Hamming Quasi-Cyclic Encryption Schemes

Jacob Ryder

Rochester Institute of Technology

17th April, 2018

Outline

Introduction

Core Concepts
- Linear Codes
- Systematic Quasi-Cyclic Codes
- Syndrome Decoding

HQC Public Key Encryption

HQC Key Encapsulation

Security Analysis
- Recommended Parameters
- Known Attacks

Advantages and Limitations

Introduction

Hamming Quasi-Cyclic is a set of encryption schemes based on code theory, and a submission to the NIST search for quantum-resistant algorithms.

The scheme relies on the computational hardness of the syndrome decoding problem over systematic quasi-cyclic BCH codes.

The authors of this scheme are Aguilar Melchor, Aragon, Bettaieb, Bidoux, Blazy, Deneuville, Gaborit, Persichetti, and Zémor. Most are researchers in French universities and organizations.

Several authors also contributed to the Ouroboros scheme, which similarly relies on decoding double-cyclic codes.

Linear Codes & Terminology

In finite binary field F with dimensionality k, a linear code C is a subspace of F. C contains n vectors, called codewords.

A matrix G is a generator for code C if the product of G and any vector from F is a codeword of C. In other words, the codewords of C span the rows of G.

A matrix H is a parity check matrix for code C if the product of H and any codeword of C is 0. In other words, H is a generator for the dual of C. The product vH is called the syndrome of v.

The Hamming weight of a vector is the number of non-zero coefficients it contains. For a binary field, this is easily evaluated.

Systematic Quasi-Cyclic Codes

A linear code C is quasi-cyclic of index s if each codeword c can be divided into a collection of s blocks, and a simultaneous circular shift on each block of c yields another codeword in C.

Traditionally, code-based cryptography suffers from inordinate key lengths. Utilizing quasi-cyclic codes sharply reduces the space needed to represent both codes and codewords, improving efficiency dramatically.

A systematic quasi-cyclic code is a quasi-cyclic code which has a specific simplified format of parity matrix, further aiding efficiency.

Syndrome Decoding Problem

Given a codeword c and an error e, the resultant vector v = x + e. The syndrome of c is 0, so the syndrome of v is the syndrome of e. For a code with tolerance δ, v can be decoded to c if $w(e) < \delta$.

Decision Syndrome Decoding Problem:
Given a syndrome y, parity matrix H, and weight w, find a vector x such that y is the syndrome x and x has weight w.

A substantial corpus over the last several decades has equated the Syndrome Decoding problem to the Shortest Vector problem and Subset Sum problem, which is NP-Complete. More contemporary research has confirmed that the optimizations presented in this scheme do not reduce the problem's complexity.

Hamming Quasi-Cyclic Public Key Encryption

The HQC encryption scheme relies on BCH error-correcting codes, which are quasi-cyclic and systematic to improve performance. A vector with a weight within δ of a codeword can be decoded as that codeword, while a larger offset cannot be decoded correctly.

HQC.PKE relies on masking messages with non-deterministic vectors based on a public/private key pair. Without knowledge of the keys, the task of identifying the message is NP-Complete.

Through a series of proofs in the form of adversarial games, this scheme is shown to be indistinguishable against chosen plaintext attacks (IND-CPA).

HQC.PKE Setup

Setup

param $= \{n, k, \delta, w, w_r, w_e\}$

n is the length of each codeword

k is the dimensionality of the code

δ is the error limit for the code

w, w_r, w_e are all weights for later calculations

HQC.PKE Key Generation

Key Generation

Generator G for Code C is published.

$x, y \leftarrow R$ such that $w(x) = w(y) = w$
$sk = (x, y)$

$h \leftarrow R$, $s = x + h \cdot y$
$pk = (h, s)$

HQC.PKE Encryption

Encryption - E(pk, m) = c

$r_1, r_2 \leftarrow R$ such that $w(r_1) = w(r_2) = w_r$
$e \leftarrow R$ such that $w(e) = w_e$

$u = r_1 + h \cdot r_2$
$v = mG + s \cdot r_2 + e$
$c = (u, v)$

In short, the encoded message mG is offset by a vector $s \cdot r_2$ based on the public key $pk = (h, s = x + h \cdot y)$ and a noise vector e.

HQC.PKE Decryption

Decryption - D(sk, c) = m

$m = C.Decode(v - u \cdot y)$

The ciphertext can be decoded with knowledge of sk = (x,y).
$(v - u \cdot y)$ expands to $(mG + x \cdot r_2 - y \cdot r_1 + e)$.
C can decode this to m iff $w(x \cdot r_2 - y \cdot r_1 + e) < \delta$.

Hamming Quasi-Cyclic Key Encapsulation Mechanism

An extension of HQC.PKE reliant on a set of secure hash functions H_1, H_2, H_3 with $H_1 \neq H_2$. Setup and Key Generation are the same as in HQC.PKE.

A symmetric key is derived from a random seed and its encryption. Securely transmitting the seed allows for a secure key exchange. To ensure this encryption is reproducible, the encryption nonces are also deterministically generated from the seed.

This specific transformation from the IND-CPA HQC.PKE is provably indistinguishable against adaptive chosen ciphertext attacks (IND-CCA2).

HQC.KEM Encapsulation

Encapsulation

$m \leftarrow R$ is selected as the shared seed.
Randomness $\theta = H_1(m)$ is used to derive e, r_1, r_2.
The seed is encrypted as $c = E(pk, m, \theta)$.
The shared key is derived as $H_3(m, c)$.
The values $(c, H_2(m))$ are shared.

Using $H_1(m)$ to generate the nonces reduces the non-deterministic nature of the encryption, but as long as $H_1 \neq H_2$ the nonces are not exposed.

HQC.KEM Decapsulation

Decapsulation

The ciphertext is decrypted as $m‘ = D(sk, c)$.
The seed is reencrypted as $c‘ = E(pk, m‘)$ and $d‘ = H_2(m‘)$.
If these validate, the shared key is derived as $H_3(c‘, m‘)$.
If not, the seed is rejected.

Security of Recommended Parameters

Instance	n_1	n_2	n	k	δ	w	w_r, w_e	security	p_{fail}
Basic-I	766	29	22229	256	57	67	77	128	$< 2^{-64}$
Basic-II	766	31	23747	256	57	67	77	128	$< 2^{-96}$
Basic-III	796	31	24677	256	60	67	77	128	$< 2^{-128}$
Advanced-I	796	51	40597	256	60	101	117	192	$< 2^{-64}$
Advanced-II	766	57	43669	256	57	101	117	192	$< 2^{-128}$
Advanced-III	766	61	46747	256	57	101	117	192	$< 2^{-192}$
Paranoiac-I	766	77	59011	256	57	133	153	256	$< 2^{-64}$
Paranoiac-II	766	83	63587	256	57	133	153	256	$< 2^{-128}$
Paranoiac-III	796	85	67699	256	60	133	153	256	$< 2^{-192}$
Paranoiac-IV	766	89	70853	256	60	133	153	256	$< 2^{-256}$

Differences in the code size n and the weights w, w_e, w_r result in differing tiers of effective security and differing decryption rates.

Known Attacks

Structural Attacks

The cyclic structure of these codes can be attacked to a degree. The Decoding One Out of Many (DOOM) attack is the best general attack against these codes, with improvement up to $O(\sqrt{n})$. The polynomials used to generate quasi-cycles can be attacked in some cases. The recommended values for n result in polynomials with very few irreducible factors, which mitigates this risk.

Quantum Resistance

The recommended parameters yield pre-quantum security levels of 128, 192, and 256 bits, each with varying decryption failure rates. A quantum adversary would be able to halve the security level, but would not be able to affect the decryption rate in the same way. Under best attacks, the HQC schemes still offer substantial security.

Advantages and Limitations

Advantages

1. Proven IND-CPA given hardness of Syndrome Decoding
2. Proven IND-CCA2 through KEM-DEM transformation
3. Not reliant on the secrecy of selected codes
4. Provides precise estimates of failure probability

Limitations

1. BCH codes give precise estimates, but suboptimal performance
2. HQC.PKE has a low encryption rate, block sizes beyond 128/192/256 would require new parameters
3. In contrast to Lattices/Ring Learning With Errors, cannot benefit from Search to Decision reduction for structured codes