Cryptography: Theory

Dr. Balázs Pejó

www.crysys.hu

Agenda

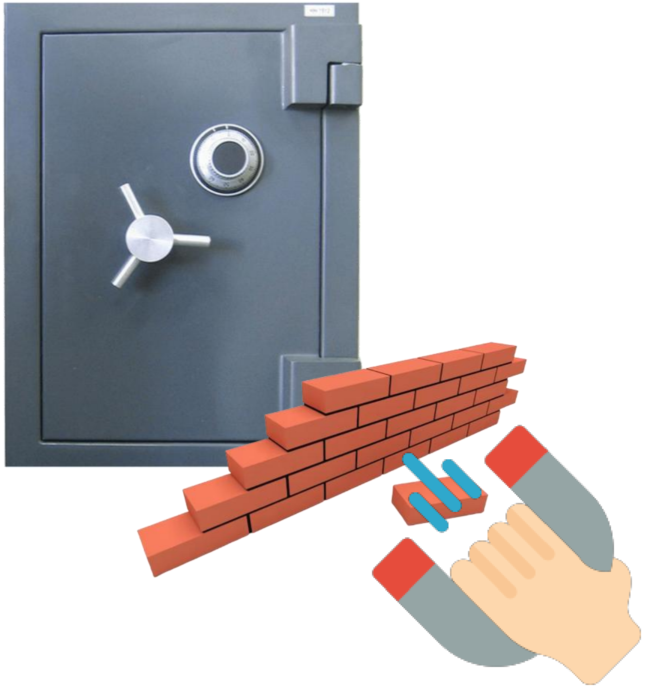
- Dark Patterns
- Tracking
- GDPR
- Machine Learning
- Deidentification
- Anonymization
- <u>Cryptography</u>

- Basics
  - CIA
  - (A)symmetric Crypto
  - Hash / MAC / Signatures
  - Diffie-Hellman / ElGamal
- Advanced
  - Cryptographic Commitments
  - Homomorphic Encryption
  - Secure Multiparty Computation
  - Secret Sharing
  - Private Set Intersection
  - Private Information Retrieval
  - Zero Knowledge Protocols
  - Post Quantum Cryptography
  - Functional Encryption

Recap

- Publishing data will lead to inevitable information disclosure.
  - There are many deanonymization attacks …
  - … as well as many anonymization techniques.

- Instead of this arm race,
  let's try to break this vicious cycle!
  - Ensuring that nothing is leaking
    with mathematical guarantees.
  - (Even Differential Privacy does not
    provide perfect protection,
    it merely limits the leakage.)

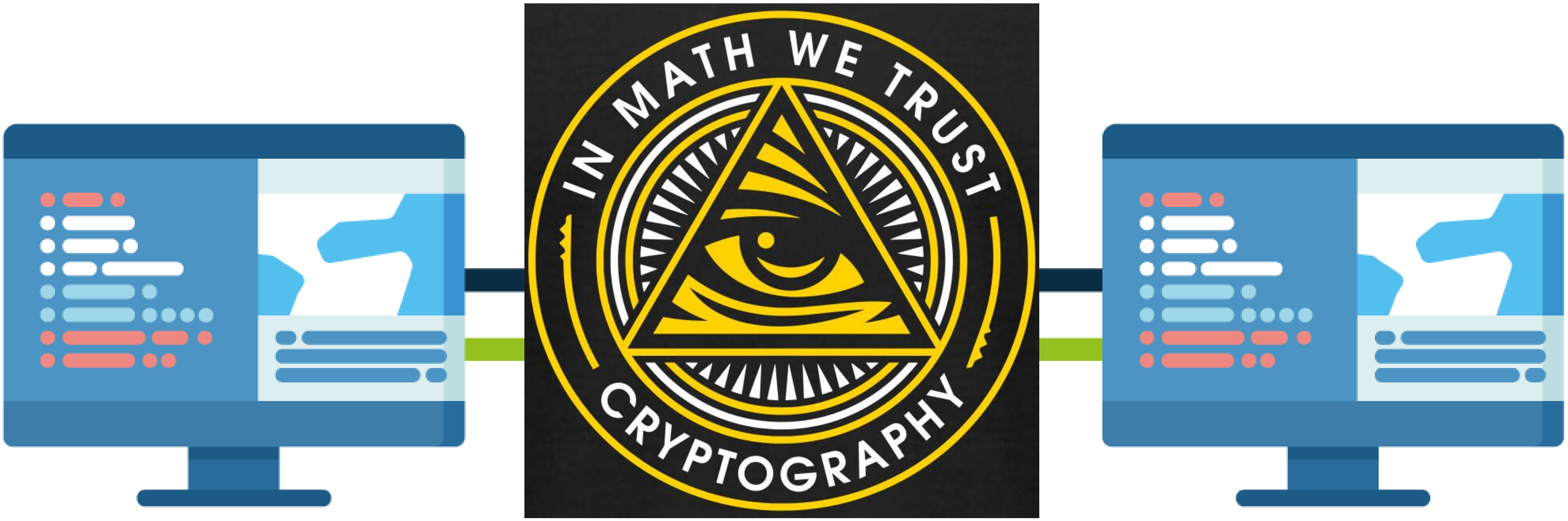


- Answer: encryption!

Cryptography 101

Privacy vs Cryptography

- Privacy is the RIGHT of an individual to control how information about him/her is collected, stored, and shared.
- Cryptography (old) is the practice and study of techniques for secure communication in the presence of adversarial behavior.
- Cryptography (new) is about replacing trust with mathematics.



CIA Triad

- A simple, yet comprehensive high-level checklist for the evaluation of your security procedures and tools.
- An effective system satisfies all three components: confidentiality, integrity, and availability.
- An information security system that is lacking in one of the three aspects of the CIA triad is insufficient.

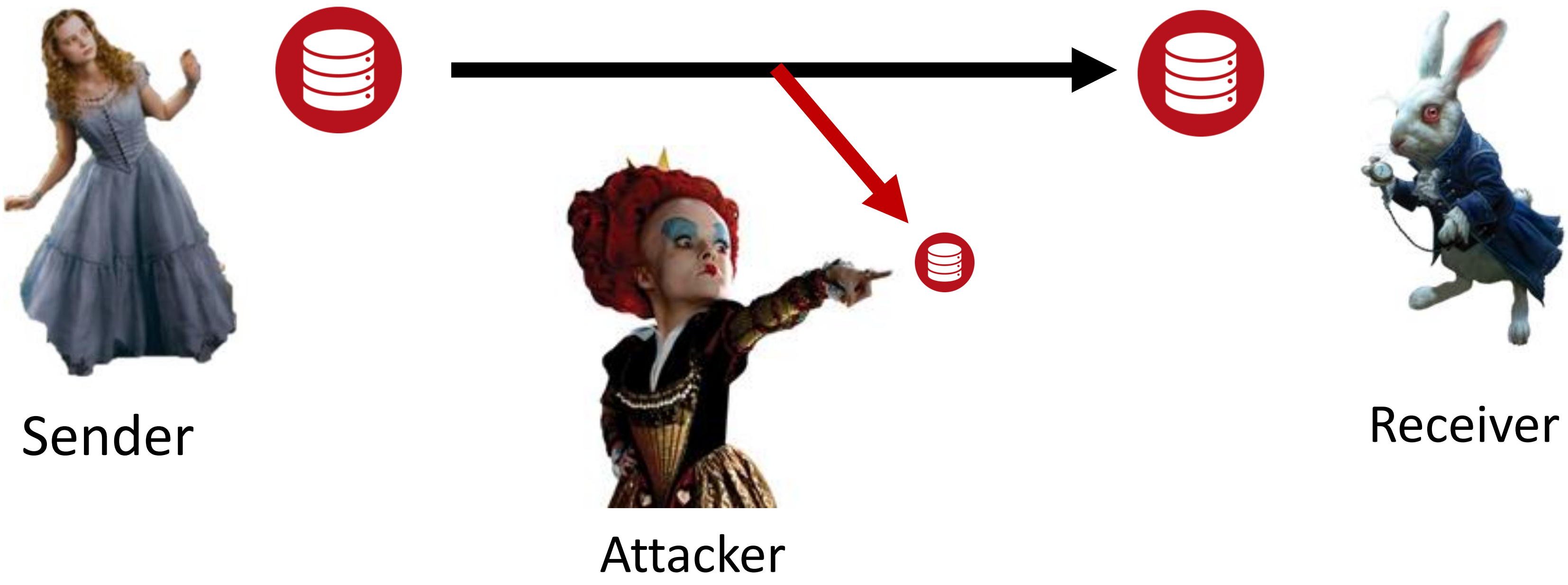
CONFIDENTIALITY
INTEGRITY
AVAILABILITY

- AAA Triad (for Access control):
  - Authentication
  - Authorization
  - Accounting

Confidentiality, Integrity, Availability

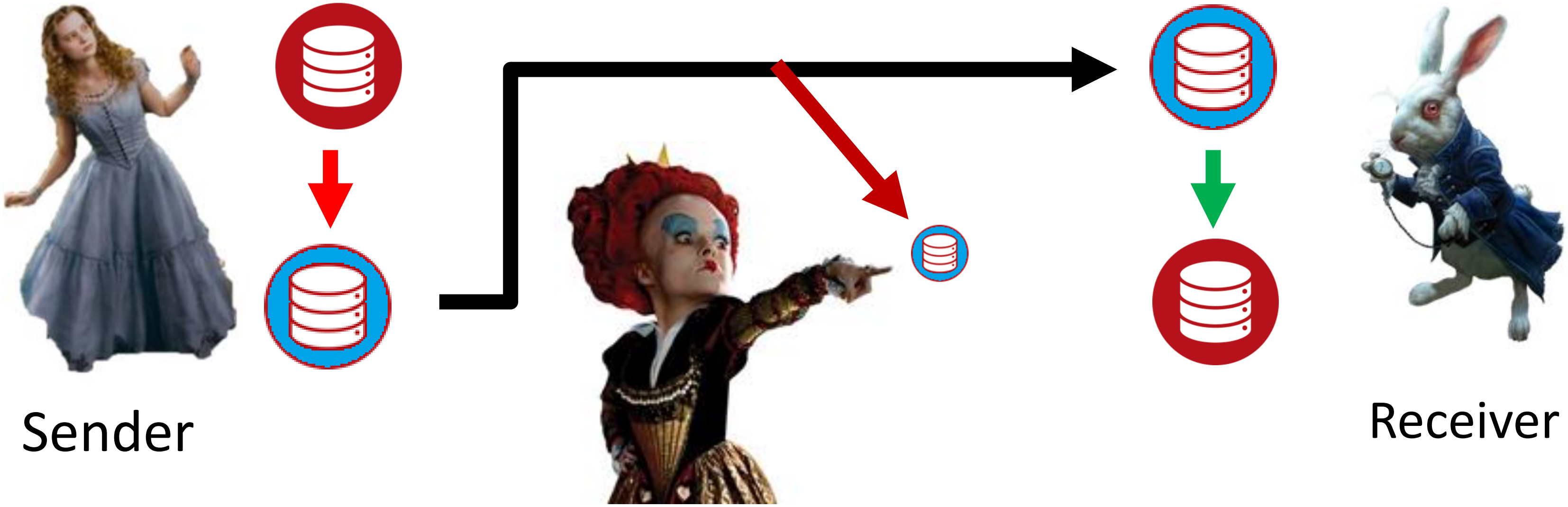
- Confidentiality measures are designed to prevent sensitive information from unauthorized access attempts.
  - Privacy can be achieved via confidentiality (and via other ways too).
- Integrity involves maintaining the consistency, accuracy, and trustworthiness of data over its entire lifecycle.
  - Data must not be changed in transit, and steps must be taken to ensure data cannot be altered by unauthorized people (for example, in a breach of confidentiality).
- Availability means information should be consistently and readily accessible for authorized parties.
  - This involves properly maintaining hardware and technical infrastructure and systems that hold and display the information.

Communication



- Alice wants to send (→) Bob a message.
- By eavesdropping (→), Eve could obtain the message if the channel is not secure.
  – Secure channels are very expensive to maintain; most communication happens on insecure channels.

Communication (Confidentiality)



Sender

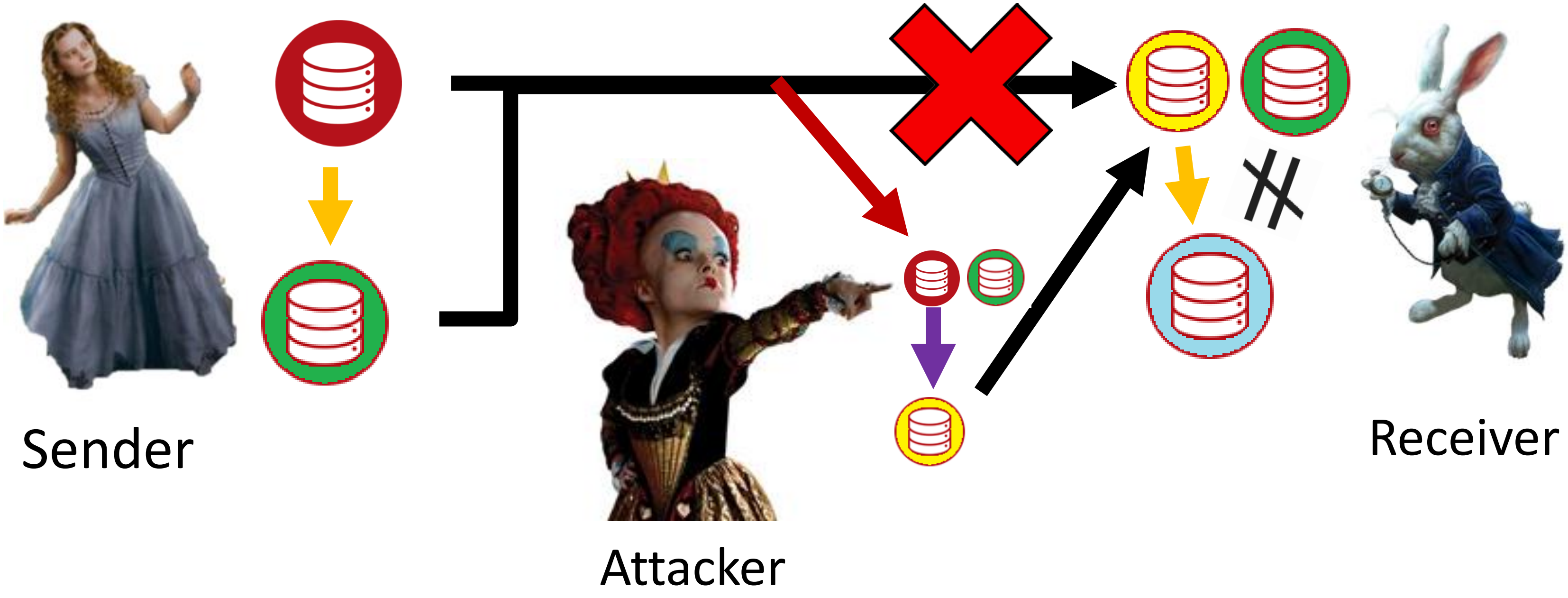
Attacker

Receiver

- Alice encrypts (→) the message which could be decrypted (→) by Bob.
- By eavesdropping (→), Eve could not obtain the original message.

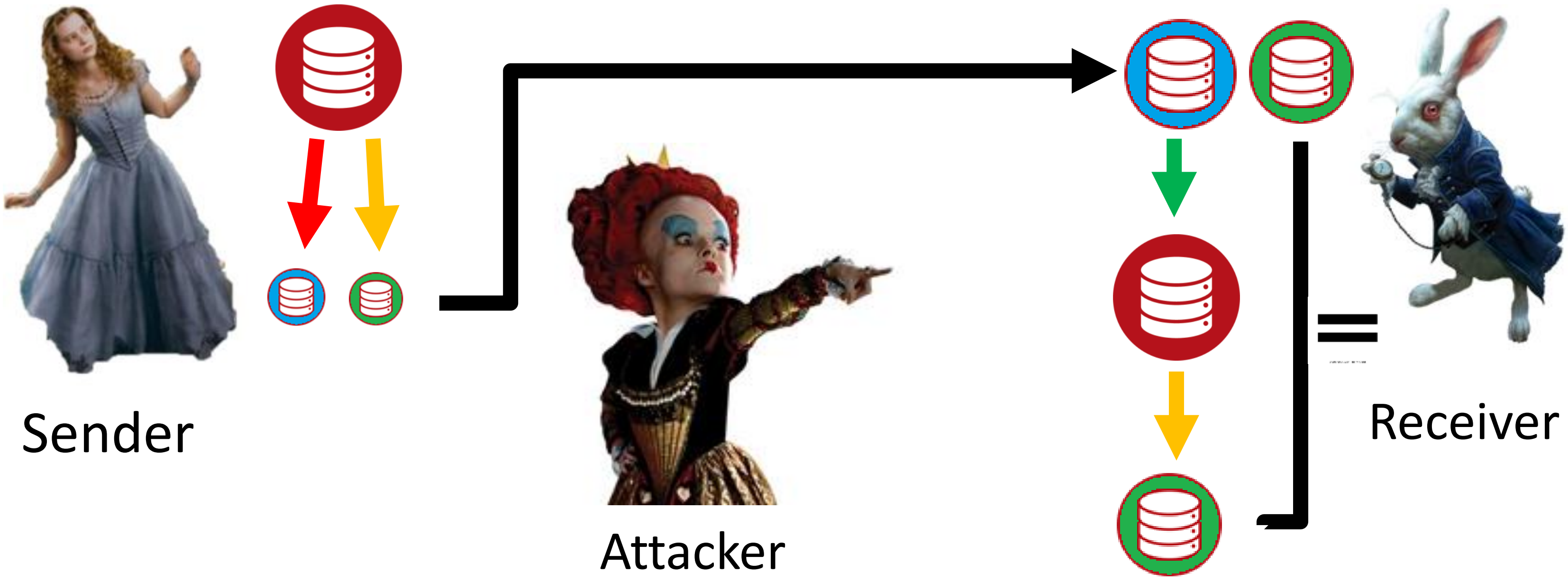
Plaintext ● Ciphertext ●

Communication (Integrity)

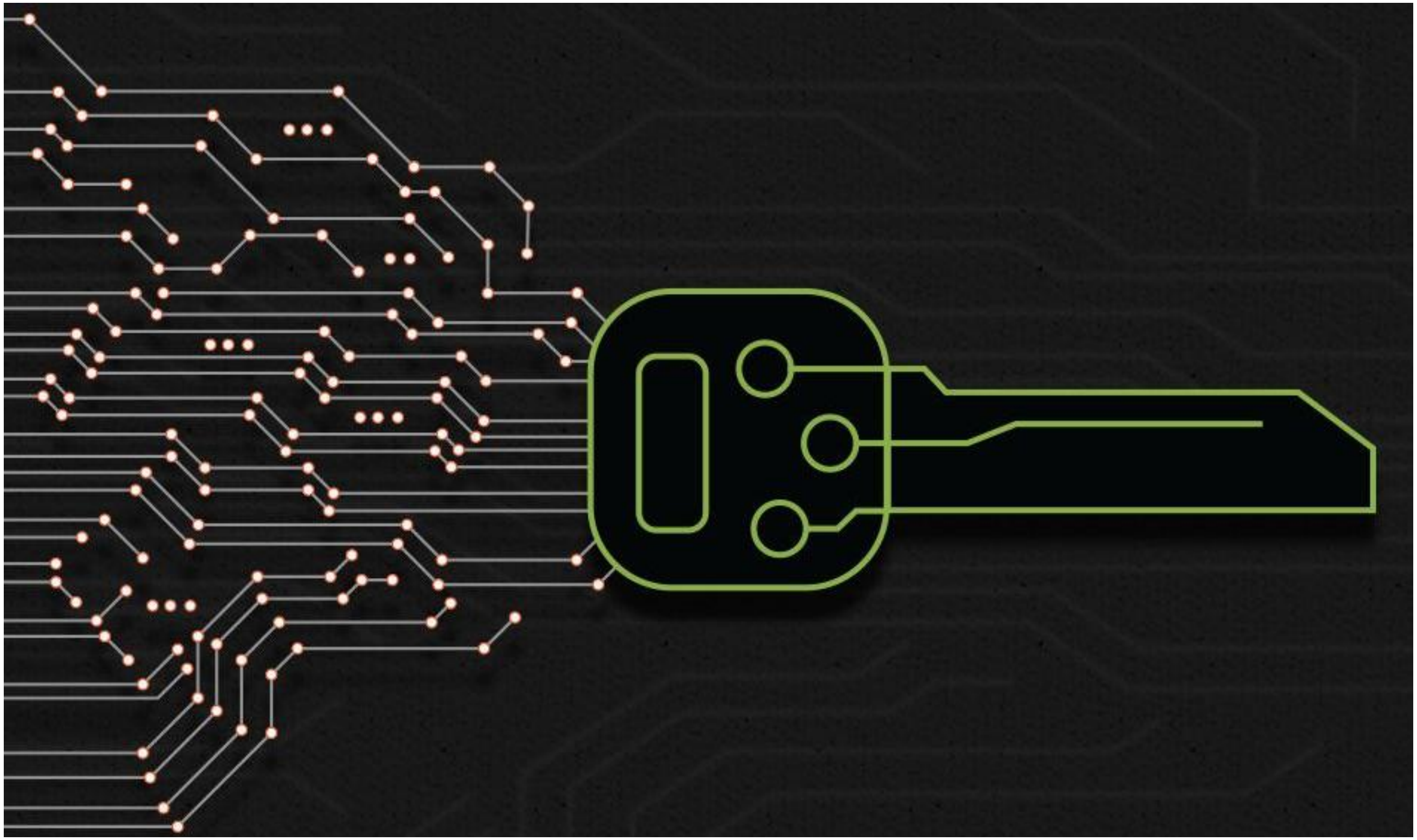


- Eve can modify (→) the message during transmission.
- How can Bob be sure, that the received message is the original?
  - Alice can append to the message a **M**essage **A**uthentication **C**ode (→), which ensures the integrity of the message.

Secure Communication



- With an encryption (→) and decryption (→) function and a MAC (→) function secure communication can be realized on an insecure channel.
  - How to realize these functionalities?

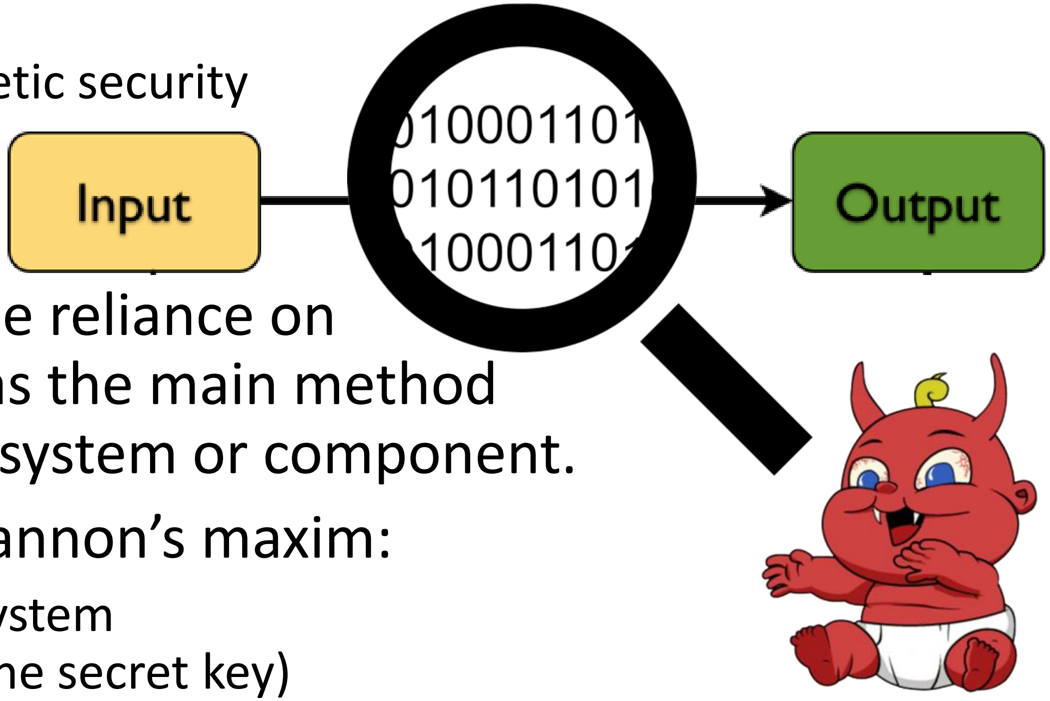


Basic Cryptographic Protocols

Basic model of Cryptography

- Attack models (for encryption)
  - ciphertext-only attack
  - known-plaintext attack
  - (adaptive) chosen-plaintext attack
  - (adaptive) chosen-ciphertext attack

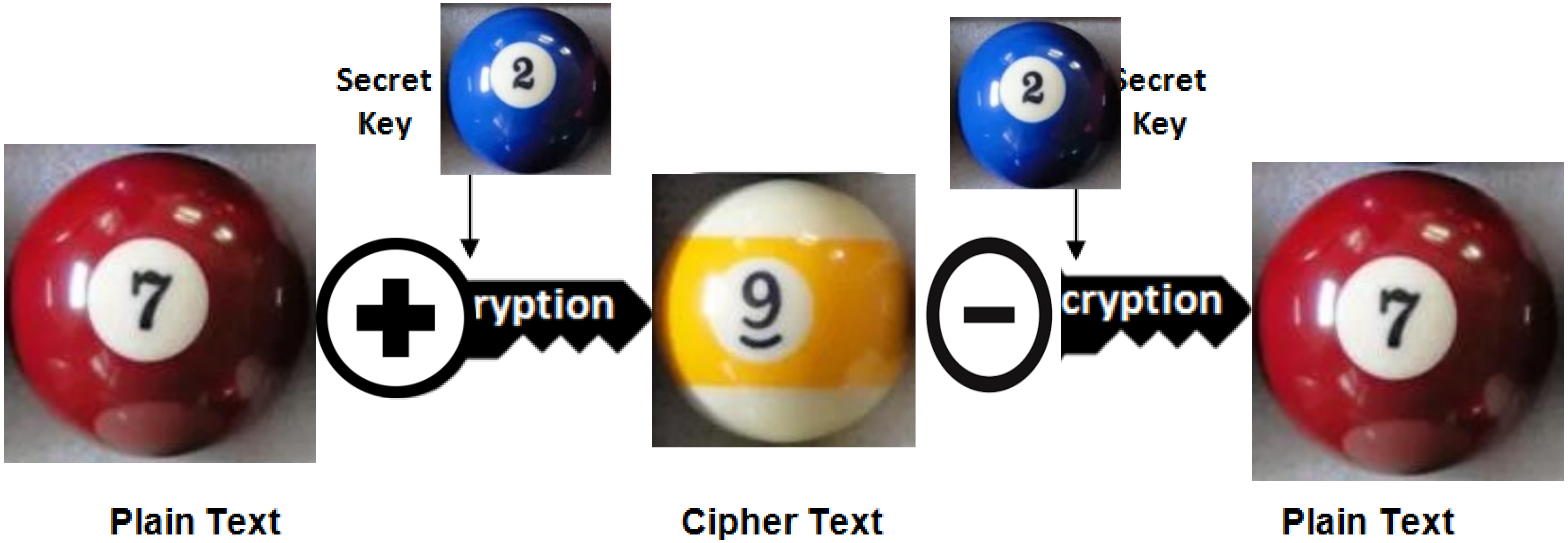
- Security requirements
  - perfect/information theoretic security
  - computational security



- Security by obscurity is the reliance on implementation secrecy as the main method of providing security to a system or component.
- Kerckhoff’s principle / Shannon’s maxim:
  - The attacker knows your system
    (except the message and the secret key)

Confidentiality: Secret Keys

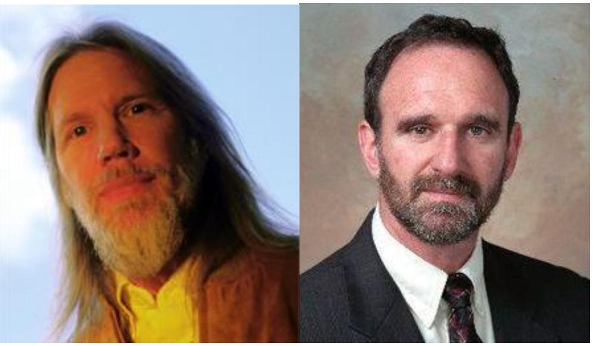
- Asymmetric Key Encryption: it uses two different key (private and public) to encrypt and decrypt the message.
  - Slow, but it provides confidentiality.
- Symmetric Key Encryption: the message is encrypted by using a key and the same key is used to decrypt.
  - Very fast, but it not provides authenticity.



Confidentiality: Diffie-Hellman

P=NP?

- Provably secure
  - Take a mathematical problem that is believed to be hard.
  - Show a cryptographic scheme/protocol.
  - Prove that breaking the scheme is not easier than solving the problem.



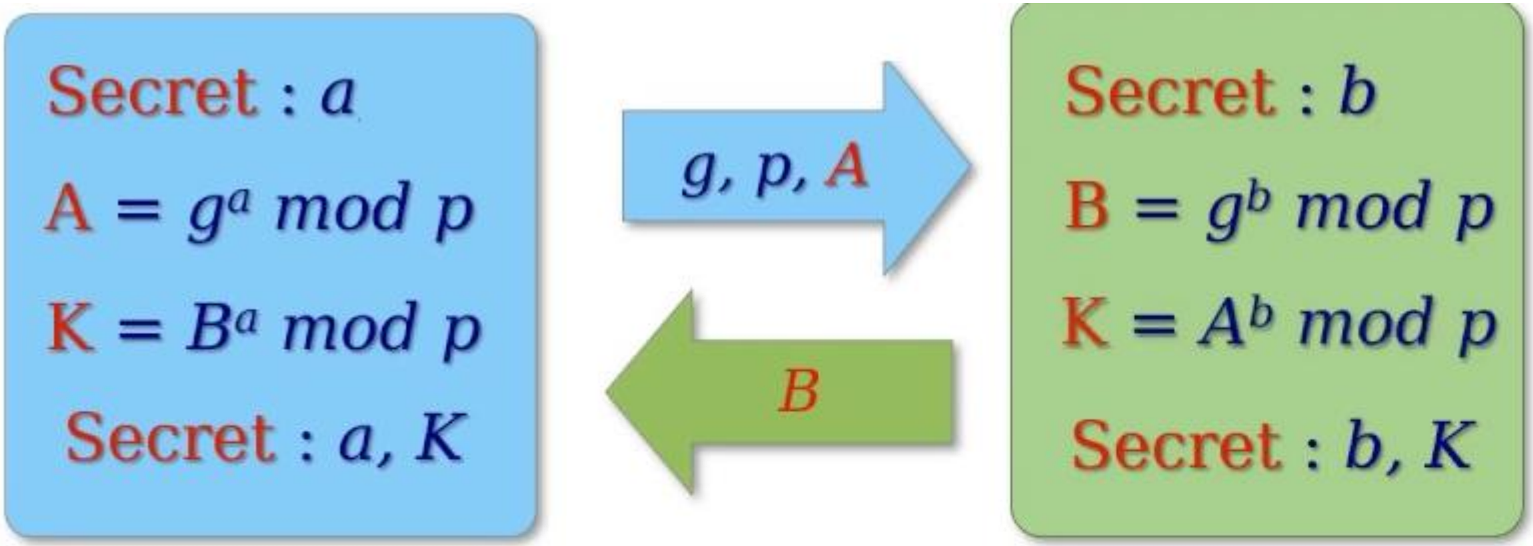
- Modular arithmetic: Z_p consists of numbers {0,1, 2, 3, … , p-1}
- g^x = y $mod\ p$

- Hard Problems
  - Discrete logarithm: to compute x from y.
  - Computational DH (CDH): to compute g^{xy} from g^x and g^y.
  - Decisional DH (DDH): to distinguish g^{xy} from a random element, if g^x and g^y are known.

Confidentiality: DH Key Exchange Protocol

- Alice selects a large p.
- g is the generator element of Z_p
  - $\forall$ x < p: $\exists$ i $\in Z$ s.t. g^i = x mod p

- Having seen A (= g^a) and B (=g^b), Eve cannot compute K (=g^{ab}) unless solving the CDH problem.



Confidentiality: ElGamal Encryption

- Setup as DH → $\{p, g\}$

- Alice selects a random $0 < x < p$ (Secret Key) and publish (p, g, g^x) (Public Key).

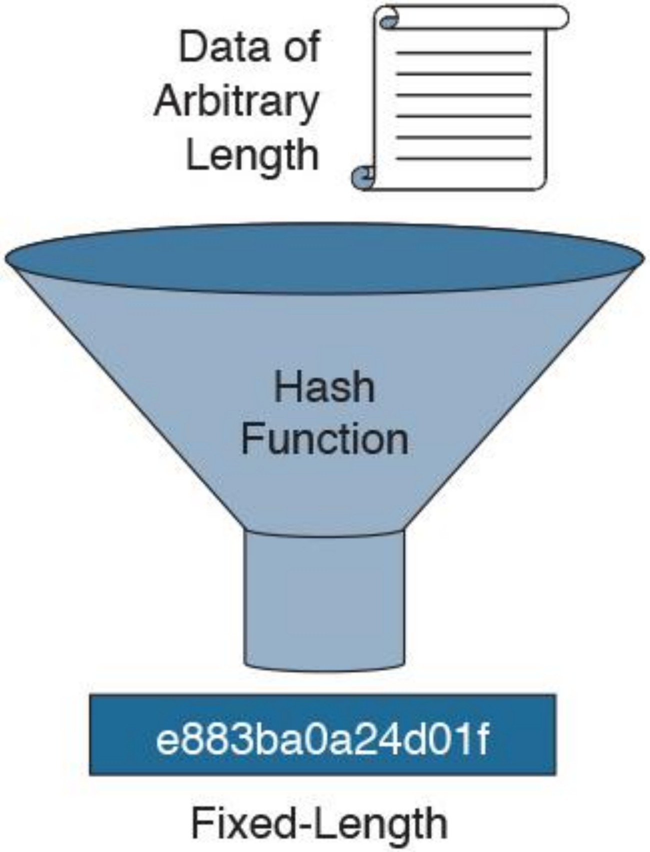
- Bob selects a random $0 < y < p$ and computes $(g^x)^y$.
- Bob encrypts and sends message: $(g^y, M \cdot (g^x)^y)$.

- Alice computes $(g^y)^x$ and decrypts $M \cdot (g^x)^y / (g^y)^x = M$.

- Instead of agreeing on a key in advance with a sender (like in DH), via ElGamal anybody could send message directly.

Integrity: Cryptographic Hash Functions

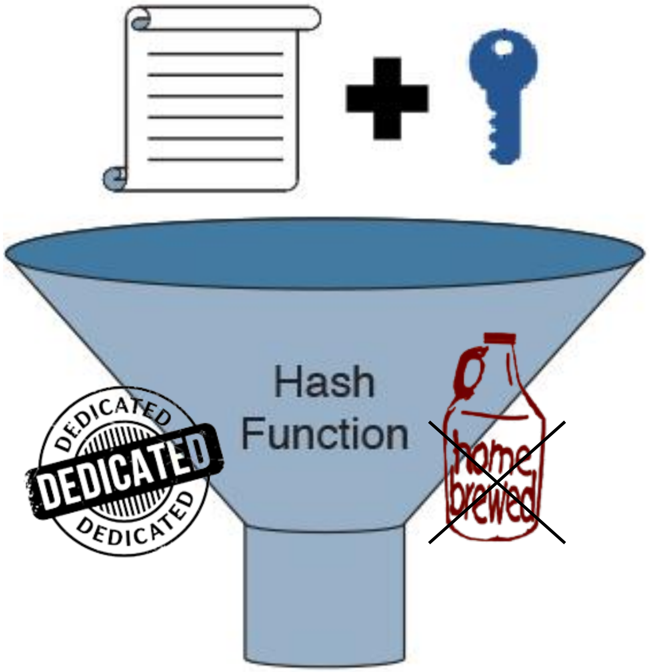
- Weak collision resistance given x, hard to find x’ s.t. H(x’) = H(x).
- Strong collision resistance: hard to find any distinct x and x’ s.t. H(x) = H(x’).
- Preimage resistance (one-way property): given y, hard to find x s.t. H(x) = y.
- Ease of computation.
- Collision resistant hash functions can be modeled as a random function.
- Serves as a compact representative image (fingerprint) of the message.



- Different form the hash functions used in data structures!

Integrity: Message Authentication Code

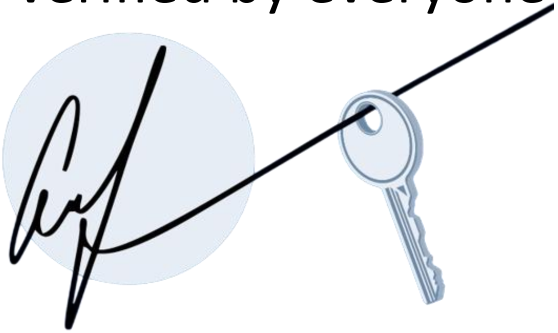
- MAC function maps an arbitrary long message and a shared key (k bits) into a fixed length output (n bits).
  - Can be viewed as a hash function with an additional input (the key).
- Message authentication: the receiver is assured that the message has been generated by the sender.
- Integrity protection: the message has not been altered in transit.



- Alice can deny the generation of the MAC.
- MAC is not verifiable by a third party without the key.

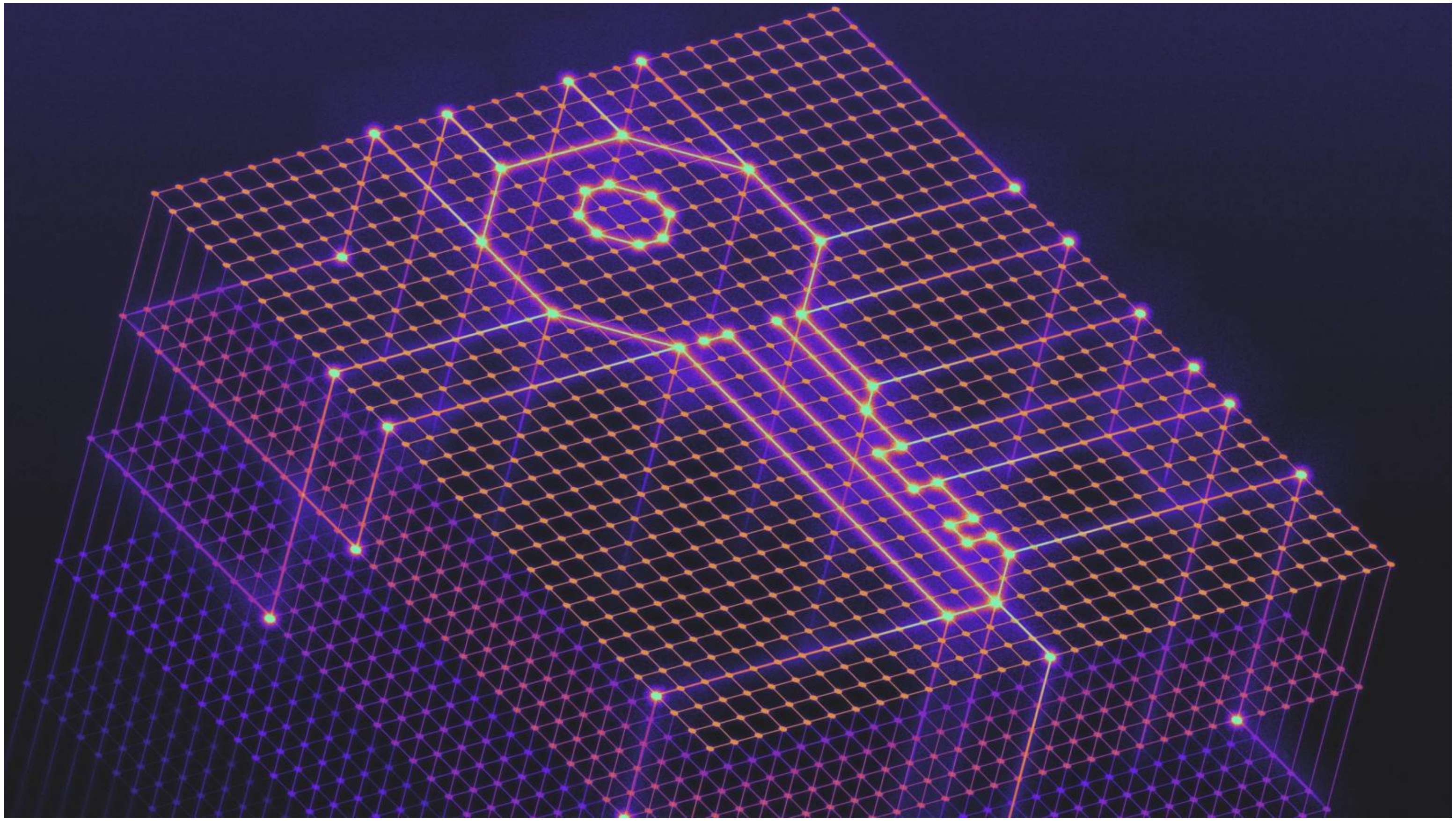
Integrity: Digital Signatures

- Instead of receiving a message which is encrypted with a public key (and could be decrypted with the private key), the message is signed with the private key which can be verified by everyone with the public key.



- Attaching such signature to the message ensures that
  - the message has been generated by the private key holder.
  - the message has not been altered.
- Non-repudiation of origin: the receiver can prove the origin of a message to a third party.

- How can Alice be convinced that the claimed public key of Bob indeed belongs to Bob?
  - Public Key Infrastructure (PKI)



Commitments

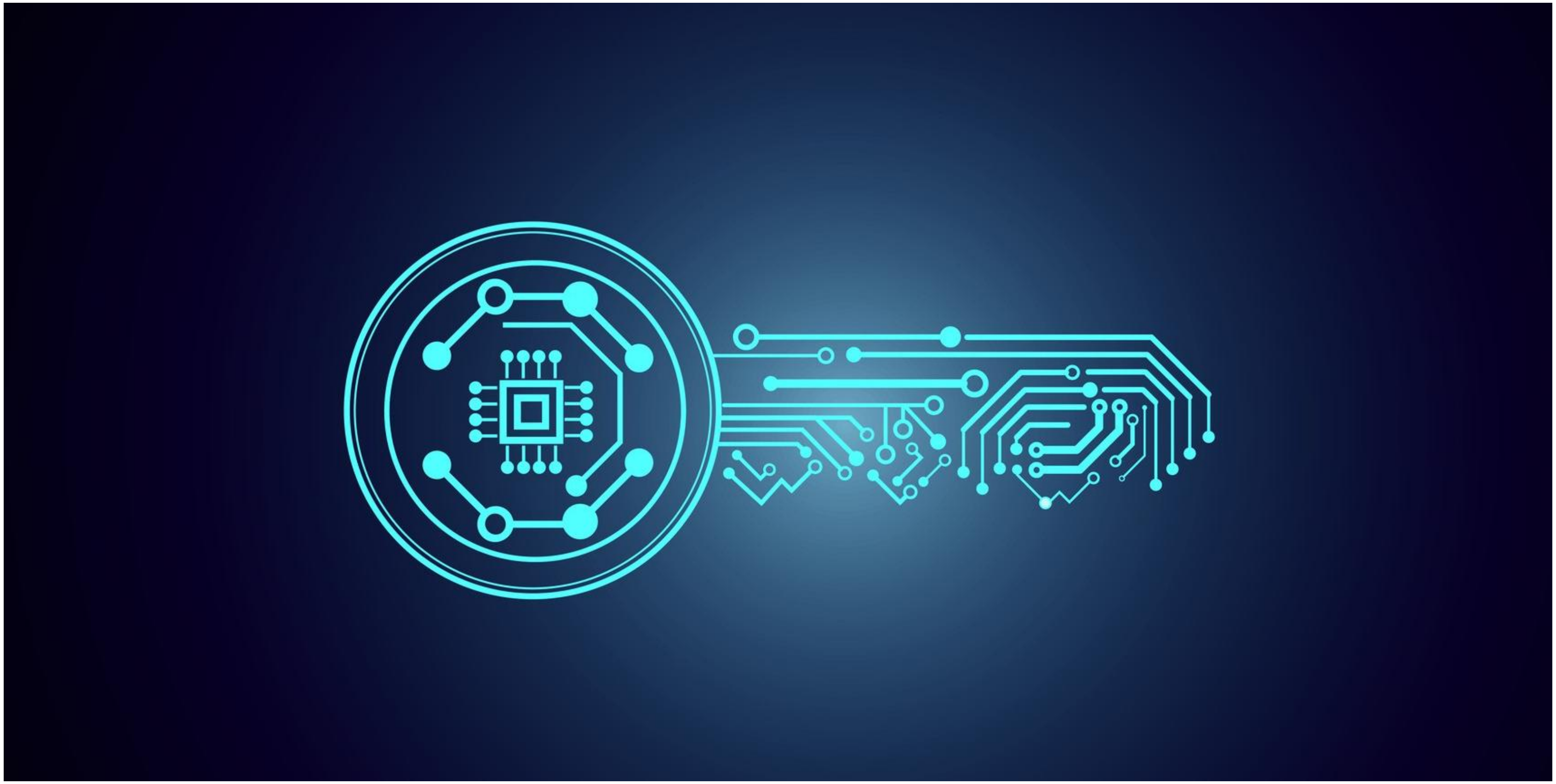
Illustration

Cryptographic Commitments

- There are two phases in the commitment protocol, the commit phase and the open phase.
  - In the commit phase, a prover makes a commit c to hide its secret m along with a random number r, i.e., c=commit(m, r).
  - In the open phase the prover reveals both the secret m and the random number r to a verifier, who can verify the validity of c.

- Hiding: in the commit phase, the verifier acquires no information about the secret m or about r from the commitment c.

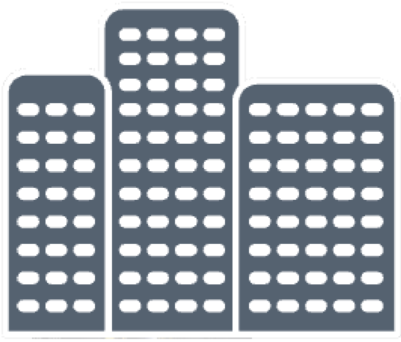
- Binding: it is computationally infeasible for the prover to open the commitment with other values than the corresponding secret m and random value r.



Homomorphic Encryption

Computation Outsourcing

$$x \rightarrow \quad f \quad \rightarrow f(x)$$

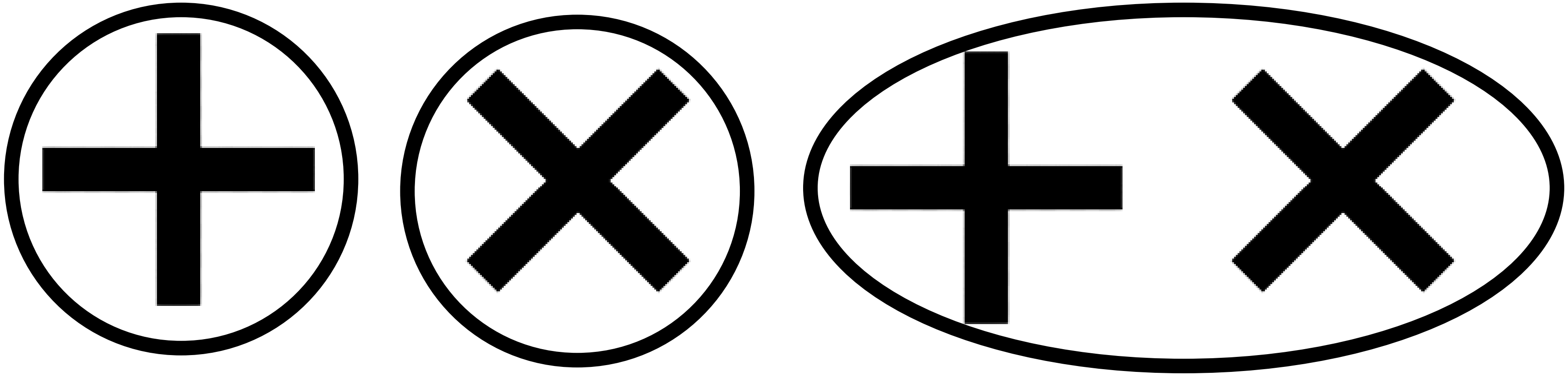


Illustration

Data Encryption

- In general, no operation can be performed on the encrypted (i.e., scrambled) data.
- Homomorphic encryption (HE) enables operations to be performed on the encrypted data and obtain the encrypted result, without ever accessing the data itself.
- Types and corresponding operations which could be performed on homomorphically encrypted data:

Partial (PHE) Somewhat (SHE) [<K] Full (FHE) [∞]



ElGamal PHE

- Multiplicative homomporphism
  - $\text{Enc}(M_1) \bullet \text{Enc}(M_2) = \text{Enc}(M_1 \cdot M_2)$

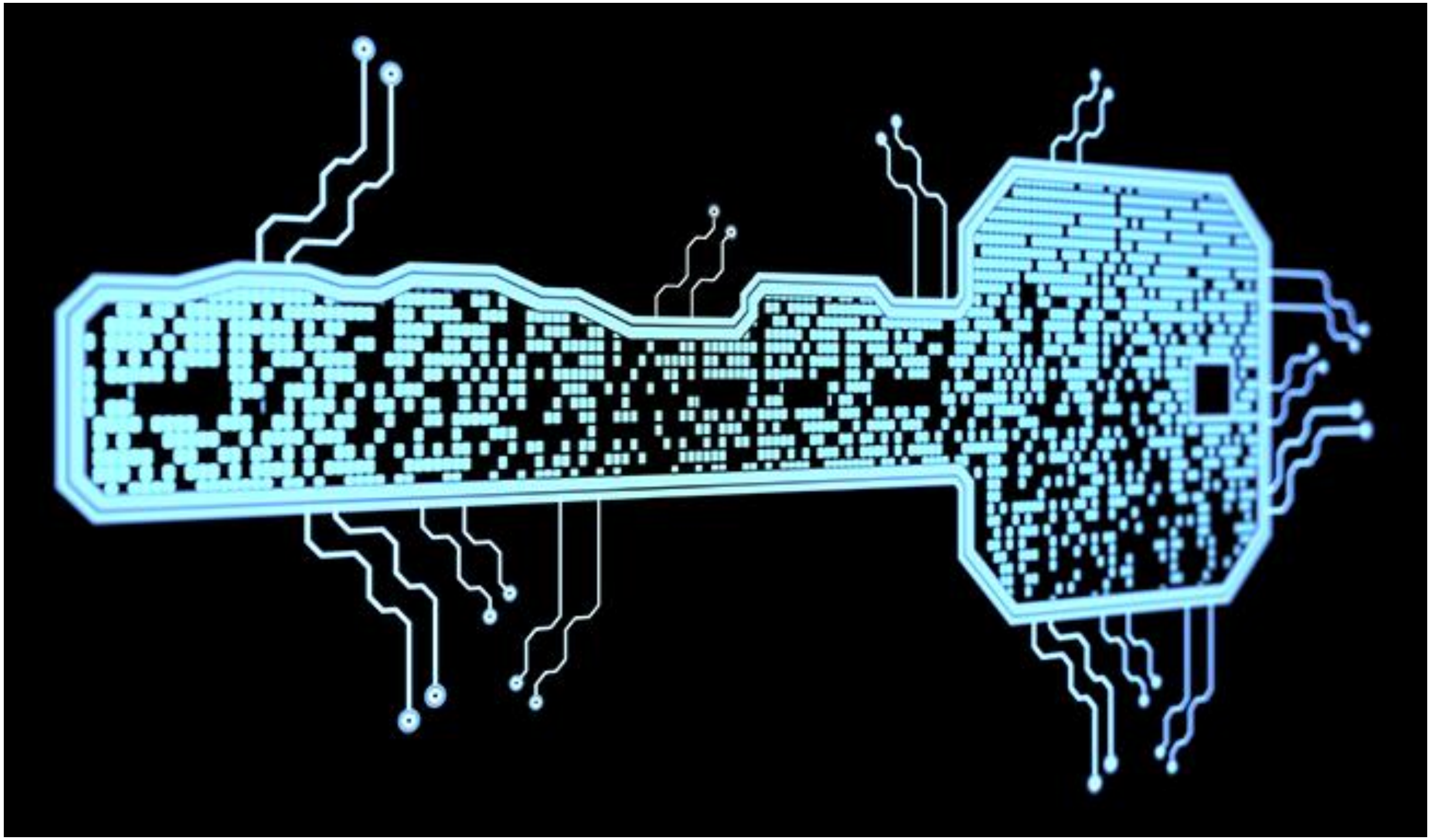
- Secret Key: x
- Public Key: (p, g, g^{x})

- $M_1 \rightarrow C_1 = \{g^{y}, M_1 \cdot g^{xy}\}$
- $M_2 \rightarrow C_2 = \{g^{z}, M_2 \cdot g^{xz}\}$

- $M_1 \cdot M_2 = C_1 \cdot C_2$
  - $g^{y} \cdot g^{z} = g^{y+z}$
  - $M_1 \cdot g^{xy} \cdot M_2 \cdot g^{xz} = M_1 \cdot M_2 \cdot g^{x(y+z)}$

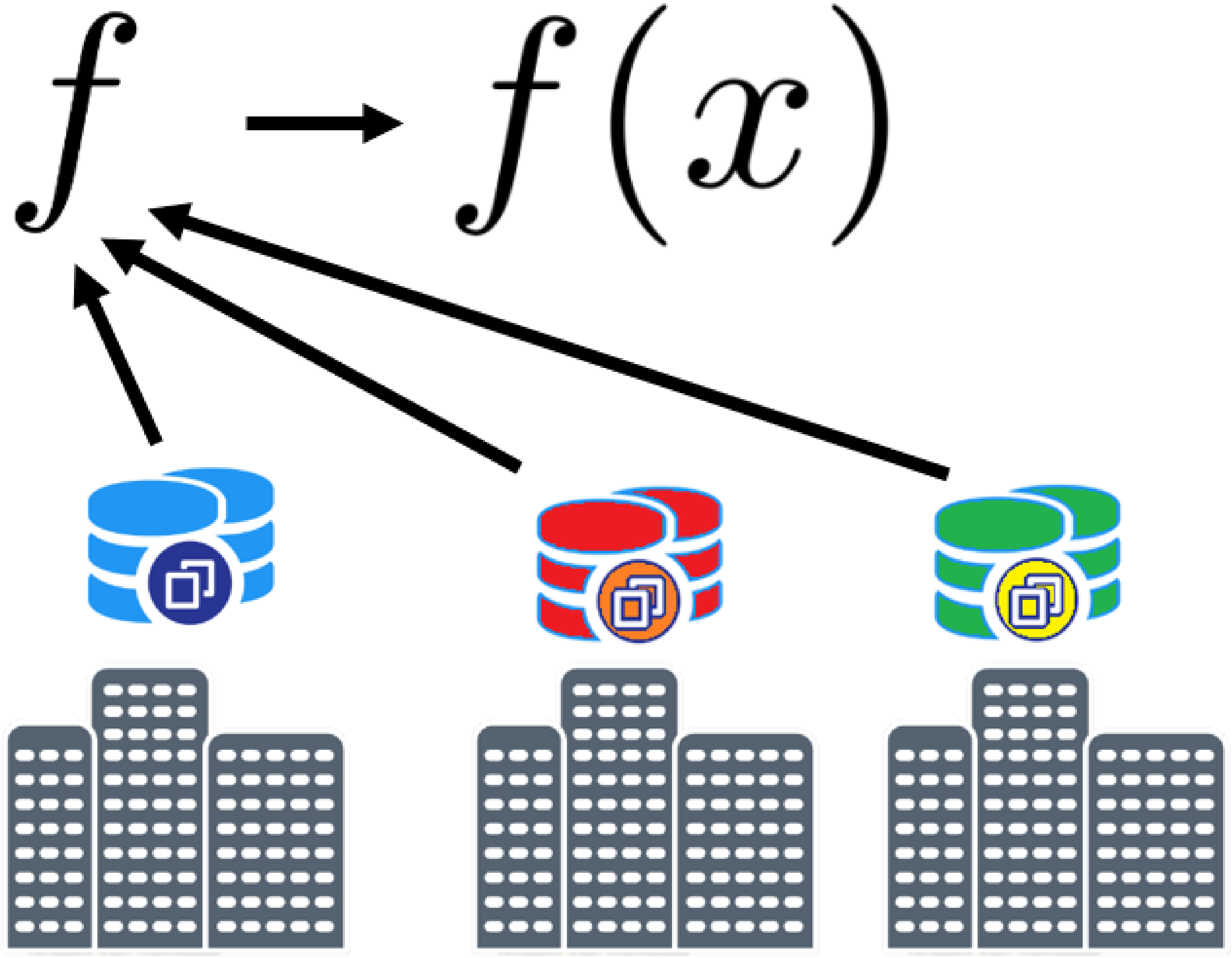
Craig Gentry

- PhD Dissertation
- FHE is possible



Secure Multiparty Computation

Multi-Party Computation



$$f \longrightarrow f(x)$$

Illustration

SMPC

- It works by enabling different parties, each with private data, to carry out a joint computation without needing to reveal their private inputs to each other.
- Although theoretically possible for decades, the computational power needed to efficiently carry out SMPC (and FHE) has only recently become available.
- In SMPC protocols the adversaries are the other participants (instead of third parties as with encryption).
  - Most common adversary models:
    Semi-honest (i.e., honest but curious) & Malicious

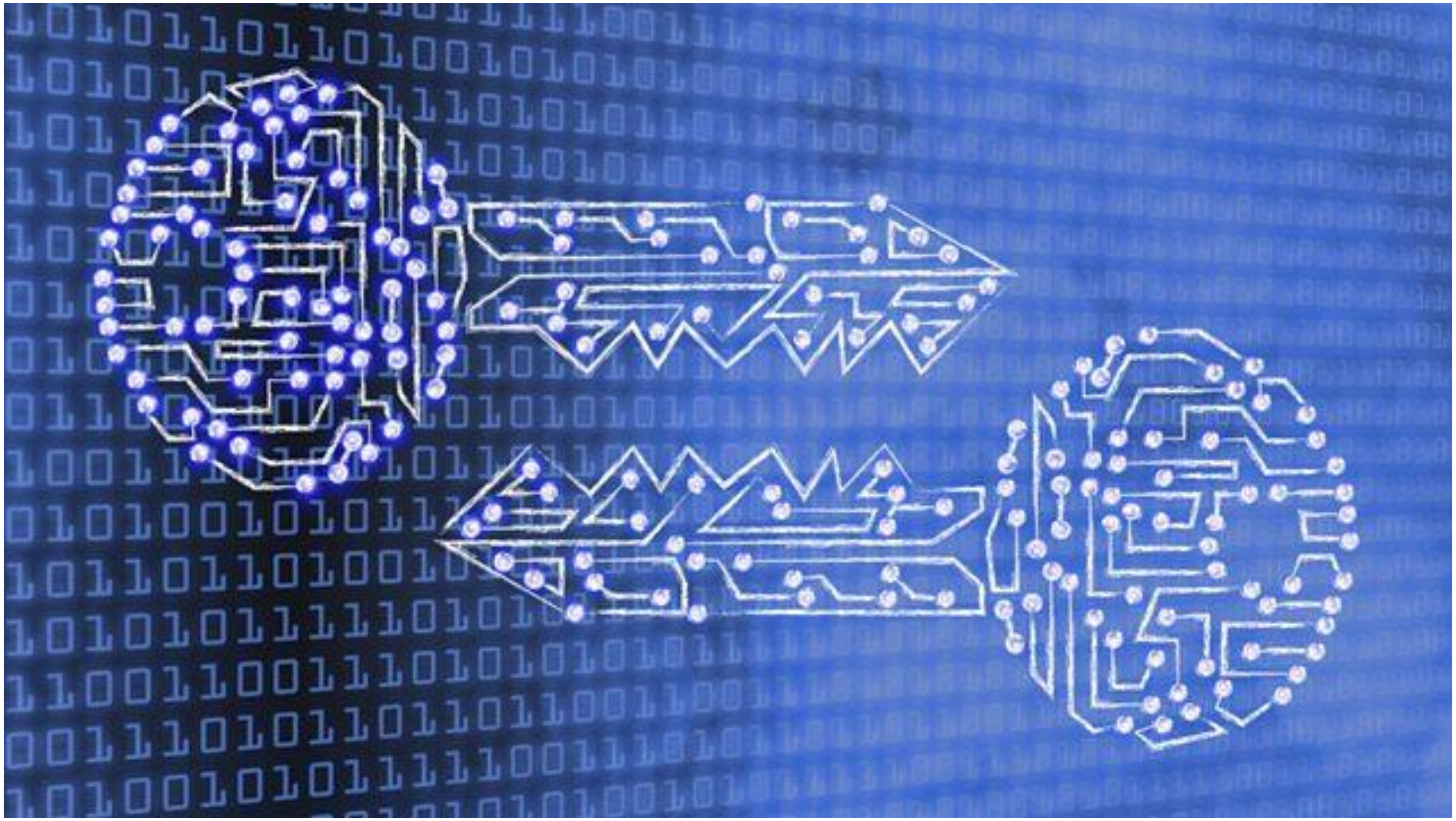
inpher · Novo Vivo · PRIVE COMMS · CYBAVO · ARPA

Secret Sharing

Illustration

K-out-of-N SS

- Secret sharing (SS) is a way to securely distribute important private information amongst a distributed network or group.
  - Secret sharing works by splitting private information into N smaller pieces and then distributing those shares amongst a group.
  - Each individual share is useless on its own but when any K shares are together, they reconstruct the original secret.
- Secure SS scheme distributes shares so that anyone with fewer than K shares has no more information about the secret than someone with no shares.
  - Shamir SS is information theoretically secure (i.e., secure against a malicious adversary with unlimited computational power) as reconstruction is not possible with fewer than K shares.
- Insecure SS allows an attacker to gain more information with each share.

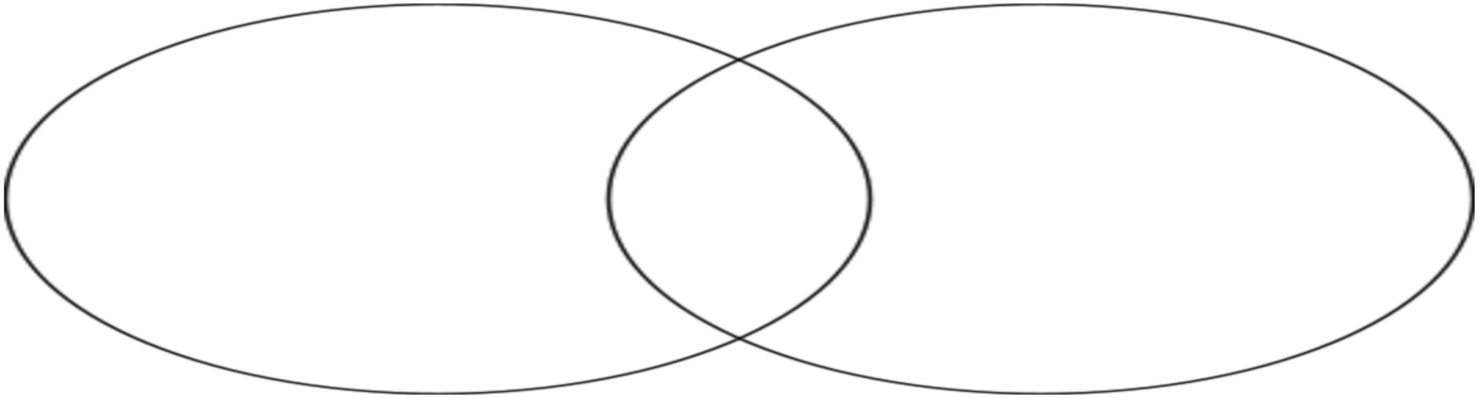


Private Set Intersection

Illustration

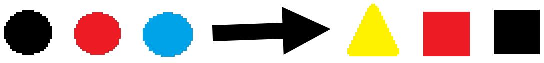
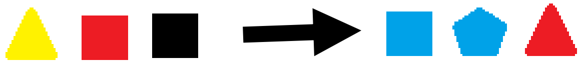
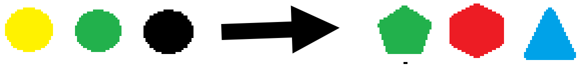
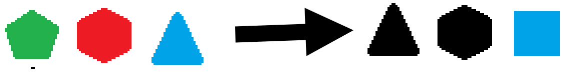
PSI

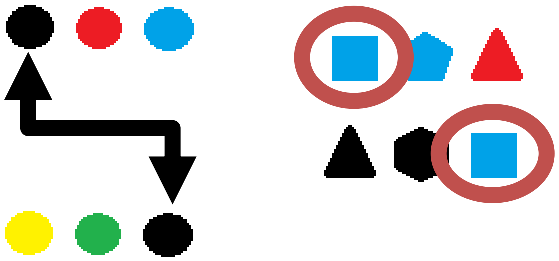
- Setup: Alice has list A, Bob has list B.
- The goal of the protocol is that Alice and Bob obtains the intersection A ∩ B, while no one learns anything about items that are not in the intersection.
- If a ∉ B then Bob learns nothing about it, i.e., Bob's apriori knowledge about whether Alice has item z, where z is not in the intersection, is not affected by the execution of the protocol.



Diffie-Hellman PSI

- Alice's input: $A = \{a_1, \ldots, a_n\}$;
- Bob's input: $B = \{b_1, \ldots, b_m\}$

- Alice select random x
  - Computes & sends $A' = \{a_1^x, \ldots, a_n^x\}$ 
- Bob select random y
  - Computes $A'' = \{a_1^{xy}, \ldots, a_n^{xy}\}$ 
  - Computes and sends $B' = \{b_1^y, \ldots, b_n^y\}$ 
- Alice computes and sends $B'' = \{b_1^{xy}, \ldots, b_n^{xy}\}$ 
- Bob intersects the two encrypted lists A'' and B''
  - For ∀ index i in A'' ∩ B'': $a_i = b_i \in$ in A ∩ B



- (Hash is insecure due to small input space.)

BONUS POINTS

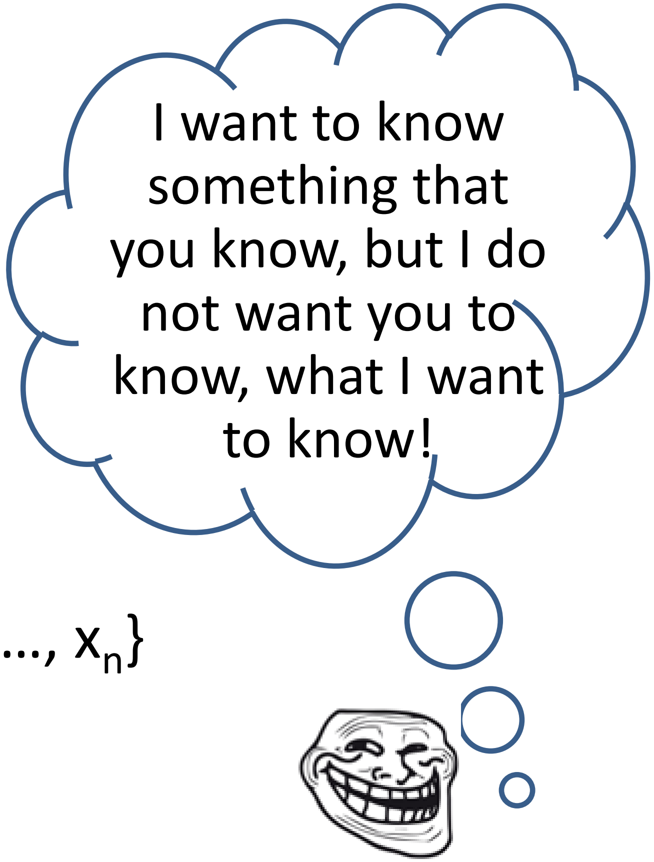
Crypto 1

Video

Private Information Retrieval

Problem Formulation

- Alice wants to obtain information from a (public) database, but she does not want the database to learn which information she wanted.
  - An investor querying a stock-market database.
  - A company querying a patent database.
- Trivial solution is to download the entire database.
- PIR solves this problem with less communication (less than O(n)).


I want to know something that you know, but I do not want you to know, what I want to know!

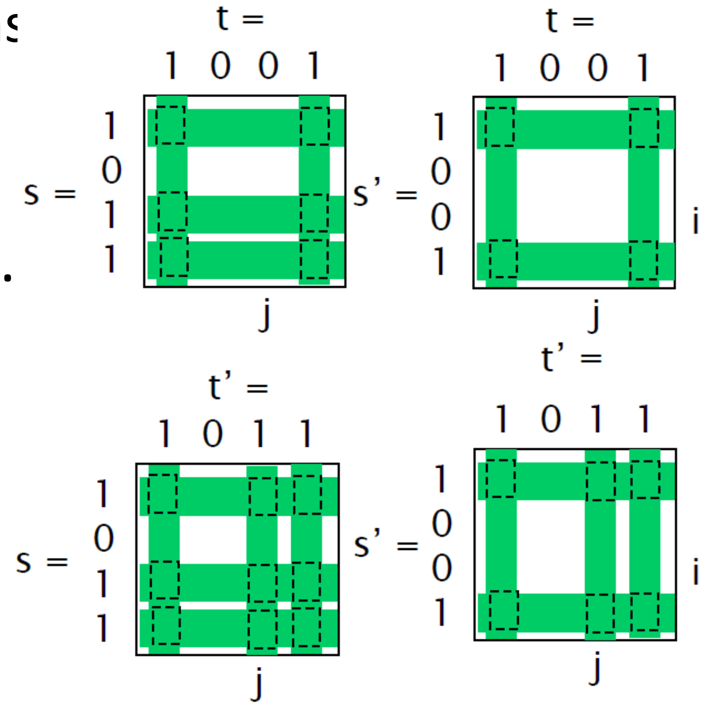
- The database is an n-bit string: $X = \{x_1, x_2, \ldots, x_n\}$
  - Alice is interested in x_i.
  - The database should not be able to learn i.

PIR Example

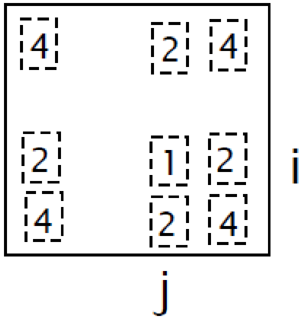
- Assume that there are 4 copies of the database; the bits of X are arranged in a √n · √n matrix.
  - Alice wants to retrieve x_{ij}.

$$\begin{array}{c|ccccc} & 1 & 2 & \cdots & \sqrt{n} \\ \hline 1 & x_{11} & x_{12} & \cdots & x_{1\sqrt{n}} \\ 2 & x_{21} & x_{22} & \cdots & x_{2\sqrt{n}} \\ 3 & x_{31} & x_{32} & \cdots & x_{3\sqrt{n}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sqrt{n} & x_{\sqrt{n}1} & x_{\sqrt{n}2} & \cdots & x_{\sqrt{n}\sqrt{n}} \end{array}$$

- Alice generates two random bit strings s and t of length √n and defines s' (t') such that they are the same as the bit in position i (j) is flipped.
- The selected bits are XOR-ed by the servers and the resulting bit is sent back.



- Alice receives a single bit from each server, these 4 bits are XOR-ed to get x_{ij}.



Zero Knowledge Protocols

Illustration

Problem Formulation

- A zero-knowledge protocol (ZKP) is a method by which one party (the prover) can prove to another party (the verifier) that something is true, without revealing any information apart from the fact that this specific statement is true.
  - Eliminates the need to reveal information to prove validity of a claim.

I want to tell you that I know something, but without telling you what I now.

- Non-Interactive ZKP (NIZK)
  - ZK-SNARK: Zero-Knowledge Succinct Non-Interactive Argument of Knowledge.
  - ZK-STARK: Zero-Knowledge Scalable Transparent Argument of Knowledge.

Properties

- Completeness: If the input is valid, the zero-knowledge protocol always returns true.
- Soundness: If the input is invalid, it is theoretically impossible to fool the zero-knowledge protocol to return true.
- Zero-knowledge: The verifier learns nothing about a statement beyond its validity or falsity.

- Classic ZKP is made up of three elements:
  - Witness
  - Challenge
  - Response

Functional Encryption

BONUS POINTS

Crypto 2

Video

Functional Encryption

- FE is a generalization of public-key encryption in which possessing a secret key allows one to learn a function of what the ciphertext is encrypting.

- While SMPC and HE could enable any operation, FE restricts the this for a specific functionality.
  - You can calculate with SMPC/HE the average, the sum, the median, etc.
  - With FE you can only compute the average (or sum, or median) if you have the corresponding function key.

Post Quantum Cryptography

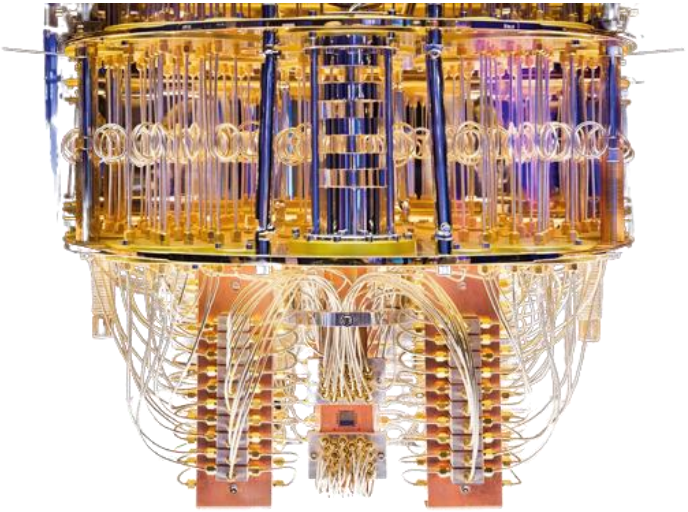
BONUS POINTS

Crypto 3

Video

Quantum Computers

- The problem with currently popular algorithms is that their security relies on one of three hard mathematical problems: the integer factorization problem, the discrete logarithm problem or the elliptic-curve discrete logarithm problem.
- These could be easily solved on a sufficiently powerful quantum computer running Shor's algorithm.
  - Luckily no such machine exists today!



- PQC is the development of cryptographic algorithms that are thought to be secure against a cryptanalytic attack by a quantum computer.

Problems

- Assume a large quantum computer can be built in 20 years.
  - We are already in trouble: what if the encryption should last for 20+ years?
- The National Institute of Standards and Technology (NIST) envisioned to transition to PQC schemes before 2035.
  - The current schemes are known to be secure as they are tested for decades.
  - Shifting to new schemes which were not tested enough might make our systems more vulnerable against currently possible attacks while making them resistant against future attacks which might never be possible.
- NSA has influence is selecting standards, and they have misaligned incentives (i.e., they want weak system so they can spy).
  - This happened already in the past with the DES encryption scheme.

Take Away

- The attacker is assumed to know your system (except your private message and secure key).
- Confidentiality and Integrity are key objectives of cryptography.
  - Encryption could tackle the former,
    MAC and Digital Signatures could tackle the latter.
- Cryptographic hash functions could be suitable building blocks of MACs and Commitment schemes (but do not use self-made).
- While HE enables operations on the ciphertext, SMPC enables joint operations amongst several participants (such as PSI).
  - FE restricts what calculation is allowed on the encrypted data.
  - SS shares secrets such that only the right combination would reveal it.
- PIR aims to hide the accessed information, while ZKP aims to prove an information without revealing that.
- Post Quantum Cryptography work on traditional systems but also secure against quantum computers.

Control Questions

- What is the CIA triad, and how symmetric and asymmetric cryptography differs?
- What is SMPC, what other cryptographic primitives can be used to realize it, and how does it differ from FE?
- What is PQC, and what problems does it have?

References

- CIA
- Symmetric Cryptography
- Asymmetric Cryptography
- Commitments
- Homomorphic Encryption
- Secure Multi Party Computation
- Secret Sharing
- Private Set Intersection
- Zero Knowledge Proofs