Matthew Townsend

CS 522 Network

CRYSTALS-Dilithium Algorithm Specifications and Supporting Documentation (Version 3.1)

The design is based on the "Fiat-Shamir with Aborts" approach. A critical design choice was the exclusive use of uniform sampling, which avoids the complexities and side-channel vulnerabilities associated with generating samples from the discrete Gaussian distribution, thereby enhancing secure implementation.

Dilithium introduces an optimization over previous schemes by shrinking the bit-representation size of the public key component t, approximately halving its size. This reduction is achieved by excluding some low-order bits of t from the public key. The verifier recovers the necessary high-order bits of Az−ct using small "hints" (h) included in the signature, which essentially encode the carries caused by the missing bits.

The specification offers the option for the scheme to be either randomized or deterministic. The deterministic version is suggested as the default, as recent results showed tighter security reductions in the Quantum Random Oracle Model (QROM) when the scheme is deterministic.

Implementation optimizations, such as the use of AVX2 instructions, dramatically improve efficiency compared to the reference code. For the NIST Level 3 parameter set, the AVX2 optimized implementation reduces the median signing time from 1,713,783 cycles (reference code) to 428,587 cycles, and median verification time from 522,267 cycles to 179,424 cycles on a Skylake CPU. Further optimizations using AES (AVX2+AES) for expansion functions yield even faster results.

The security of Dilithium is based on the hardness of two primary lattice problems: Module Learning with Errors (MLWE) and Module Short Integer Solution (MSIS). Key security results are based on the ability to prove the scheme's security against these problems in the Random Oracle Model (ROM) and the Quantum Random Oracle Model (QROM).

The standard security objective is Strong Unforgeability under Chosen Message Attacks (SUF-CMA). The reduction proofs show that breaking SUF-CMA is equivalent to solving MLWE or the specialized SelfTargetMSIS problem, which involves MSIS combined with the hash function. While classical security proofs use the non-tight "forking lemma," evidence is mounting that the construction remains secure in the QROM, even with a tight reduction when the scheme is deterministic.

The concrete security analysis provides estimates for the BKZ block-size needed for the best known lattice attacks (Core-SVP) against MLWE and MSIS for various NIST levels. For instance, NIST Level 3 requires a BKZ block-size of 624 (LWE) and 638 (SIS) for the best classical Core-SVP attacks. Parameters can be increased (e.g., using Level 5 or Extended Sets 5+, 5++) by changing the matrix dimensions (k,ℓ) to increase lattice hardness.

Transitioning to a Quantum-Resistant Public Key Infrastructure

The source advocate for the use of hybrid digital signature schemes, which employ multiple algorithms (a traditional one and one or more post-quantum ones) in parallel. This strategy serves two primary purposes:

1. Hedging Bets: It mitigates risk, ensuring confidentiality or authenticity as long as at least one of the component algorithms remains unbroken.

2. Backwards Compatibility: It allows for security in post-quantum-aware systems while still functioning with not-yet-upgraded software.

The research addresses two key security properties for hybrid signatures:

1. Unforgeability (EUF-CMA): The widely accepted security standard for digital signatures is existential unforgeability under chosen message attack (EUF-CMA). The paper introduces a novel security hierarchy (XyZ) that distinguishes four types of adversaries based on whether they are classical or quantum during the signing interaction period (X), the type of access they have to the signing oracle (y, classical 'c' or quantum 'q'), and the type of adversary after the signing period (Z). This hierarchy includes CcC (fully classical) up to QqQ (fully quantum).

2. Non-Separability: This property is unique to hybrid schemes and aims to prevent an adversary from separating a valid hybrid signature into a valid signature recognizable as a single component scheme, which could misrepresent the signer's original intention. Non-separability relies on a "recognizer algorithm" (R) to help distinguish separated hybrid signatures.

The research evaluated methods for deploying hybrid signatures within three key standards: X.509, TLS, and S/MIME.

- X.509v3 Certificates: The standard permits only one subject public key and one CA signature. A promising backwards-compatible approach (Approach 2) is to construct a Dnest hybrid by having the CA sign the traditional certificate (c1) using Σ1, and then embed c1 as a non-critical extension inside a second certificate (c2) signed using Σ2.

- TLS (Transport Layer Security): While TLS 1.2 permits negotiation of a single signature algorithm, emerging features like the "post-handshake authentication" mode in TLS 1.3 drafts allow for multiple client authentications, functioning as a concatenation combiner (C‖).

    ◦ Testing revealed that large certificates pose compatibility issues: while TLS data structures can accommodate up to 16MiB, many libraries and web browsers failed to handle certificates containing extensions simulating large post-quantum schemes, such as those exceeding 43.0 KiB (SPHINCS size) or 90 KiB (for GnuTLS/OpenSSL), let alone the largest lattice schemes (1.3MiB).

- CMS and S/MIME: Using parallel SignerInfo objects (Approach 1, C‖) was found not to be fully backwards-compatible because several tested applications rejected messages with multiple SignerInfo objects. However, embedding the second signature as an optional attribute within the first SignerInfo (Approach 2.b, Cstr-nest) generally succeeded.

In summary, maintaining backwards compatibility is most easily achieved when post-quantum objects are placed as non-critical extensions or attributes of pre-quantum objects.

PUBLIC-KEY CRYPTOSYSTEM BASED ON ISOGENIES

Elliptic Curves and Isogenies: Elliptic curves over Fp are called isogenous if they have an equal number of points. The set of these isogenous curves, U, forms a category where isogenies act as morphisms. The size of this set is determined by the class number $h(D\pi)$.

- Isogeny Star: An isogeny star is a graph consisting of a prime number of elliptic curves connected by isogenies of prime degrees that satisfy the Elkies criterion.

- Routes: Movement within the isogeny star is defined by a route (R), which is a set of steps by specified isogenies in a positive direction determined by Frobenius eigenvalues. Routes are commutative.

- Encryption Scheme: The cryptosystem implements the ElGamal public-key encryption technique.

  - The private key is a route, $Rpriv$. The public key is an elliptic curve, $Epub$, computed by applying the private route to an initial curve, $Einit$ ($Epub=Rpriv(Einit)$).Encryption involves choosing a random route $Renc$, calculating $Eenc=Renc(Epub)$, and an additional elliptic curve $Eadd=Renc(Einit)$. The ciphertext s is computed as $m \cdot jenc$ (mod p). A variant exists where rational point mapping ($Pinit$) is used instead of the j-invariant for generating the ciphertext. Decryption uses the private route $Rpriv$ applied to the additional curve $Eadd$ to recover $Eenc$ and then the cleartext m.

The strength of the cryptosystem depends on the difficulty of finding an isogeny between the initial curve ($Einit$) and the public-key curve ($Epub$).

- Meet-in-the-Middle Attack: The most effective known technique for searching for an isogeny is the meet-in-the-middle attack. This attack selects $m \approx O(\log n)$ isogeny degrees, where $n=\#U$ (star size).

- Complexity Estimate: The complexity of the meet-in-the-middle attack is estimated at $O(\sqrt{n})$ isogeny computations. Since the size of the isogeny star n is approximately $O(\sqrt{p})$, the strength of the cryptosystem on isogenies of elliptic curves over Fp is estimated at $O(\sqrt{n} \sim O\sqrt{p})$.

- Quantum Resistance Hypothesis: The supposition that this system is hard to break with a quantum computer rests on the idea that chaining a series of q isogenies requires q consecutive solving of necessary equations , meaning computations cannot be easily parallelized, unlike the requirement for the Shor's algorithm black box.

Selecting parameters involves choosing discriminants $D\pi$ that yield a large class number $hD\pi$ or one with a large prime divisor r, in which case the cryptosystem strength is estimated at $O(\sqrt{r}$. Small-degree isogenies are efficiently computable, with complexity $O(l(logp)^2)$.

How to not break SIDH

Isogeny-based cryptography is a relatively new post-quantum approach offering advantages like relatively tiny key sizes and suitability for functions like non-interactive key exchange. An isogeny is a surjective morphism of elliptic curves that preserves the identity, acting as a group homomorphism given by rational maps.

SIDH (Supersingular Isogeny Diffie–Hellman), proposed in 2011, was the first practical isogeny-based key exchange. SIDH’s core contribution was resolving a correctness challenge in the chaotic isogeny graph structure by incorporating auxiliary points (the images of the other party's torsion basis) to help Alice and Bob obtain a shared secret,

The sources detail several approaches that failed to improve upon this generic attack:

1. Exploiting the Fp-subgraph: This approach attempts to use the subset of curves defined over the smaller field Fp. While finding a path inside the Fp-subgraph can be reduced to O~(p1/4) using multiple isogenies, this is not better than the standard meet-in-the-middle attack because the secret SIDH isogenies are too short, and almost none of the curves along the short path are defined over Fp. Finding the Fp-subgraph itself by brute force costs O~(p[redacted]), which is also too expensive.

2. Lifting to Characteristic Zero: Although ordinary elliptic curves have a canonical lift to characteristic zero that preserves the endomorphism ring, this method is not applicable to generic supersingular elliptic curves. For generic supersingular curves, the only known endomorphisms are scalar multiplications, and finding a suitable non-scalar endomorphism is computationally equivalent to breaking the scheme.

Failed Attacks Utilizing Auxiliary Points

1. Interpolation Techniques: Because isogenies are rational maps, one might hope to recover the coefficients using the input/output pairs provided by the auxiliary points. This is computationally infeasible because the degree of the rational map is exponentially large, meaning that even printing the interpolated result would take exponentially large time.

2. Purely Group-Theoretic Approaches: Attempting to extrapolate the known action of the isogeny on the *ℓBnB*-torsion to either a coprime torsion subgroup or a higher *ℓB*-power torsion subgroup is "doomed to fail". The group-theoretic structure of elliptic curves dictates that coprime torsion subgroups are independent, providing no non-trivial relations for extrapolation.

In summary, many conceptually appealing attacks that leverage standard cryptanalytic techniques or structural properties of elliptic curves have been explored but found to be inefficient or inapplicable to the specific parameters and setup of the SIDH/SIKE protocols.