FUJITSU

shaping tomorrow with you

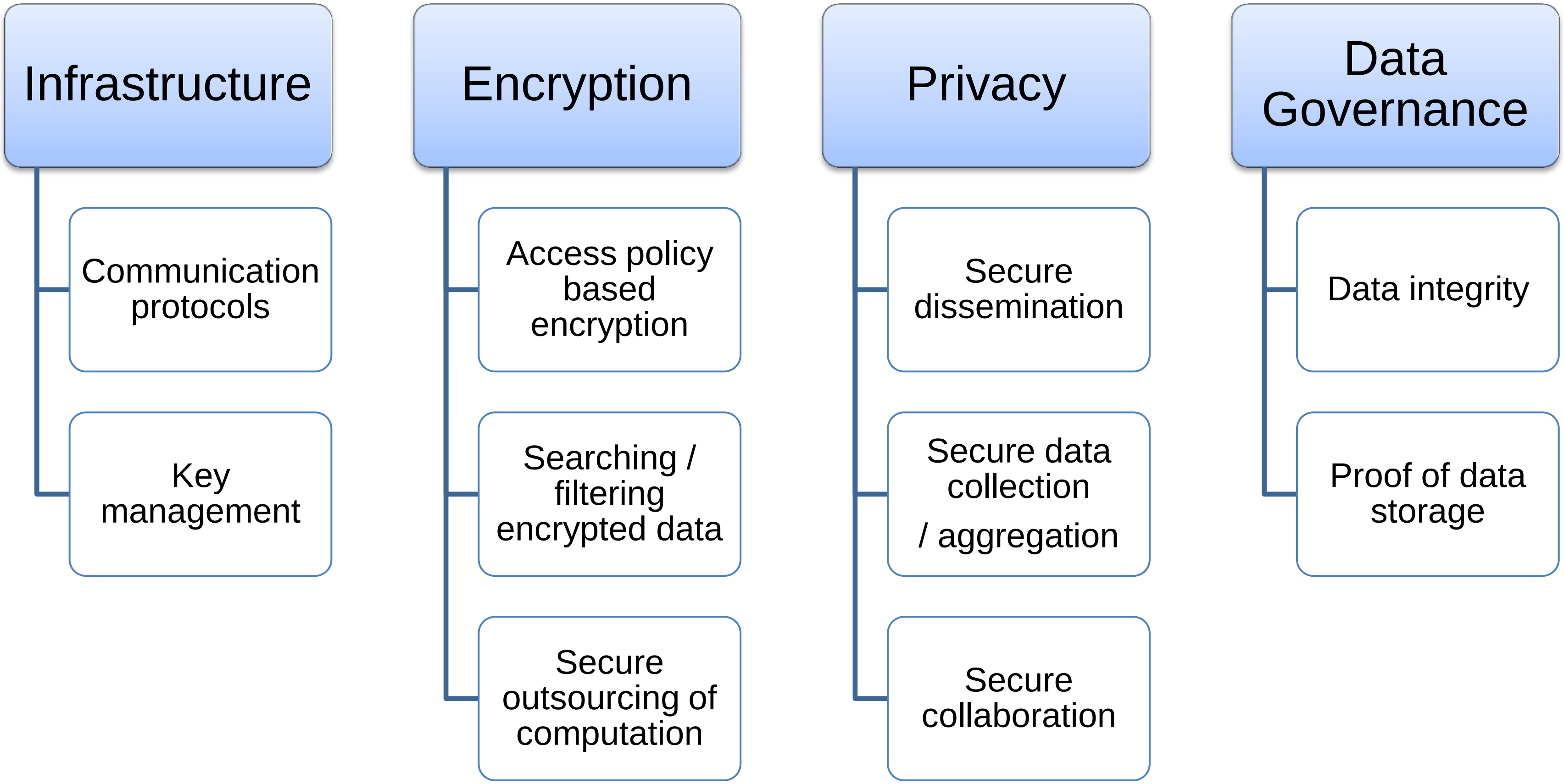
Top 10 Challenges in Cryptography for Big Data

Arnab Roy
Software Systems Innovation Group
Fujitsu Laboratories of America

How is cryptography for Big Data different?

- BIG
  - Scale up existing solutions for volume, variety and velocity
  - Retarget to Big Data infrastructural shift
- DATA
  - Balance privacy and utility
  - Enable analytics and governance on encrypted data
  - Reconcile authentication and anonymity

Top 10 Challenges identified by CSA BDWG

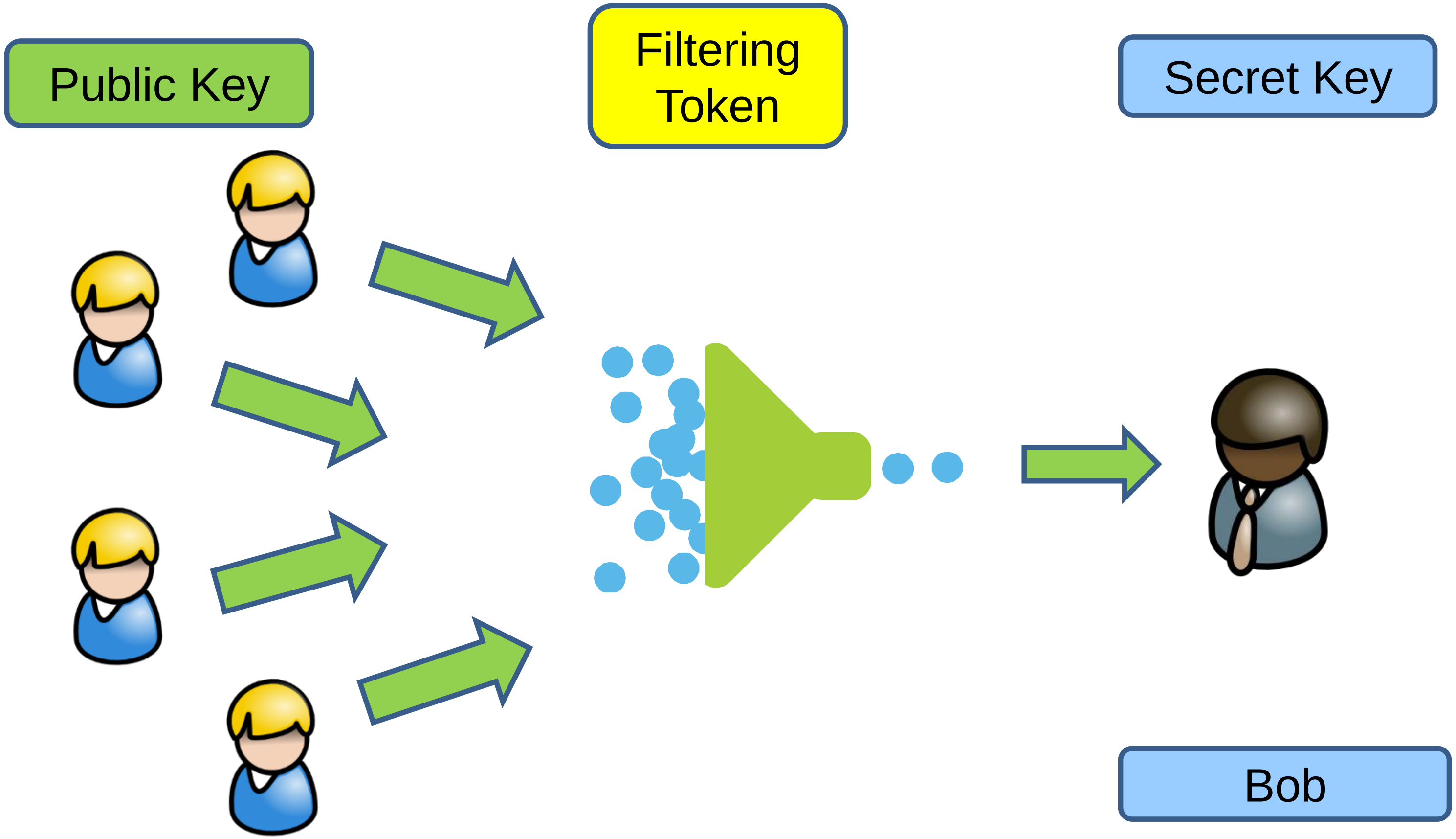


- Infrastructure
  - Communication protocols
  - Key management
- Encryption
  - Access policy based encryption
  - Searching / filtering encrypted data
  - Secure outsourcing of computation
- Privacy
  - Secure dissemination
  - Secure data collection / aggregation
  - Secure collaboration
- Data Governance
  - Data integrity
  - Proof of data storage

Data Privacy

- Anonymization is not privacy!
- Suppose, there is a public movie database
  - User Joe Smith likes movies U, V, W
- Suppose, there is a private movie database
  - Anonymous likes movies M, N, U, V, W, X, Y
- For rare combination of movies U, V, W, it becomes extremely unlikely that someone other than Joe Smith likes the same exact 3 movies
  - So, it is likely that Joe Smith likes movies M, N, X, Y as well
  - This may reveal private information about Joe, like his political or religious inclinations
- This is the type of exploit used in de-anonymizing the Netflix dataset by Narayanan and Shmatikov

Searching and Filtering Encrypted Data



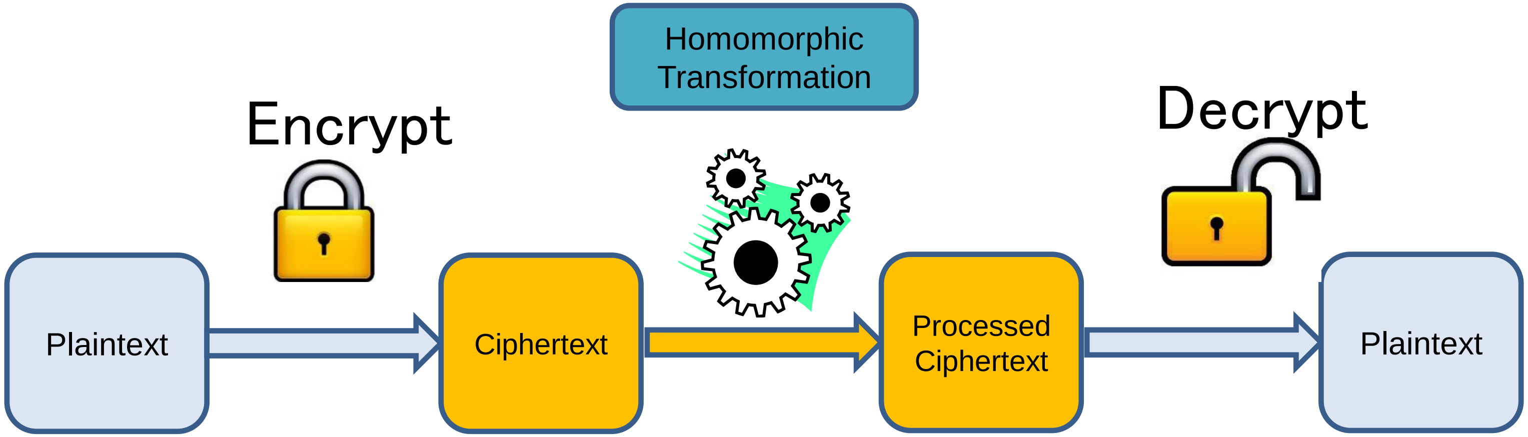
Public Key

Filtering Token

Secret Key

Bob

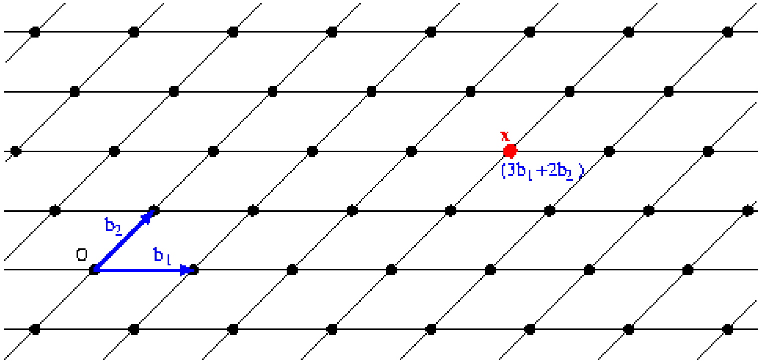
Fully Homomorphic Encryption (FHE)



- With FHE, computation on plaintext can be transformed into computation on ciphertext
- As a use case, a cloud can keep and process customer’s data without ever knowing the contents
  - Only customer can decrypt the processed data
  - End to end security of customer data

How does FHE work?

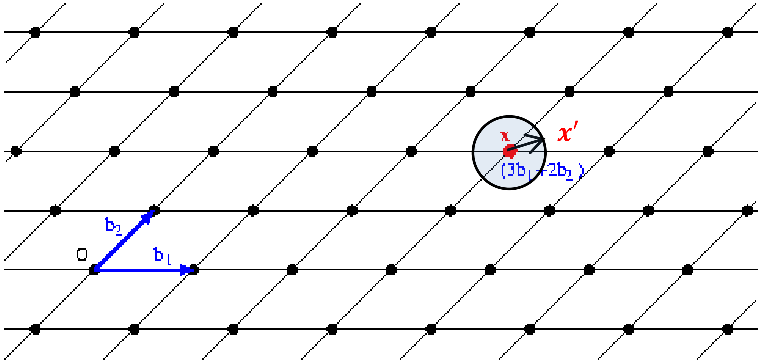
- Intuition:
  - Represent programs as circuits
    - Sequence of additions and multiplications
  - Transform the input data to a high dimensional ring (popularly, lattices)
    - Exploit ring homomorphism with respect to +,×



Source: http://cseweb.ucsd.edu/~daniele/lattice/lattice.html

How does FHE work?

- How do we ensure that the transformed representation hides the plaintext?
  - Solution: add some noise to the representation
  - In sufficiently high dimensions, it is considered hard to derive the closest lattice point, when noise is added
- Now, we have a different problem
  - With each +,×, noise grows!
  - At some point, data may be irrecoverable
  - Solution: noise reduction techniques
    - Bootstrapping, Modulus switching



Source: http://cseweb.ucsd.edu/~daniele/lattice/lattice.html

■ Securely Retrieving Sensitive Information via Encrypted String Search Technology

Booth 24

FUJITSU

Advanced Technology **Demonstration Only**

■ Software Technologies Laboratories

■Enabling Searches of Text Patterns in Encrypted Data without Decryption

■We have developed searchable encryption technology that enables the searching of encrypted string data for specific string patterns without the need for decryption. The technology makes it possible to safely search DNA information, biochemical information, medical information, and other kinds of highly-sensitive data via the cloud.

Key Features

- **Searches on data encrypted with homomorphic encryption**
  Using the confidential matching functions of homomorphic encryption, makes it possible to search on any text, even text strings without preset keywords, so that encrypted text strings can be searched (4 patents pending).
- **Fast searching of arbitrary text strings using batch-mode computations**
  Using Fujitsu's proprietary batch-mode encryption and inner-product calculations, this technology can search 10,000 characters of encrypted text per second (world's fastest).

Anticipated Customers and Services

- Systems that need to search while maintaining the secrecy of cloud databases
- Including: drug R&D with research-subject records, personal medical information, etc.

Anticipated Availability

- Technology applications are now being studied.
- Aiming for commercial implementation in 2015

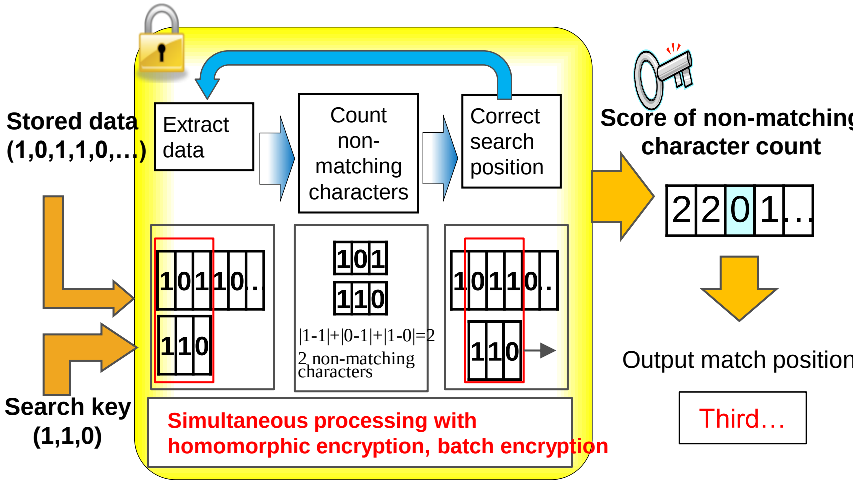


Figure 1: Principles of Encrypted String Search

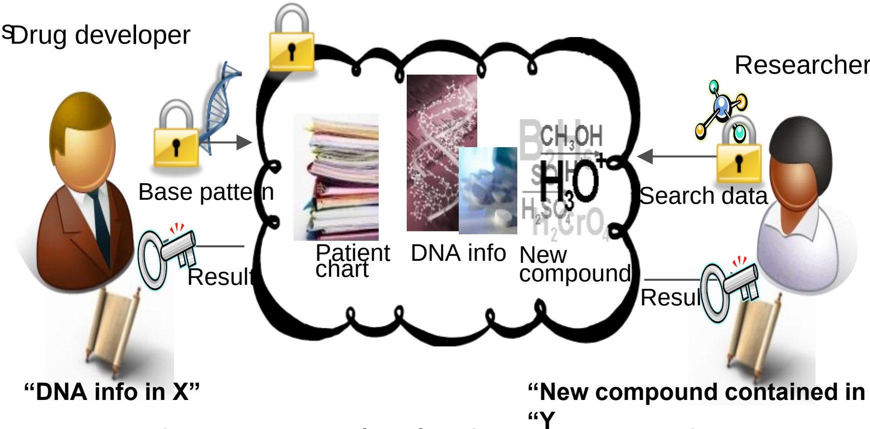


Figure 2: Sample Cloud Usage Scenarios

Why should we do crypto research?

- Cryptography is an enabler
  - There is a realization across the industry that cryptographic technologies are imperative for cloud and big data
  - Mathematical assurance of trust gives people more incentive to migrate data and computation to cloud
- Significant opportunities to explore
  - Mere anonymization is not privacy
    - Systematic and mathematical considerations need to applied when responding to queries on personal data
  - Sophisticated techniques are in research stage or have limited deployments, which enable rich transformations and management of encrypted data