The Payne Institute for Public Policy

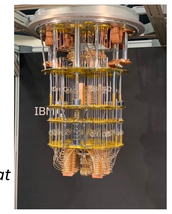
COLORADOSCHOOLOFMINES
EARTH • ENERGY • ENVIRONMENT

Applications of Post-Quantum Cryptography – Survey and Application of Machine Learning

Mack Osborne

Abstract

Abstract - Quantum Computing poses a considerable threat in the world of cyber security. Policy makers are largely unprepared for a post-quantum world, significantly due to a lack of understanding and awareness. The goal of this paper is to improve understanding and provide a new and effective way to analyze post-quantum cryptography, for researchers and security engineers alike. This is done by providing a background of quantum computing, a survey of the state of technologies and relevant policies, and a novel application of machine learning to perform analysis of quantum-ready encryption. The machine learning research will provide a Multinominal Naïve Bayes for discrete analysis of the RSA and CRYSTAL-Kyber encryptions.



The IBM Q quantum system at Semicon West

Methods

- Kyber512, Kyber768, Kyber1024, and RSA Algorithms were implemented entirely in Python 3.7. The programs could simulate a nonsense text (*gp*), random real-text(rp), and static real-text(kp). Texts were then encrypted and flagged for creating training models.
- Baseline was established with undecrypted, equivalent datasets (The model was able to classify with 100%)
- WEKA was used to build models based off training data using Multinominal Naïve Bayes
- Model then tested against unflagged ciphertexts to create a confusion matrix.
- All testing was done on an Intel(R) Core(TM) i7-10700K CPU @ 3.80GHz with 16 GB RAM running Ubuntu 22.04

Version	Security Level	Private Key Size	Public Key Size	Ciphertext Size
RSA	**AES128**	**384**	**384**	**384**
Kyber512	**AES128**	**1632**	**800**	**768**
Kyber768	**AES192**	**2400**	**1184**	**1088**
Kyber1024	**AES256**	**3168**	**1568**	**1568**

Future Direction

- Implementations of Chi Square, Reinforcement Learning, and Deep Learning could improve model
- Similar Areas of Application
  - -Spam Detection
  - -Snort Deep Packet Inspection
  - -Cryptocurrencies

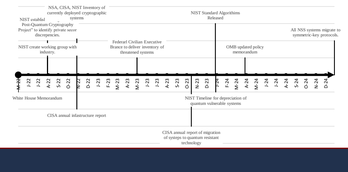
Introduction

Background

Superposition is the concept that a particle can be in multiple states at once. It comes from the notion that we cannot say with certainty where a subatomic particle is at any given time, instead we can only predict where it may be with some degree of probability, such as a probability distribution. Quibits are the units of quantum computing, as opposed to classical bits. While bits represent 1 or 0, qubits can represent 1, 0, or a linear combination of 1 and 0.

This leads us to the concept of quantum entanglement. Intuition serves that measuring a singular qubit is much less useful than a single bit. But quantum computers utilize quantum entanglement and observe the behavior of quibits on one another. Consider the probability distribution of a qubit as a wave, by observing the way waves interact, either negating or adding to one another, we can collect data.

Through this system we can collect the data of 16,000 bits in just 10 qubits. And the data representable by 500 qubits would require more bits than there are atoms in the known universe [4]. We can easily see how modeling quantum systems is too expensive for a classical computer. While it is not so simple to say quantum computers dwarf classical computers entirely, in fact there are many applications which show the opposite, one key advantage of quantum computers is factoring large prime numbers. The hardness of factoring large sudo-prime numbers is the basis of RSA encryption, and therein lies the problem. Public key encryption is now, effectively, unsafe.



CRYSTALS-Kyber

Kyber is an IND CCA2- secure key-encapsulation mechanism. The security of KYBER is based on the hardness of solving the learning-with-errors in module lattices (MLWE). The NP-hard reduction of MLWE to the Shortest-Vector Problem (SVP) is useful for understanding lattice based cryptographic systems. SVP requires one to find the shortest, non-zero vector in a matrix width n with regard to a conventional norm, most commonly Euclidian. The strongest results are known in the $\ell_{(\infty)}$ form. SVP($\ell_{(\infty)}$) is NP-hard under deterministic reductions for quasi polynomial approximation factors $2^{O\left(\frac{\log n}{\log\log n}\right)} = n^{O(1)}$

An example application follows[7]. Consider the vector as a polynomial with small coefficients ***s***. This serves as the private key and is unique per key-generation.

$$s = (-x^3 - x^2 + x - x^3 - x)$$

The public key consists of the matrix width n=2, which is generated with a series of polynomials with random coefficients modulo ***q***. Consider matrix ***A***:

$$A = \begin{bmatrix} 6x^3 + 16x^2 + 16x + 11 & 9x^3 + 4x^2 + 6x + 3 \\ 5x^3 + 3x^2 + 10x + 1 & 6x^3 + x^2 + 9x + 15 \end{bmatrix}$$

The public key contains another vector, ***t***, to generate this we introduce error noise via error vector, ***e***.

$$e = (x^2, x^2 - x)$$
$$t = A * s + e$$
$$t = \left(16x^3 + 15x^2 + 7{,}10x^{3{,}12x^2} + 11x + 6\right)$$

We have now completed key generation for *Private-Key(s)* and *Public-Key(A,t)*. Text encryption using this keypair requires 3 additional vectors, randomizer ***r***, and error vectors $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$. Consider the example:

$$r = (-x^3 + x^2, x^2 - 1)$$
$$e_1 = (x^2 + x, x^2), e_2 = (-x^3 - x^2)$$

Encryption begins by converting message ***m*** to a binomial. Consider ***m*** = 6. $\boldsymbol{m_{bin}}$ =110. $Poly(m_{bits}) = \mathbf{1}x^2 + \mathbf{1}x^1 + \mathbf{0}x\text{^}0$). The coefficients are scaled using $\left\lfloor \frac{q}{2} \right\rfloor$, where ***q*** is the same value used during matrix generation. Consider ***q*** =17, the final $Poly(m_{bits}) = (9x^2 + 9x)$. Ciphertexts consist of two values ***u*** and ***v*** whereas:

$$u = A^T r + e_1$$
$$v = t^{Tr} + e_2 + Poly(m_{bits})$$

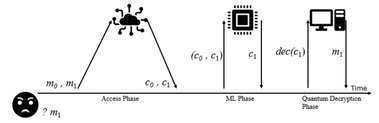
Thus we complete ciphertext(***u***,***v***). In this example:

$$u = (11x^3 + 11x^2 + 10x + 3{,}4x^3 + 4x^2 + 13x + 11)$$
$$v = (-2x^3 + 6x^2 + 8x + 6)$$

Recovering s from the public key would require solving MLWE. In real Kyber, implementations matrixes are expressed with module lattices, meaning a small matrix of a constant size polynomial ring.

Semantic Security

Semantic security is the notion that an adversary is unable to distinguish between two ciphertexts. It is often evaluated in turn with indistinguishability and randomness. Consider this high-level example. There are two messages m¬0 and m1. And adversary obtains ciphertexts c0 and c1 but is unsure of the mapping. In this example, if the adversary is unable complete the mapping with over 50% accuracy, then the encryption is semantically secure. This may seem like a foolhardy method for an attacker, but there are methods to exploit this in encryption systems. For example, if the user can obtain a plaintext or isolate a specific ciphertext, then the system is vulnerable to chosen plaintext attacks (CPA) or chosen ciphertext attacks (CCA).



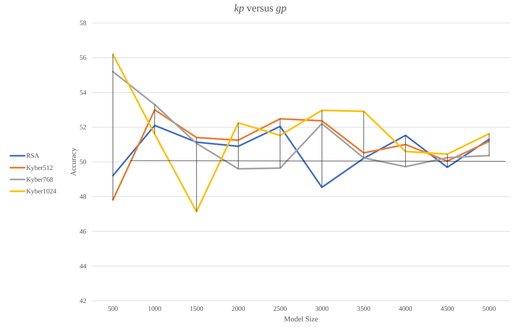
This figure presents such a scenario and is based on Boneth's research, Secure Signatures and Chosen Ciphertext Security in a Quantum Computing World. Boneth proves that in quasi-length preserving schemes it is possible to map the core encryption of a plaintext to a smaller subset of ciphertexts. Quantum computers are able to process this mapping and eventually access a 1 to 1 mapping of plaintexts and ciphertexts. Gagliardoni et.al. presents a solution of instilling randomness prior to encryption to mapping unobtainable by quantum systems.

Results

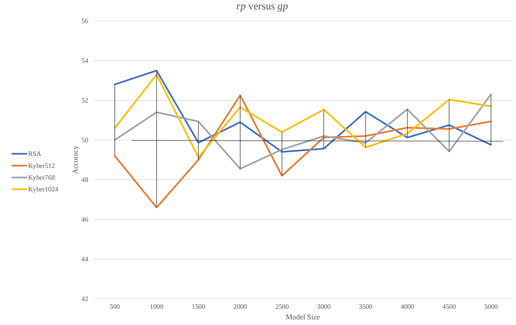
kp versus gp				
Model Size	RSA	Kyber512	Kyber768	Kyber1024
500	✓ 49.2	✓ 47.8	✗ 55.2	✗ 56.2
1000	✗ 52.1	✗ 53	✗ 53.3	✗ 51.6
1500	✗ 51.1333	✗ 51.4	✗ 51.0667	✓ 47.133
2000	✗ 50.9	✗ 51.25	49.6	✗ 52.25
2500	✗ 52.04	✗ 52.48	49.64	✗ 51.52
3000	✓ 48.5333	✗ 52.3667	✗ 52.2	✗ 52.967
3500	50.2	✗ 50.5143	50.2286	✗ 52.914
4000	✗ 51.525	✗ 51	49.725	✗ 50.6
4500	49.6889	50.0444	50.2444	50.444
5000	✗ 51.3	✗ 51.18	50.36	✗ 51.62

rp versus gp				
Model Size	RSA	Kyber512	Kyber768	Kyber1024
500	✗ 52.8	✓ 49.2	50	✗ 50.6
1000	✗ 53.5	✓ 46.6	✗ 51.4	✗ 53.3
1500	49.8667	✓ 49	✗ 50.9333	✓ 49.133
2000	✗ 50.9	✗ 52.25	✓ 48.55	✗ 51.65
2500	✓ 49.4	✓ 48.2	49.52	50.4
3000	49.5667	50.1333	50.2	✗ 51.533
3500	✗ 51.4286	50.2	49.8571	49.629
4000	50.125	✗ 50.625	✗ 51.55	50.325
4500	✗ 50.7556	✗ 50.5556	✓ 49.4222	✗ 52.044
5000	49.76	✗ 50.94	✗ 52.3	✗ 51.7

kp versus gp



rp versus gp



Discussion

There are a few observations observable in this data. We see that are a significant number of tests that finished above 50% accuracy. Models built on smaller training sets have a larger variance between algorithms. This is not reflective of the algorithm but of the quality of the model. This trend is more visible in the line graphs, which suggest effective model for such ciphertext classification benefit from larger data sets.

We observe far more semantically secure models in Case 2 (*rp* VS *gp*), that is models that cannot classify *rp* versus *gp* with accuracy >=50% within a 0.512 relative absolute error (*RAE*) average across models. This is a predictable result as it shows the principal advantage of a known plaintext.

RSA noticeably outperforms Kyber in this testing. This is likely attributed to the increased ciphertext size required in Kyber. The larger text allows more points of analysis by multinominal naïve bayes.

We can say this method of classification from Alshammari is a valid application to ciphertext identification. Continued research could improve the understanding of the vulnerability. However, the model was unable to significantly outperform a 50/50 chance and models that were closest to doing so are less dependable due to their variance.

References

[1]Alshammari, Riyad & Zincir-Heywood, A.. (2009). Machine learning based encrypted traffic classification: Identifying SSH and Skype. 1 - 8. 10.1109/CISDA.2009.5356534.
[2]Boneh, D., Zhandry, M. (2013). Secure Signatures and Chosen Ciphertext Security in a Quantum Computing World. In: Canetti, R., Garay, J.A. (eds) Advances in Cryptology – CRYPTO 2013. CRYPTO 2013. Lecture Notes in Computer Science, vol 8043. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40084-1_21
[3]Gonzalez, Ruben. "Kyber - How Does It Work?" Approachable Cryptography, Cryptopedia, 14 Sept. 2021, https://cryptopedia.dev/posts/kyber/.
[4] Grimes, Roger A. Cryptography Apocalypse: Preparing for the Day When Quantum Computing Breaks Today's Crypto. Wiley, 2020.
[5]"National Security Memorandum on Promoting United States Leadership in Quantum Computing While Mitigating Risks to Vulnerable Cryptographic Systems." The White House, The United States Government, 4 May 2022, https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/04/national-security-memorandum-on-promoting-united-states-leadership-in-quantum-computing-while-mitigating-risks-to-vulnerable-cryptographic-systems/.