</td>
</tr>
<tr>
<td>Frodo Recomm.</td>
<td>ECDSA</td>
<td>923</td>
<td>892 ± 59      878 ± 70</td>
<td>843 ± 68</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>703</td>
<td>700 ± 6.2     698 ± 1.8</td>
<td>635 ± 16</td>
</tr>
<tr>
<td>Frodo Recomm.</td>
<td>Hybrid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18.3</td>
<td>18.1 ± 0.5    20.7 ± 0.6</td>
<td>31.5 ± 2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.7</td>
<td>23.0 ± 3.4    24.5 ± 3.9</td>
<td>35.2 ± 4.2</td>
</tr>
<tr>
<td>RES+ECDSA</td>
<td>ECDSA</td>
<td>736</td>
<td>735 ± 3.8     728 ± 6.4</td>
<td>664 ± 7.3</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>567</td>
<td>567 ± 1.0     559 ± 2.4</td>
<td>503 ± 3.2</td>
</tr>
<tr>
<td>NewHope+ECDSA</td>
<td>ECDSA</td>
<td>1095</td>
<td>1075 ± 58     1017 ± 16</td>
<td>776 ± 11</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>776</td>
<td>775 ± 3.7     765 ± 2.4</td>
<td>686 ± 8.6</td>
</tr>
<tr>
<td>NTRU+ECDSA</td>
<td>ECDSA</td>
<td>590</td>
<td>589 ± 1.0     578 ± 3.2</td>
<td>539 ± 5.4</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>468</td>
<td>467 ± 0.4     456 ± 3.6</td>
<td>424 ± 24</td>
</tr>
<tr>
<td>Frodo+ECDSA</td>
<td>ECDSA</td>
<td>735</td>
<td>716 ± 1.5     701 ± 20</td>
<td>667 ± 7</td>
</tr>
<tr>
<td></td>
<td>RSA</td>
<td>552</td>
<td>551 ± 1.9     544 ± 1.6</td>
<td>516 ± 1.8</td>
</tr>
</tbody>
</table>

Connection time and handshake size. Wireshark was used to measure the time for the client to establish a TLS connection on an otherwise unloaded server, measured from when the client opens the TCP connection to when the client starts to receive the first packet of application data, and the size in bytes of the TLS handshake as observed. Connections were initiated using the `openssl s_client` command.

We considered two scenarios in this measurement. In scenario 1, client-server connections were initiated when the server was under no load. The results reported under the column “w/o load” in Table 5 were the average of ten executions. In scenario 2, client-server connections were initiated when the server was under approximately 70% CPU load (also due to concurrent TLS connections). Connection time naturally had a larger variance under load and we report the 75th percentile results over twenty experiments under the column “w/ load” in Table 5.

Discussion. Due to HTTPS loads imposing stress on many different parts of the system, slower compute time of LWE compared to R-LWE (NewHope) has a much lower impact.
on overall performance. This is especially true when considering hybrid ciphersuites. Our results are not necessarily generalizable to all applications, but do provide a good indication of how TLS-based applications with a variety of load profiles are affected.

Since establishing a TLS connection involves several round trips as well as other cryptographic operations, the connection times (for a server without load) for all ciphersuites we tested all range between 12ms and 21ms, which are an order of magnitude smaller than the load times of typical web pages, so perceived client load times will be minimally impacted in many applications. Even under reasonable load, the increase in connection times for LWE follows similar increases across the different ciphersuites profiled.

Our key exchange does affect server throughput, but the gap between LWE and R-LWE is smaller and (naturally) decreases as page size increases. For 1 KiB pages, LWE can overtake R-LWE (NewHope), but this gap narrows to 1.2× fewer for 100 KiB pages, and outperforms NTRU. When measuring server throughput against other cryptographic operations, the overhead of LWE is further reduced to just about 15% for 100 KiB pages when compared to NewHope.

While LWE does have an impact on both TLS client and server performance and on the handshake size, the complex nature of TLS connections and various system bottlenecks when serving webpages mute the perceived gap seen between LWE and R-LWE in the microbenchmarks. For many applications, such as web browsing and secure data transfers, the LWE latency and communication size are small compared to typical application profiles.

Going forward, if we choose to deploy a post-quantum ciphersuite that combines LWE and ECDHE instead of a (non-post-quantum) ECDHE ciphersuite, the additional overhead in serving typical webpages between 10 KiB and 100 KiB will only decrease server throughput by less than a factor of two. This might be a small price to pay for long-term post-quantum security based on generic lattices without any ring structure, in order to avoid the possibility of attacks that this structure gives rise to.

Acknowledgments
JB and LD were supported in part by the Commission of the European Communities through the Horizon 2020 program under project number 645622 (PQCRYPTO). DS was supported in part by Australian Research Council (ARC) Discovery Project grant DP130104304 and a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. The authors would like to thank Adam Langley, Eric Grosse, and Úlfar Erlingsson for their inputs. A large part of this work was done when VN was an intern with the Google Security and Privacy Research team.

8. REFERENCES