CMPS 10 Lecture Notes: Lecture 2 (1-7-2016)

Introduction/logistics

Office Hours: 3:30-4:30. Day is still undecided (no strong preference either way from class)

- Class Demographics
  - How many are taking class to fulfill GE (maybe about 10)
  - Engineering (not that much)
  - bio/physics/math (seemed like the most)
  - humanities (a handful)
  - arts (just a couple)

MIDTERM: 18TH of FEBRUARY If you can't make it, let teacher know

- Idea of information is far more tangible than what the everyday usage of the word is
  - information is as tangible as mass!
  - Computation is not something that just computers do. it happens in nature all the time!
- So far, only 40-50 people have participated in the lab survey.
  - do it by the end of the day, because otherwise you will just be assigned to one.

How many people have finished homework (maybe 25 percent
Last week we talked about privacy, and how the great availability of computation and the capacity to store information, dramatically changes standard notions of privacy

Today we will talk about a different idea: secrecy (distinct from privacy): the two parties work that maintain secrecy with the expectation that people will acting adversary.

Secrecy

- http:// vs https:// :: sometimes s sometimes not! S stands for secure!
  - when there IS an s, it is supposed to mean that communication is secure
  - What communication, what does secure mean? Can sometime tell me?
  - Silence from class. Then someone answers, but too quiet to hear.
- So, you are typing at your computer (when https was introduced, used for commercial transactions)
  - And someone far away is trying to sell you something, and so you fill out a form that asks for your credit card information.
  - You obviously do not have a direct connection between your computer and the website selling you something.
  - There are a ton of intermediate computers in your connection.
  - Merchant computer does not even live at merchant site!
- So, somewhere on this webform, you write down your credit card information (e.g. 5493 7265 6914 2319 May 19 723)
  - And you send it and it travels and it goes to merchant computer.
  - Now, even if you were willing to trust your merchant (just like you trust your waiter at a restaurant)
  - But HERE you do not have any idea who all those intermediate computers are!
  - So at first, people were really paranoid.
- So, what S means, is that even though this is the first time you are visiting this site, and you send them information, no one else in between will be able to understand what this information means.

  - How does this happen?
    * student answer: encrypt it.
  - The key is you have no idea who this person is. You do not call Amazon and say “hey, we‘re going to set up this secret key“ and they say “yeah sure¡‘
  - what happens is you go to this computer, that is EXACTLY as strange to you as you are to them, yet they will be able to figure out your info but the intermediates won‘t!

- So, let‘s make sure we are all on the same page about what it means to encrypt something.
  - Starting with the simplest

Encryption

One-Time Pad

- two parties meet in a dark corner in secret, and they assume that no one can hear it.
  - They take out a coin, and they flip it 100 times, and they record the coin flips.
  - And THAT is the secret. They take two copies of those recorded coin flips and put them in the pockets.
  - And then, let‘s say, one of them know they are the sender, the other is the receiver.
  - Let‘s say that the information to be transmitted is a binary string of length 100 (binary means either 0 or 1).
    * So the sender has the “true message“ that they want to send.
    * But the problem is that if someone other than the receiver hears this then terrible things happen (maybe the message is “this is where we will deploy the troops“).
    * So something besides the true message is actually being sent. You compare the string of coin flips to the true message, bit by bit.
      · If the coin flip was heads, you type the actual thing from the true number.
      · If the coin flip was tails, you type the ‘opposite‘ thing than the true number.

Checked in with the class to make sure everyone was on the same page.

- So now we have the “Sent“ message, which is what the person will actually send.
  - Now imagine you are an eaves dropper, and that is what you get to hear. What can you assume about the original message.
  - Can you extract any information from the original message about this?
  - Is it clear that the answer is no?
    * Since the coin flips are independent! (independent means: no matter how many coin flips you have flipped so far you have no way of knowing what the next coin flip is going to be. That is why going to the casino and betting after 5 blacks was a bad idea)
    * So even if you could decrypt the 1st bit it does not help you decrypt the rest.
    * And you know absolutely nothing as to why the 1 is a 1! It could be because you were meant to send a 1, or it could be because you meant to see a 0!
  - This is perfect scrambling! It has not been proved in class yet, but there is hopefully enough intuition that this is ultimately secure.
  - But because the receiver knows the coin-flip string, they can then successfully decode it, again, bit by bit.
  - Question: is transformation independent from the message? Answer: yes!

- Disadvantages:
  - Can you use it twice?
    * Is it risky to use it twice? Maybe maybe not? Why will they figure out if you use it twice but not once?
    * They might be able to find patterns
    * What the eavesdropper WANTS is the coin-flip string. Because once they get that then they get everything.

  - If you use it twice,bad guys MAY get an advantage (not necessarily going to happen)
    * They might already know for whatever reason that first bit has 60 percent chance of being 0, and second one has 30 percent chance of being 0.
    * Even when they have this knowledge, when they only see it once, they have no additional knowledge, they have not learned anything.
    * But if you do this many many times, you should be getting 0s 60 percent of the time for that first bit!
      · Example: first bit is going to represent north or south, and because they know things about your position, they know it is more likely to be north than south.
      · So when you do it a 1,000 times, and you are expecting a 1 in the first bit, if you GET a 1, then you know that the first coin flip was heads, and if you get a bunch of 0s, then you know that the first coin flip was tails.

- The above is known as the ONE-TIME PAD
  - PAD referes to the fact that you are doing the flipping
  - ONE-TIME means that if you want to be really safe you can only do it once.
  - Click me for Wikipedia article
- Pretty inefficient: it requires the two parties to have that secret channel with each other
  - The analog of calling Amazon and saying "let's flip some coins first before I send you my credit card info"

Cyclic Shift

People have been doing cryptography since ancient times.

- Let us say that you have all of the letters, A, B, C, D, all the way to Y, Z
  - And let's say that A is 1, B is 2, ... and Z is 26.
- Cyclic Shift: Add 4 to everything (A becomes E, B becomes F, D becomes G, and it wraps around again when you get to the end of the alphabet.
- So here, the ONLY secret information is the length of the shift. A single number between 1 and 26.
- Click for Wikipedia article
- Of course, this is a very easy code to break. How could we make it harder?
  - Student answer: we could add another thing.
- Before we add another thing: we could say that EACH letter gets assigned to a new one RANDOMLY.
  - Now there is more information that needs to be sent (what each letter translates to)
  - BUT the 'secret part' is now relative to the size of the ALPHABET as opposed to the size of the MESSAGE (so just 26 pieces of information is enough to transcribe an entire book in this case!)
- You could also have letters translate to made-up symbols! The important thing is that needs to be a 1-1 function (the function needs to be an injection).
- Code is independent of message length (good!), but it is not terribly secure (bad).
- But in actuality, when you exchange information with a website, there is no secret code whatsoever.
  - So when you see that S in https, it means that it has to be secure!
  - But you do not share ANY secret information with the merchant!
  - But the claim is that you can start telling them stuff, but what you tell them can be intelligible to the merchant but not to anyone else.
- The other claim is that *only* the merchant gets to learn your credit card information, and no one else in between.
  - Why is the merchant different than the intermediate computers? They are just as unknown to you as them.
  - But it IS possible!

Public Key Cryptography

- This is known as PUBLIC KEY CRYPTOGRAPHY
  - Let us say that a group of people, as big as you like (bigger than 2, let us say 5), see each other for the first time in their lives
  - They sit in a room.
  - the only form of communication is saying things out loud. So everyone can hear everything.
  - And so they start talking.
  - And the claim is that they can start talking: after some preliminary talking, they can reach a state, in which one of them will be able to say something, such that ONLY ONE other person in the room understands what they are saying.
  - HOW CAN THIS BE POSSIBLE!?
  - It turns out it is NOT possible!?!
  - So it turns out public key cryptography isn't possible!?!
  - But it IS possible!
  - EXCEPT IT IS POSSIBLE!?!?!!?!!?!
    - Professor just has the goal of making it clear to us that it is a miracle! And it is a miracle that is possible!
    - The key point is to define the word "is"
- If the people in the room can do arbitrarily complicated mathematical equations, with infinite speed, then the world described is impossible.
- IF, though, there are limits to how much computation that they can perform, then it IS possible.
  - Let us say that the amount of time it takes to solve the problem is more than a billion years.
  - Then the idea is that even if you HAVE all of the information in front of you, it takes too long to be able to actually solve it.
- To draw the distinction again..
  - in 1 world, people are infinitely smart. public key cryptography can not exist.
  - in the other, people are not infinitely smart. And so in the time they could decrypt the information, the sun will have collapsed, and so it is irrelevant.
- And so now that we know that, we will cover what is the mathematical function that they are going to be solving.

Reminder on Prime Numbers: there are integer prime numbers, that can only be divided by themselves and 1, e.g 2, 3, 5, 7, 11, 13.

- Here is the conjecture: If we pick two large prime numbers (and when we say large, we mean in the bazillions, NOT trillions).
  - And then we multiply them together.
  - We will get a number which is their product. Just prime1 * prime2. Just regular multiplication.
- We believe, if someone does this, and announces the product, that the task of figuring out the two constituent primes, is a very difficult task.
  - This is known as integer factorization, or just factoring.
  - If something is prime, you say it has no factors. If it is 18, the factors are 2 times 3 times 3.
  - So, if a number is the product of only two primes, both of roughly the same size, both are rather large, the problem of forming the factorization is very difficult.
  - All of modern cryptography rests on this assumption! That this is a hard problem to solve!
  - BUT we know that this is an EASY problem for quantum computers! So if we ever make quantum computers we'll have to do something different!
- And so now that we know that, we will cover what is the mathematical function that they are going to be solving.
- So what does the above have to do with what we have been talking about?
  - The first thing that people in the room utter is the product of the primes
  - Everyone knows the product, but only one brain in the room knows the two primes.

  - So, person 1 picks the two primes, multiplies them together, and speaks out the product of the two numbers.
  - And everybody takes out a piece of paper and writes that number down.
  - Then, whenever anybody wants to say something to person 1, they use that product, as the code, to encrypt what it is they want to say.
  - So, they take their phrase, for example the word "hello" and you pass it through person one's product, and out comes a string of gibberish.
    * turning that gibberish back into "hello" can only be done IF you know what prime 1 and prime 2 are. And the only person who KNOWS what prime 1 and prime 2 are is person 1!
    * So person 1 is the only person that can take the gibberish encoding of "hello" directed at them and turn it back into hello.
  - So person 1, announcing their product, means that now everyone else can SEND INFO to person 1 (and the only person who can receive anything at this point is person 1).
  - And so that product IS your Public Key.

- An alternative metaphor: You can imagine that Prime 1 and prime 2 are two chemical compounds. You combine them to make a new chemical compound, that can not be dissolved.
  - You can then coat things in the new chemical compound,

- So If you are using an old browser, than your "s" is being encoded with prime numbers that aren't very big, and consequently it isn't very secure.
  - So before you give the credit card information, you ask for the merchant's public key. It is like putting it in an envelope
  - So no one else can see "inside the envelope" so only the merchant can get it.
  - UNLESS they are really good at math, if they can do the prime factorization, then they can see your credit card info JUST as good as the merchant.
  - so the fundamental assumption behind secure, modern communications, is that the communication is secure under adversaries that are computationally limited.
    * Modern day systems would use prime numbers that are 1024 bits, numbers that are of the size 2 to the 1024th power. to give you an idea: 2 to the 300th power is the number of estimated particles in the Universe.
    * And this is WAAAAAAY more times that that!
    * But can be encoded in just a kilobyte!

BREAK

- Experiment: Everybody pick a large prime number:
  - Write it down silently
  - Everybody pick another large number
  - Write it down too-Now everyone multiply it together.
  - And now everyone announces that to the class.
    * Then you decide what you want to send. This would be just regular text. Maybe "hello Ben."
    * then you will do something with Ben's big number (his key) with "hello Ben" and that will give you a different big number, which is the encrypted message.
    * Then you tell Ben that big number, and he (and everyone) hears you say it.
    * And then Ben figures out how to decode it using his two prime numbers (that only he has)
    * If anyone else ever figures out those prime numbers, then the key is broken. It is like they are literally broken in half.

Gauss, smartest guy ever

Pardis teaches us how to successfully be unsuccessful in the class!

Slides Available on the course Piazza page: https://piazza.com/class/ii6jfojck806la?cid=13