Medical and Genomic Data Privacy

Tingting Chen
Assistant Professor
Computer Science Department
Cal Poly Pomona
Outline

• Medical Data Privacy
• Efforts by Government
  – HIPAA, CMIA, EHR and HITECH
• Medical Data Privacy Protection
  – Technical strategies
• Cross Multiple Institutions
  – CONNECT, the DIRECT Project
  – Privacy preserving medical data mining
• Genomic data security and privacy
  – Unique challenges
Medical Data Privacy

• Medical data
  – Any information that relates to health or condition of an individual, the provision of health care to an individual, or the payment for the provision of care to an individual.

• Privacy
  – The claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others.

• Medical data privacy
  – Individual
  – Institutional
HIPAA and CMIA

• HIPAA
  – It establishes a federal floor of safeguards to protect the confidentiality of medical information.
  – 18 HIPAA identifiers.

• CMIA in California
  – Confidentiality of Medical Information Act
  – Requires patient authorization for release of information unless release otherwise permitted or required by law
Government Investment

• U.S. government’s investment in Health Information Technology
  – Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009
  – $29 billion Electronic Health Records (EHR) reimbursement program
  – Contracts of $18.6 million to develop prototypes for the National Health Information Network (NHIN) architecture.

• NHIN
  • Support accurate, timely, appropriate, and secure health-care information exchange.

Nationwide Health Information Network (NHIN)
Medical Data Privacy Protection

• Data collection
  – Encryption: most vendors use AES. Some accelerations, such as Intel AES-NI.

• Data storage
  – Authentication
  – Access control
  – API security

• Data utility
  – Publishing the de-identified data.
Medical Data Privacy Protection

- Anonymization before publishing
  - Masking methods: Suppression, generalization, etc.

Example: Value Suppression
Medical Data Privacy Protection

- Anonymization before publishing
  - Masking methods: Suppression, generalization, etc.

Example: generalization
Data Across Multiple Institutions

• When personal medical information moves across hospitals, doctors’ offices, insurers or third party payers, and state lines.

• The Direct Project (launched 2010)
  – Developing technical standards and services required to enable secure, directed health information exchange among trusted providers.
  – XDR and XDM for direct messaging.
  – SMTP, S/MIME, and X.509 certificates are used.

• CONNECT
  – Free, open source software solution that supports health information exchange

• Nationwide Health Information Network Exchange
  – First community that implements the NHIN standards.
  – Including CDC, DoD, Kaiser Permanente, MedVirginia, etc.
Data Across Multiple Institutions

Top threats include “Sharing data with third parties”.

Privacy preserving medical data mining across multiple organizations

– Collaborative medical research leveraging big data.
– Both individual and institutional privacy should be protected.
– Solutions include applications of homomorphic cryptographic algorithms, data transformation, etc.

Survey by KMPG of 223 healthcare executives in 2015
Genomic Data Privacy

• Applications of human genome data makes more personalized treatments and preventive healthcare possible.

• Personal genomic data contains highly sensitive information, e.g., prediction of a certain disease.

• Unique challenges in protecting genomic data privacy.
  – Genomic data has a long life span as DNA almost does not change over time. Brute force attackers have more time.
  – Genome itself is the ultimate identifier. Traditional anonymization methods that remove personally identifiable information are not effective.
  – Harder to resolve privacy concerns of genomic data owner to participate in medical research.