Privacy Engineering: Goals and Implementation Techniques

Tatyana Ryutov

Outline

- Information security goals and implementation techniques
- Privacy protection goals and implementation techniques
- Genomic privacy issues
- Emerging genomic privacy protection technology:
  - Dynamic privacy preserving encryption
  - Watermarking of genomic data

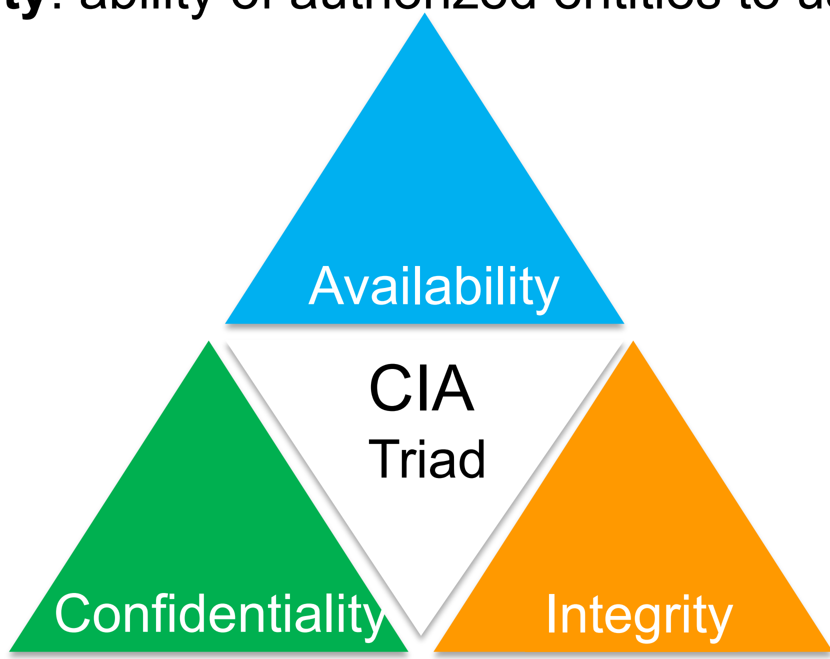
Privacy vs. Data Protection

- The terms of privacy and data protection are not synonyms
- Privacy is about people
  - Sense of being in control
  - A right to be protected
- Data protection is about protecting identifiable data
  - Refers to the organizational perspective
- Information security traditionally addressed the data protection goals

USC Viterbi School of Engineering

Information Security Goals

- CIA Triad:
  - **Confidentiality**: protection from unauthorized disclosure of information
  - **Integrity**: protection from unauthorized modification of information
    - Includes **non-repudiation** that prevents an entity from denying previous commitments or actions and ensures the contents cannot be disputed
  - **Availability**: ability of authorized entities to use the information



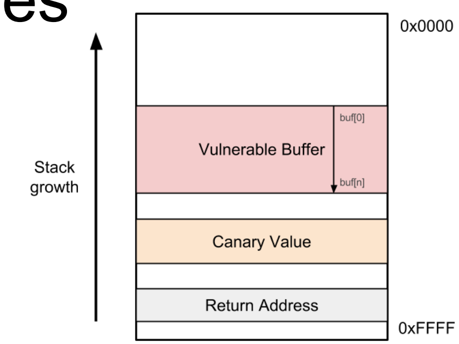
Confidentiality: Examples of Implementation Techniques

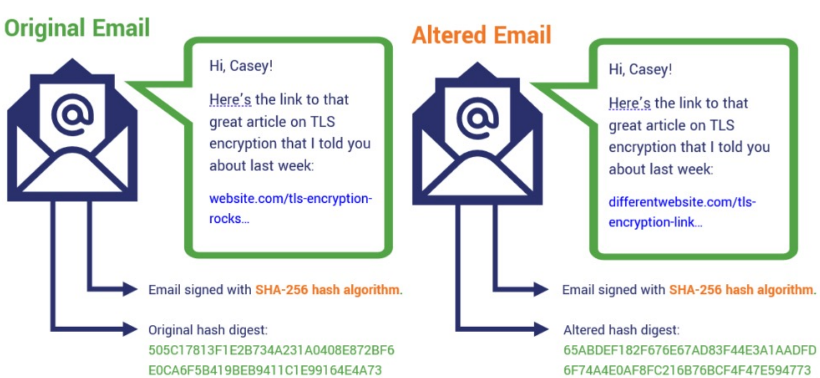
- Data minimization
- Data segregation
  - Secret sharing, secure multiparty computations
- Access control enforcement
  - DAC, MAC, RBAC, ABAC, etc.
- Data encryption
  - in transit (TLS, IPSec, VPN ...)
  - at rest (PGP, BitLocker, TrueCrypt, ...)
  - in processing (homomorphic encryption, confidential computing, …)

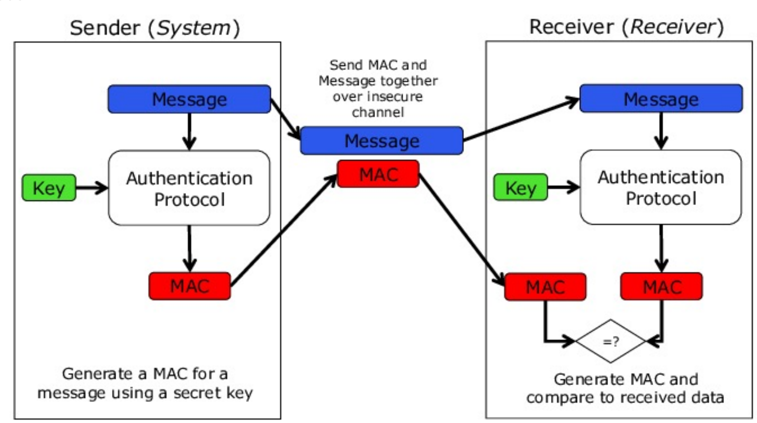
USC Viterbi
School of Engineering

Integrity: Examples of Implementation Techniques

- Digital Signatures
  - RSA, ECDSA
- Message Authentication Codes
- Hash values
- Access control enforcement
- Watchdogs/Canaries
- Two-Man rules

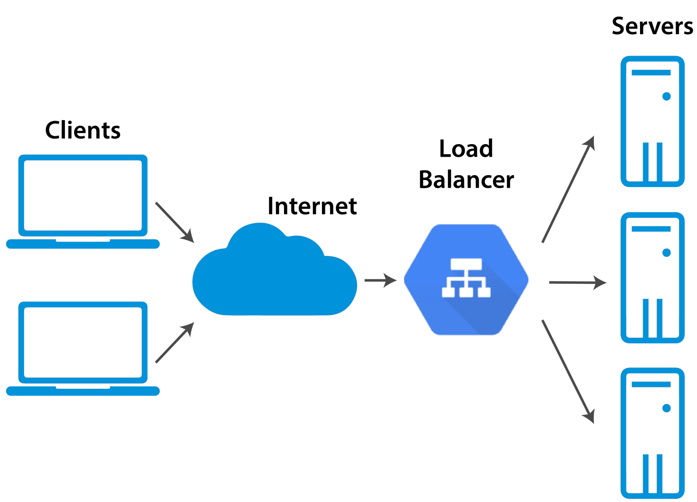


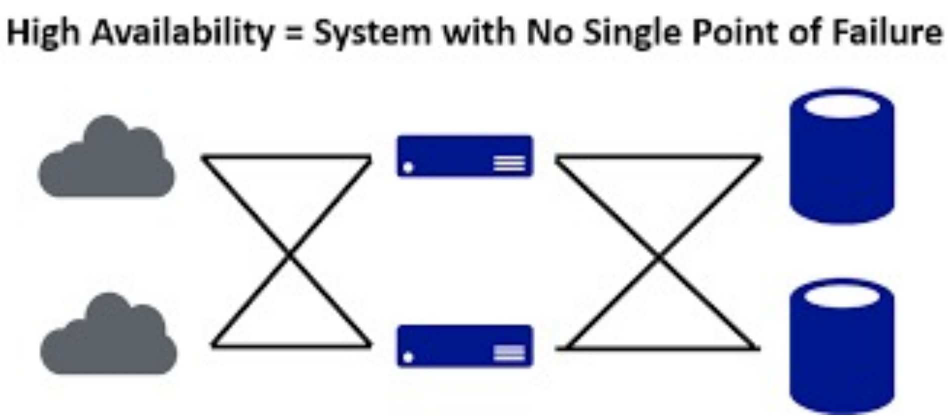




Availability: Examples of Implementation Techniques

- Backups
- Load Balancers
- Failovers
- Redundant Components
- Avoidance of Single-Points-of-Failure





USC Viterbi School of Engineering

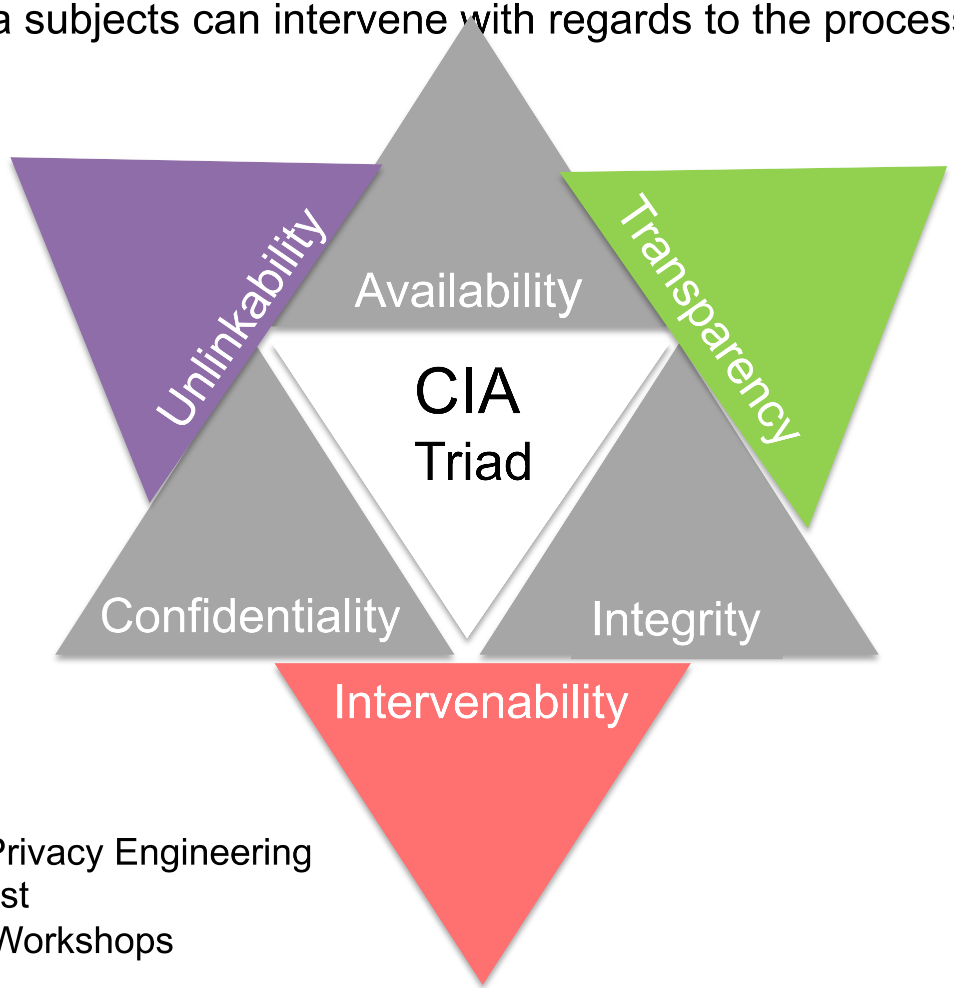
Outline

- Information security goals and implementation techniques
- Privacy protection goals and implementation techniques
- Genomic privacy issues
- Emerging genomic privacy protection technology:
  - Dynamic privacy preserving encryption
  - Watermarking of genomic data

USC Viterbi School of Engineering

Complementing CIA with Privacy Goals

- **Unlinkability/Untraceability:** privacy-relevant data cannot be linked across domains
- **Transparency/Openness**: all privacy- relevant data processing can be understood and reconstructed at any time
- **Intervenability**: the data subjects can intervene with regards to the processing of their data

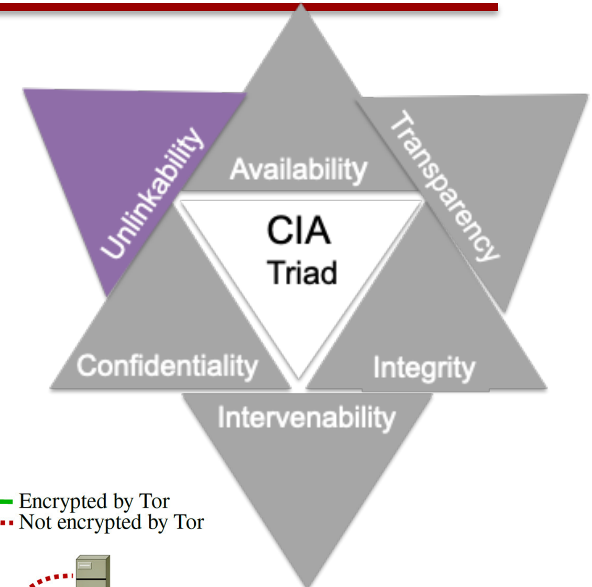


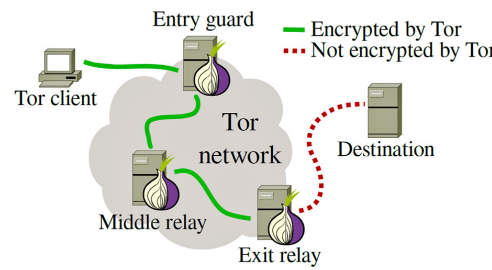
Reference: Protection Goals for Privacy Engineering
M. Hansen, Meiko Jensen, M. Rost
2015 IEEE Security and Privacy Workshops

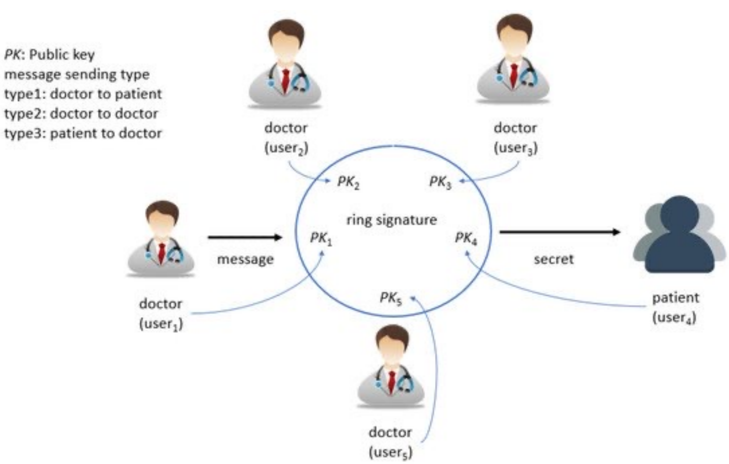
USC Viterbi School of Engineering

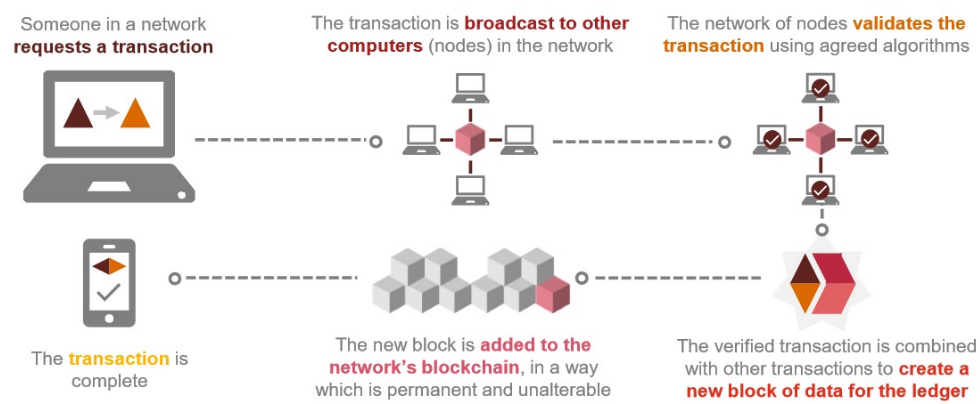
Unlinkability: Implementation Techniques

- Aggregation
  - Anonymization, e.g., K-anonymity, Differential Privacy
- Anonymous communication
  - Crowds, Mix-networks, Onion routing
- Cryptographic techniques
  - Group signature, ring signature
  - ZKP (Zero Knowledge Proof)
  - Digital currency (bitcoin, ….)



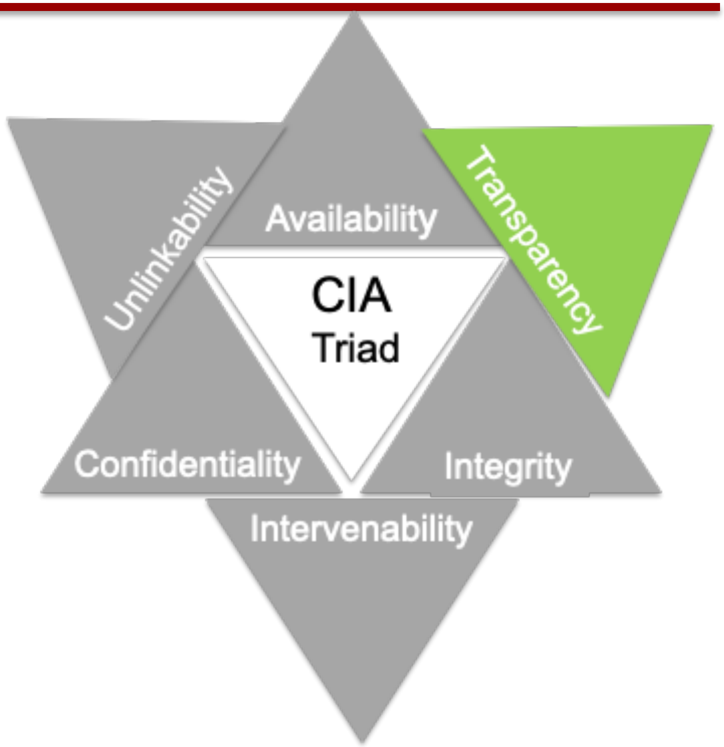






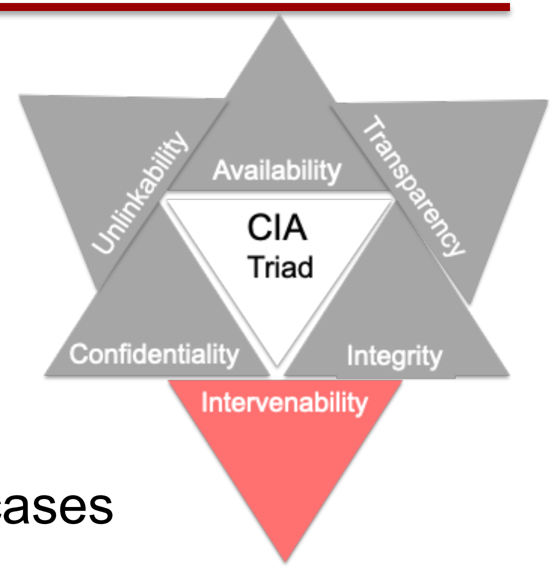
Transparency: Examples of Implementation Techniques

- Logging and reporting
- User notifications
- Transparency services for personal data
- Data breach notifications



USC Viterbi School of Engineering

Intervenability: Examples of Implementation Techniques



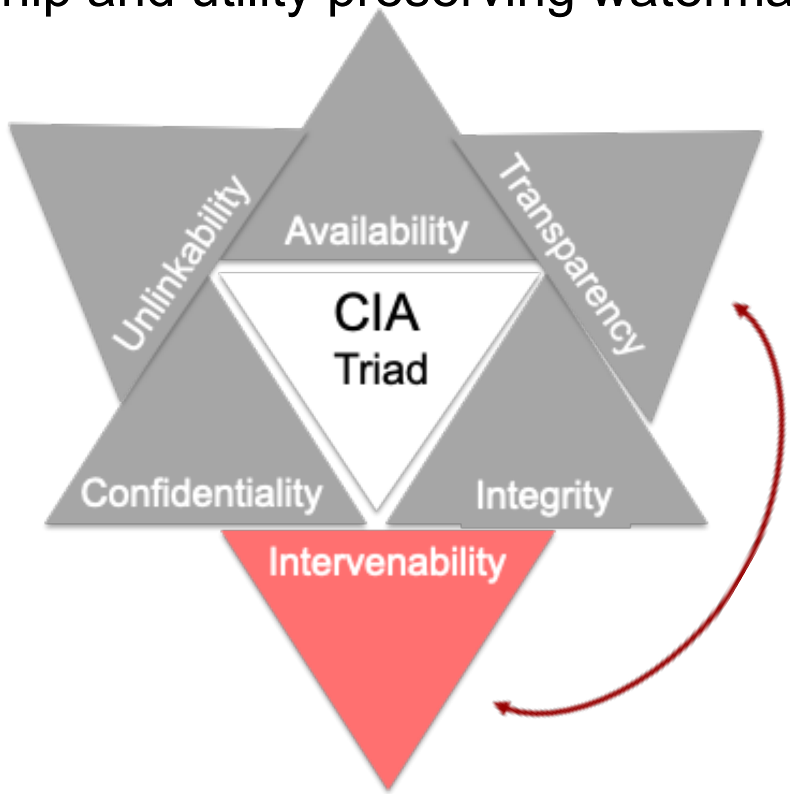
- For individuals (data subjects):
  - Goal: control disclosure of data, but how?
    - Lodge a claim
    - Submit a dispute for arbitration
    - Go to court
- For data controllers:
  - Break-glass procedures known in healthcare: facilitation of a privileged access in emergency cases
  - Alert procedures
  - Manual override of automated decisions
  - Incident management, change management
- No practical existing IT mechanisms!
  - Intervenability techniques go **beyond** IT solutions

Outline

- Information security goals and implementation techniques
- Privacy protection goals and implementation techniques
- Genomic privacy issues
- Emerging genomic privacy protection technology:
  - Dynamic privacy preserving encryption
  - Watermarking of genomic data

Focus: Intervenability

- The challenge: supporting intervenability
- The context: genomic privacy
- Proposed solution:
  1. Dynamic privacy preserving encryption
  2. Ownership and utility preserving watermarking



USC Viterbi
School of Engineering

Genomic Privacy Issues

- Genomic privacy is very important but difficult to protect
  - Misuse of genomic sequencing data may lead to dire consequences, including discrimination against the individual
- Data Protection Regulation (GDPR)
  - A regulation in EU law on data protection and privacy
  - Two essential GDPR rights:
    1. The right to be forgotten
    2. The right to revoke consent

No practical existing solutions!

- Issues:
  - De-identification could not protect privacy of genomic data
  - Informed consent, is the most fundamental aspect of modern biomedical research and basis of the common rules, protects the institution instead of the patient
  - Current practices do not provide individuals with control over their data
- We need to support intervenability - ownership-based governance!
- Robust informatics solutions needed to implement ownership-based governance are currently lacking!

USC Viterbi
School of Engineering

Genomic Privacy

- *In December 2018, LunaDNA received precedent-setting approval from the U.S. Securities and Exchange Commission (SEC) to recognize an individual's health data as currency with which to acquire shares of ownership in the company. This may open up many possibilities when it comes to **monetizing** the data, and a way to incentivize the patients to share the data.*
- We need robust informatics solutions that support:
  - Intervenability
    - Giving and withdrawing consent
  - Individualized privacy needs
    - Example:
      - Dr. James Watson published his fully sequenced genome online, but asked his ApoE4 genotype, which is associated with Alzheimer's disease, to be withheld

USC Viterbi School of Engineering

Outline

- Information security goals and implementation techniques
- Privacy protection goals and implementation techniques
- Genomic privacy issues
- Emerging genomic privacy protection technology:
  - Dynamic privacy preserving encryption
  - Watermarking of genomic data

USC Viterbi School of Engineering

Proposed Solution

- We designed and implemented two algorithms that enable truly ownership-based governance of genomic sequencing data:
  1. Dynamic privacy preserving encryption
     - Dynamic Encryption/Decryption of Genomic Sequencing Data – Ryutov A., Ryutov T., Gai X., U.S. Provisional Patent Application No. 62/859,575
  2. Ownership and utility preserving watermarking
     - Watermarking of Genomic Sequencing data - Ryutov A., Ryutov T., Gai X, U.S. Provisional Patent Application No. 62/891,830
- Support interoperability with existing systems and pipelines for processing genomic data
  - Binary Alignment Map (BAM) format for storing genomic sequences, and Variant Call Format (VCF) format for storing genomic variations
- The algorithms are implemented - LUBA-PrivET
  - LUBA - Lightweight Utilities for Bioinformatics Analysis
  - PrivET - Privacy Enabling Technology
- Using these mechanisms, the data subject (data owner) can specify and **revoke** authorizations for data access and usage

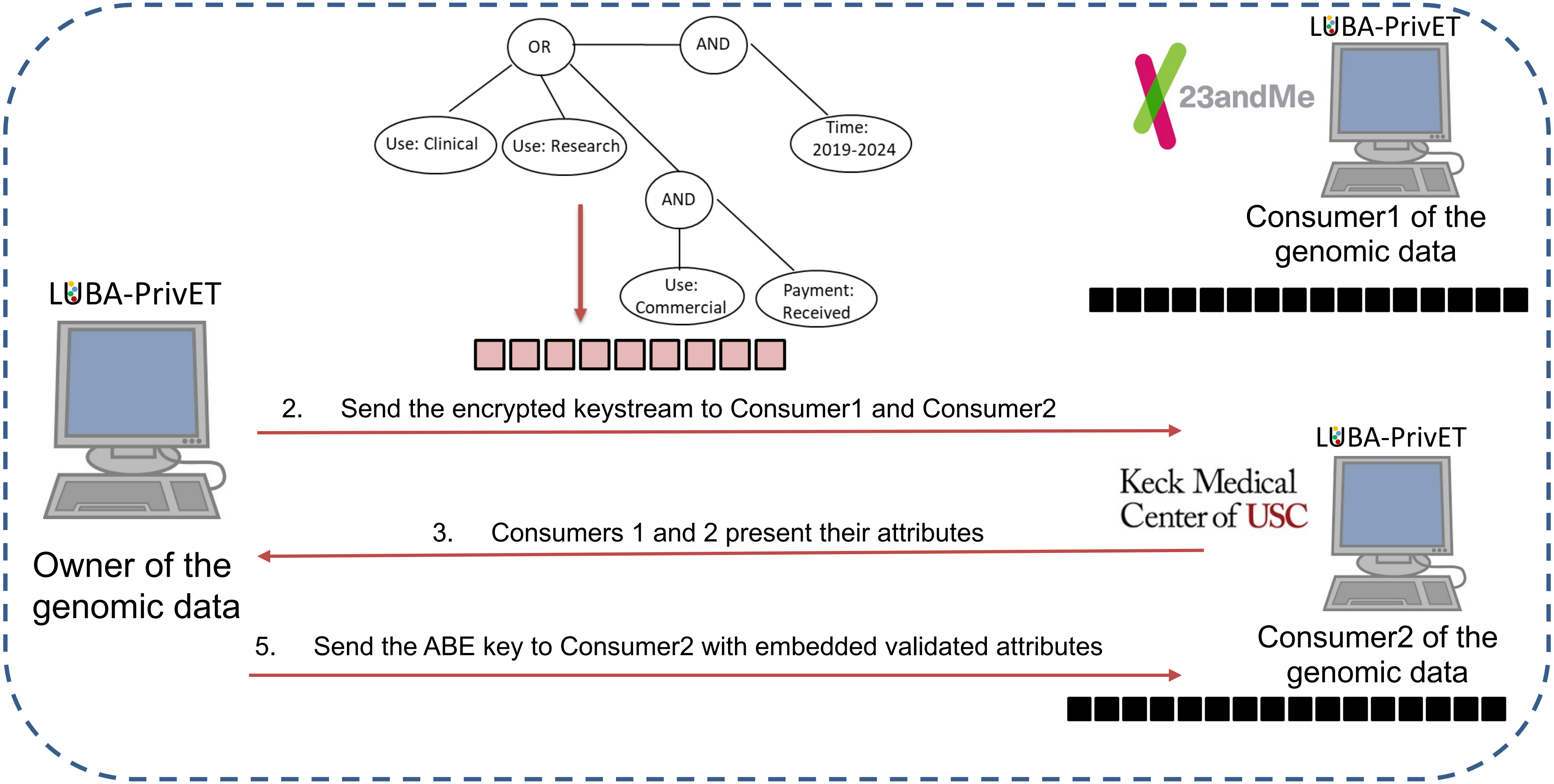
LUBA-PrivET

USC Viterbi School of Engineering

Overview

- Ciphertext Policy (CP) Attributed-Based Encryption (ABE), is an extra layer on top of our existing implementation of the encryption schema
  - The attribute-based policies are embedded in the encrypted data objects (ciphertext), thus making it impossible to remove or modify the policies
  - The data requestor will need to prove the possession of the requestor-related attributes, in order to decrypt the ciphertext
- The dynamic encryption algorithm, combined with CP-ABE allows fine-grained controls over access to the genomic sequencing data
  - Share **specific** genomic regions (without the need to re-encrypt the data), under **certain** conditions (defined as set of attributes)
- The watermarking scheme provides ownership protection, prevents collusion attacks, and enables traceability and audit controls
  - Preserves the **utility** of watermarked data

Example: Dynamic Encryption Usage



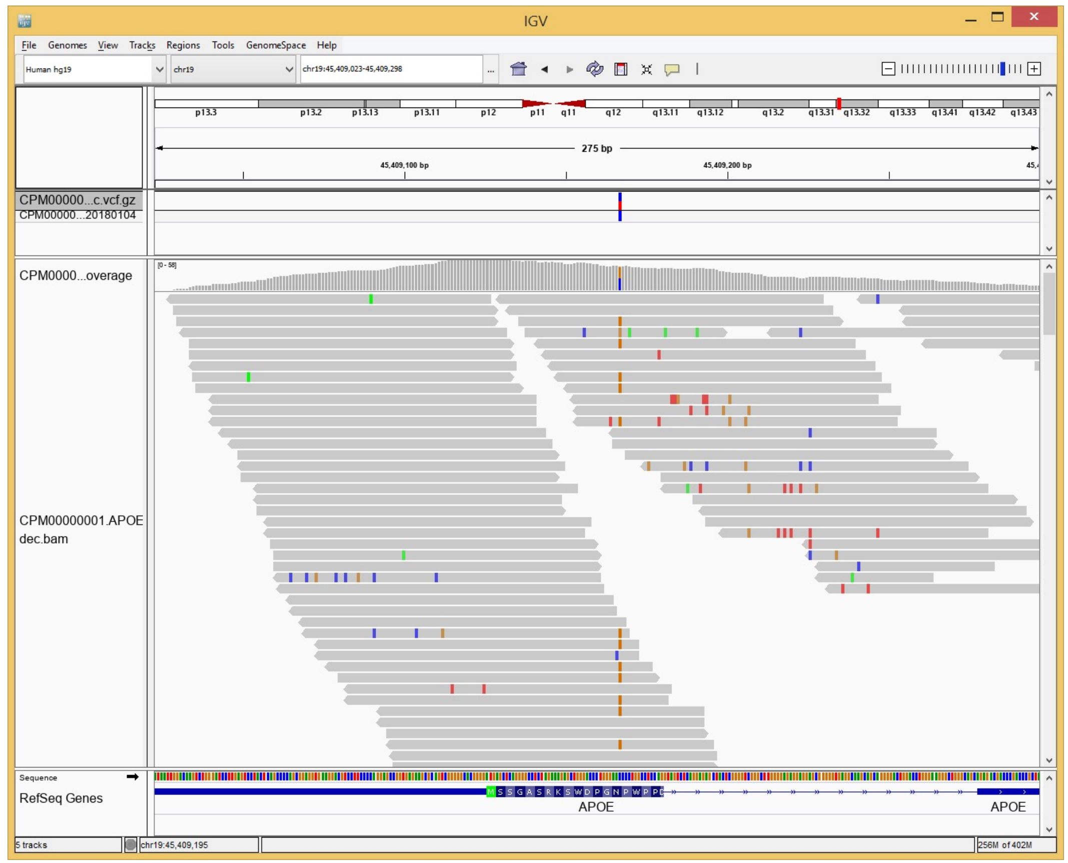
1. Encrypt the keystream that opens a genomic region (e.g., APOE gene) using ABE, the access control policy is imbedded in the ciphertext

4. Validate the attributes, decrypt only if Consumer's attributes satisfy the policy

6. Decrypt the APOE gene

Accessing decrypted region

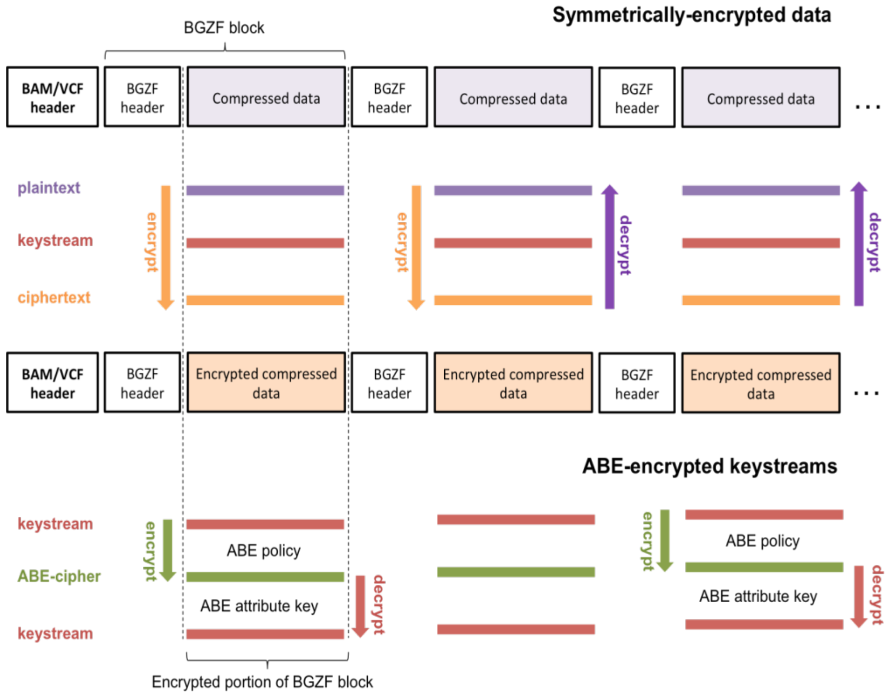
- Decrypted genomic data can be analyzed using the Integrative Genomics Viewer (IGV) is an interactive tool for the visual exploration of genomic data

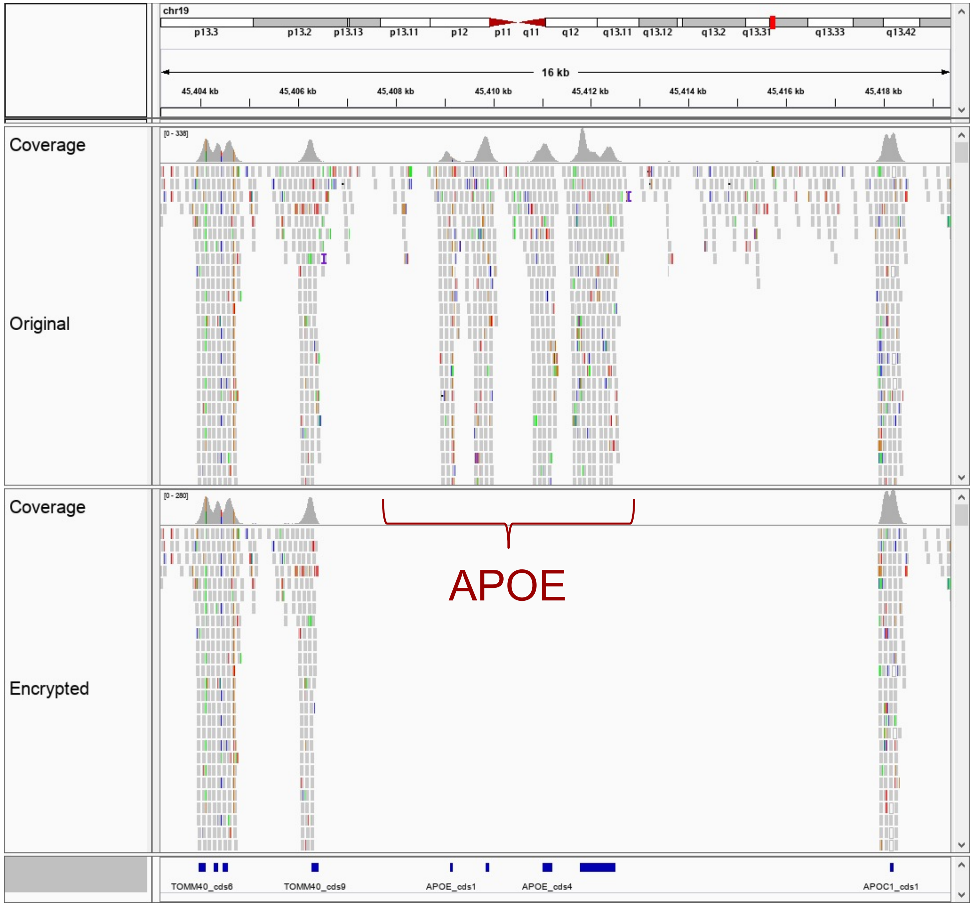


Integrative Genomics Viewer

USC Viterbi School of Engineering

Dynamic encryption and decryption of a BAM file





LUBA-PrivET

USC Viterbi School of Engineering

Outline

- Information security goals and implementation techniques
- Privacy protection goals and implementation techniques
- Genomic privacy issues
- Emerging genomic privacy protection technology:
  - Dynamic privacy preserving encryption
  - Watermarking of genomic data

USC Viterbi School of Engineering

Motivation for Watermarking of Genomic Data

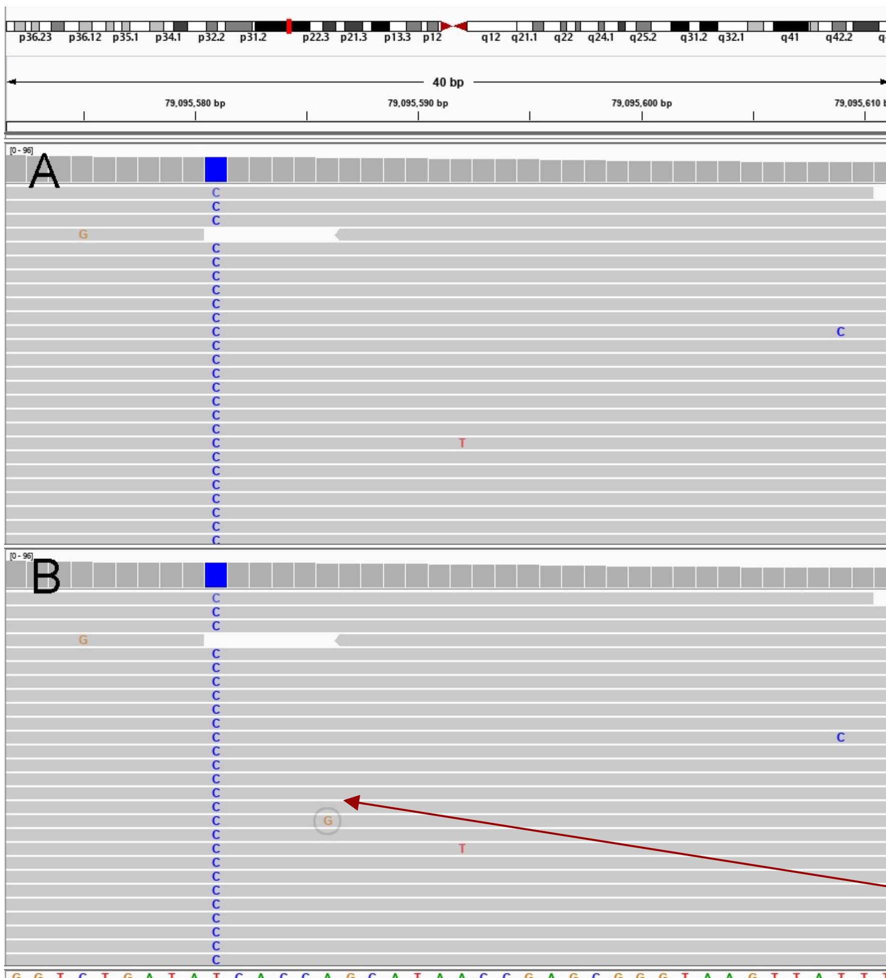
- Practical limitations of the enforcement of ABE policies
  - Some constraints can be enforced only through auditing data transactions and keeping track of the shared genomic data
- Watermarking scheme enables:
  - Support for detection of unauthorized data sharing and use
    - For example, encrypted BAM/VCF file was accessed/shared without the data owner consent (or violating some constrains defined in the consent)
  - Support for consent revocation and enforcement of time based constrains
    - For example, encrypted BAM/VCF file was accessed/shared before release date or after the expiration date
  - Support for accountability
    - Identification of entity who tampered with the watermark and/or violated access control policies.
- No practical existing solutions to watermark genomic data!
  - Either destroy genomic data utility or require proprietary formats that impede interoperability

USC Viterbi
School of Engineering

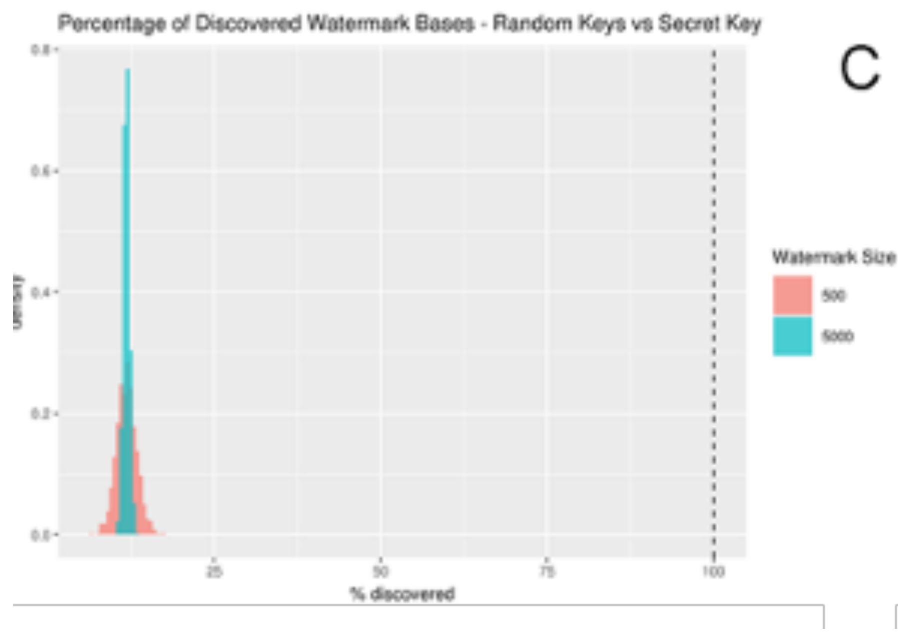
Dynamic Watermarking

- Goal: deter unauthorized use and sharing of data with third parties
- This scheme:
  - Uses **public** watermarking algorithm
    - the only unknown variable is a secret key used to generate the watermark
  - Employs **long** watermark that preserves data **utility** while retaining robustness in protecting whole as well as partial data
  - Provides high robustness:
    - Prevents the identifiability and modification of the watermark by
      - relying only on a secret key, making watermark detection prohibitively expensive
      - hiding the watermark within the inherent noise in the data
    - Provides resistance to collusion attacks
      - colluding parties can detect a portion of the watermark specific only to each party, but not watermark elements **common** to all
  - Detects parties responsible for the unauthorized sharing with a high probability (even when they share a portion of the data or when they modify the data in order to damage the watermark) by selecting watermark positions based on the **identity** of the entity the data was shared with

Dynamic watermarking establishes the ownership



bases matching the reference genome



Impossible to discover/remove all watermarks

A is replaced by G in one of the reads – one watermark element

USC Viterbi School of Engineering

Conclusions

- Information security goals (confidentiality, integrity, availability) should be supplemented with the privacy goals (unlinkability, transparency, intervenability)
- The importance of genomic privacy is increasing
- One of the main challenges is supporting intervenability
- We need technologies that can help address existing technical, legal and ethical hurdles associated with sharing genomic data effectively, securely, and transparently
- Emerging technologies:
  - Dynamic encryption and watermarking of genomic data
    - enable individuals to revoke consent by either disabling data sharing, or by modifying the existing policy
    - supports ownership-based genomic data governance, and may become essential for a genomic data exchange or a market place