Quantum zero knowledge proofs

Nono Horiuchi

May 2022

1 Introduction

The goal of this paper is to outline a loose map of concepts that connect zero knowledge proofs to quantum mechanics/computing, then offer a new perspective and potential application of those concepts. First, I will briefly summarize the history of cryptographic systems from Diffie-Hellman-Merkle to zero knowledge proofs, then explain necessary modifications to the latter to make them, lastly leaving a discussion on how quantum zero knowledge proofs could be used to solve the common sense problem of AI.

2 Securely exchanging private information

As technology constantly develops over time, so do means of breaching information privacy. The problem of securely exchanging private information over a public channel is ever present.

3 Diffie-Hellman-Merkle protocol

The Diffie-Hellman key exchange is one of the first public key protocols which allows users to communicate private information through a public channel. To illustrate this process in terms of modular arithmetic, assume that there are two parties, Alice and Bob, who have no information about each other. In order for Alice and Bob to share a secret key y, they proceed as follows:

1. Alice and Bob both agree to use modulus m and generator g, which are made public.
2. Alice picks a random private number a and sends p to Bob, where $p = g^a$
3. Bob chooses a random private number b where $q = g^b$ and sends q to Alice
4. Alice then computes $y = q^a \pmod{n}$
5. Bob computes $y = p^b \pmod{n}$

Now Alice and Bob both share a private key y.

4 Proof systems

A proof is a sort of protocol in which a prover convinces the verifier that a statement is true. Furthermore, there are two types of proof systems: interactive and non-interactive.

Interactive proofs, as implied in the name, need both the prover and verifier to be present at the same time to exchange information until the statement is successfully proved to the verifier. Interactive proofs require completeness (if the statement is true, it can be proved to the verifier by following protocol) and soundness (if the statement is false, it cannot be proved to the verifier even without following protocol). The complexity of interactive proofs (IP) and its quantum counterpart (QIP) both equal PSPACE.

Non-interactive proofs do not need the prover and verifier to interact. This gives non-interactive proofs the advantage of zero knowledge. Zero knowledge means that neither the prover or verifier learn anything new about each other during the proof protocol. The zero knowledge property has made zero knowledge proofs popular in cryptography and cyber security, as it allows for the prover to confirm to the verifier that they have access to a secret key without needing to reveal any information about the key itself or parties involved in the proof. Non-interactive zero knowledge proofs exist for all problems in NP [1], which means it is easier than problems solved with interactive proofs.

Interactive zero knowledge proofs take the best of both worlds. While non-interactive proofs have the advantage of zero knowledge, some problems, such as 4-colorability or SAT can only be proved with interaction due to PSPACE being harder than NP. Which brings us to the next topic:

5 Fiat-Shamir Heuristic

In order to achieve interactive zero knowledge proofs, we must be able to convert an interactive proof to a non-interactive proof. The Fiat-Shamir heurustic accomplishes this. The protocol is similar to that of Diffie-Hellman-Merkle key exchange(DHM), although the aim differs slightly in that DHM is about securely sharing a private key, while Fiat-Shamir is about sharing the knowledge that the prover has the private key. In the following is an explanation using modular arithmetic:

1. Alice (prover) and Bob (verifier) publicly agree on generator g, Alice and Bob know a secret key y, but Bob does not know whether Alice is being honest about knowing the secret, Alice must convince Bob.

2. Alice wants to convince Bob that there is a number x such that $y = g^x$.

3. Alice privately picks a random number s and sends $t = g^s$ to Bob through a public channel.

4. (a) (interactive) Bob chooses a random number n, and sends this to Alice through the public channel. OR

   (b) (non-interactive) Alice uses a random oracle model to access a random n.

5. Alice sends Bob a number r such that $r = s - nx$.

6. Bob checks that $t = (g^r)(y^n)$, which if true convinces him that Alice knows y because $(g^r)(y^n) = g^{s-nx}g^{nx} = g^s$

As can be seen in the steps above, the Fiat-Shamir heuristic allows proofs to be converted between interactive and non-interactive.

6 Quantum zero knowledge proof

Classical zero knowledge proofs do not hold up against quantum attacks. Mathematical details are outlined in this paper [2], so I will attempt to explain it in a more conceptual sense.

The introduction of quantum computers changes keys from numbers to quantum states. For simplicity, assume that there is an unreliable/cheating verifier that has access to quantum algorithms and uses quantum states.

The protocol follows the same steps as the Fiat-Shamir algorithm shown in the previous section, the only changes being that instead of choosing random numbers s and n, Alice and Bob get a state that is the superposition of all possible s and n, measure the register, and send the resulting bit to the other party. By the end of the protocol, if the state Alice sends Bob and the secret key that Bob knows are equal, then Alice has proved her trustworthiness. If not, then there is a possibility that a malicious third party is interfering or natural errors are occurring. In that

case, because measurements have already been done, the starting state has collapsed, and Alice and Bob cannot go back. The solution is to apply the inverse of some of the steps in the protocol to get a state that is, as [2] puts it, "partially unwound". This will allow Alice and Bob to run the algorithm again from a new state.

7 Discussion

The use of zero knowledge proofs in cryptography is mainly focused on the aspect of communicating knowledge of information without giving away that piece of information. In a quantum context, zero knowledge could mean checking whether a quantum state exists without observing the actual state. With that interpretation, quantum zero knowledge proofs could be a potential solution to the "common sense problem" of AI. As Stephen Pinker writes: "...the brain's processors are slower and fuzzier, but there are hundreds of billions of them, each connected to thousands of others. That allows the human brain to recognize complicated patterns in an instant. Computers have to reason out every niggling detail one step at a time." [3] Human minds are built to filter out information that is irrelevant to a task in almost an instant without using a lot of energy, something that machines currently cannot do without someone explicitly coding in instructions.

For example, imagine an AI robot given a task to put a box away inside a room. If a person were doing this tasks instead, they would consider only a few relevant variables, such as whether the box would be secure and convenient to access. An AI built with a modern computer would not only be able to have foresight like this without having is hard-coded in, it would have to scan the room and compute every open space to determine whether the box can fit in it. The limitation of only being able to process instructions "one step at a time" is called the Von Neumann bottleneck, named after the Von Neumann architecture which all modern computers are based upon. Quantum computers, if technology reaches a point where they are comparable to or surpass classical computers, could be used to overcome the bottleneck.

A quantum computer AI could first take the superposed state of all possible places where the box can be placed in the room, then use a non-interactive quantum zero knowledge proof to determine whether an ideal spot for the box (the secret key) exists, without having to compute and test for each of the possibilities.

This method results some mysteries that I do not know how to answer:

1. "All possible places the box can be placed in the room" is still a set of specific states, so how would one filter out every other state with a reasonable amount of computation? It seems like the "common sense problem" all over again.

2. Will the AI decide on its own what the "ideal spot for the box" will be, or will a person code it in? It seems to me that it has to be the former case, because in the latter case the "common sense" is being injected into the robot manually by people, undermining the purpose of the AI's design, which is to be able to make "common sense" decisions on its own.

It is to be noted that my conception of "AI" comes from the current AI trained with machine learning. The way AI is approached my change in the future. For instance, AI may be two or more computers communicating to make decisions (not too unlike the relationship the prover and verifier have in interactive proofs), instead of the one computer which handles all of the data and computation. Depending on where the approach goes, the above questions may not need to be answered, or new problems may arise.

References

[1] D. J. W. Henry Corrigan-Gibbs, Sam Kim. Interactive proofs and zero-knowledge. Ac-

cessed: 2022-05-01.

[2] J. Katz. Quantum zero knowledge. Accessed: 2022-05-01.

[3] S. Pinker. Can a computer be conscious? Accessed: 2022-05-01.