



IBM Research

# Security and Privacy Technology Enablers for Electronic Healthcare

Tyrone Grandison PhD  
**IBM Healthcare Center of Excellence**  
Almaden Research Center  
San Jose, California  
tyroneg@us.ibm.com

2007 AMIA Spring Congress | May 22-24, 2007

# Introduction

## ■ Caveat

- As medical information move to electronic platforms, policy and social education programs **must** be augmented by appropriate, corresponding technology<sup>1</sup>.

## ■ Objectives

- Define the addressable.
- Define the current major problems.
- Outline technological solutions to each of these problems.

<sup>1</sup>Christopher Johnson, Rakesh Agrawal, "Intersections of Law and Technology in Balancing Privacy Rights with Free Information Flow", Proceedings of the Fourth IASTED International Conference on Law and Technology, Cambridge, Massachusetts, USA, October 2006.

# Scope of Current Technical Enablers

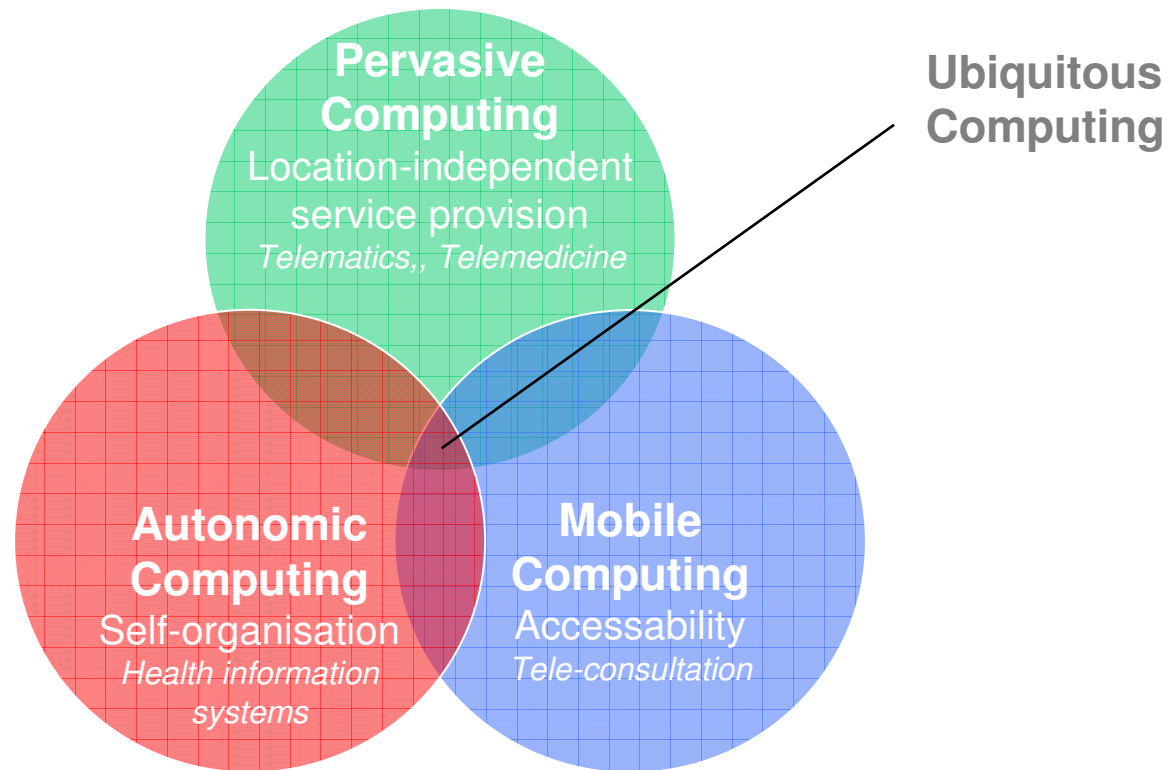
## ■ The Problem Space

- Tightly Coupled Complex Systems
- Each Silo'ed System has its own Protection Mechanisms
- Conflicting Priorities and Policies
- New (and changing) Technology

## ■ Solution Requirements

- Reduce the complexity and work-load in integrating and deploying systems, i.e. allow systems to worry about their core function and leverage security and privacy controls in the data system.
- Do not impact the performance/efficiency of the currently running system
- Enable the current (clinical) workflow and do not require it to change.

# What the Future Holds for Healthcare?



- Bernd Blobel, Head, German National eHealth Competence Center, University of Regensburg Medical Center, Regensburg, Germany

## System Requirements – Current & Future

- Appropriate security and privacy services
- Openness
- Flexibility
- Scalability
- Portability
- User acceptance
- Service orientation
- Distribution at Internet level
- Lawfulness
- Based on standards
- Service-oriented interoperability

– Bernd Blobel, Head, German National eHealth Competence Center, University of Regensburg Medical Center, Regensburg, Germany

## Current Major Problems

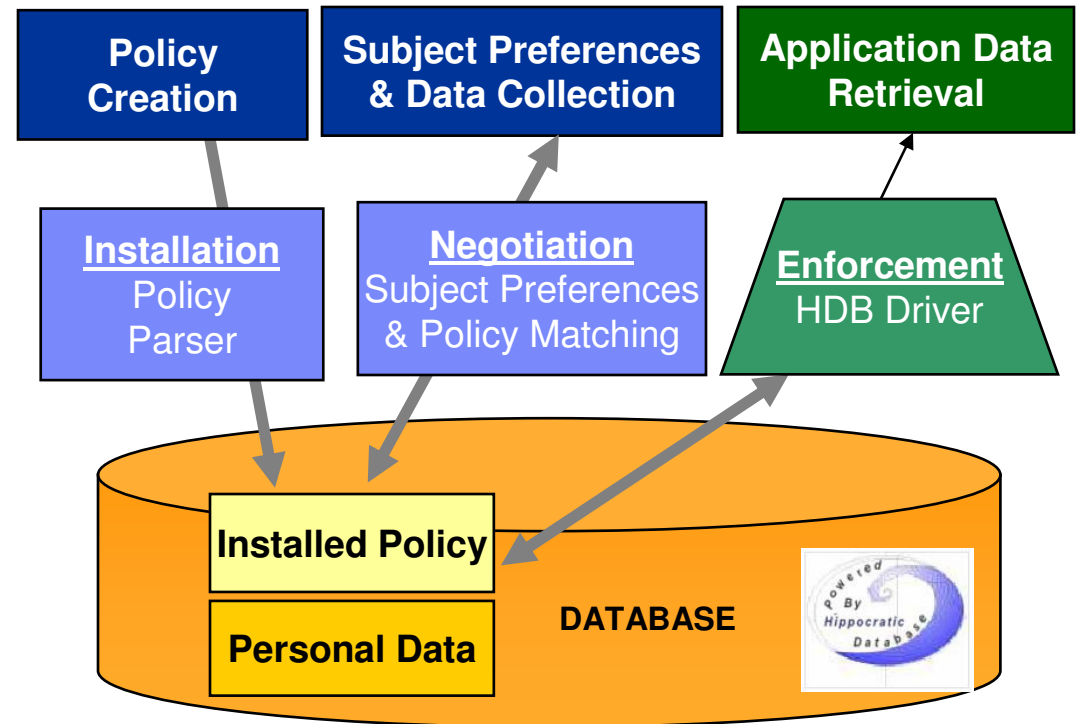
- Policy-based Private Data Management.
  - How does one enforce data disclosure policies and patient preferences?
  - How does one enable privacy-preserving data mining?
- Secure Information Exchange
  - How does one selective share the minimum amount of data necessary for a task?
  - How does one de-identify data for information exchange?
- Efficient Data Access Tracking
  - How do you efficiently track access and disclosure?
  - How do you protect data sent to outsourced agents?

## Technology Solutions

- Policy-based Private Data Management.
  - Active Enforcement
  - Privacy-Preserving Data Mining
- Secure Information Exchange
  - Sovereign Information Sharing
  - Optimal  $k$ -anonymization (de-identification)
- Efficient Data Access Tracking
  - Compliance Auditing
  - Database Watermarking

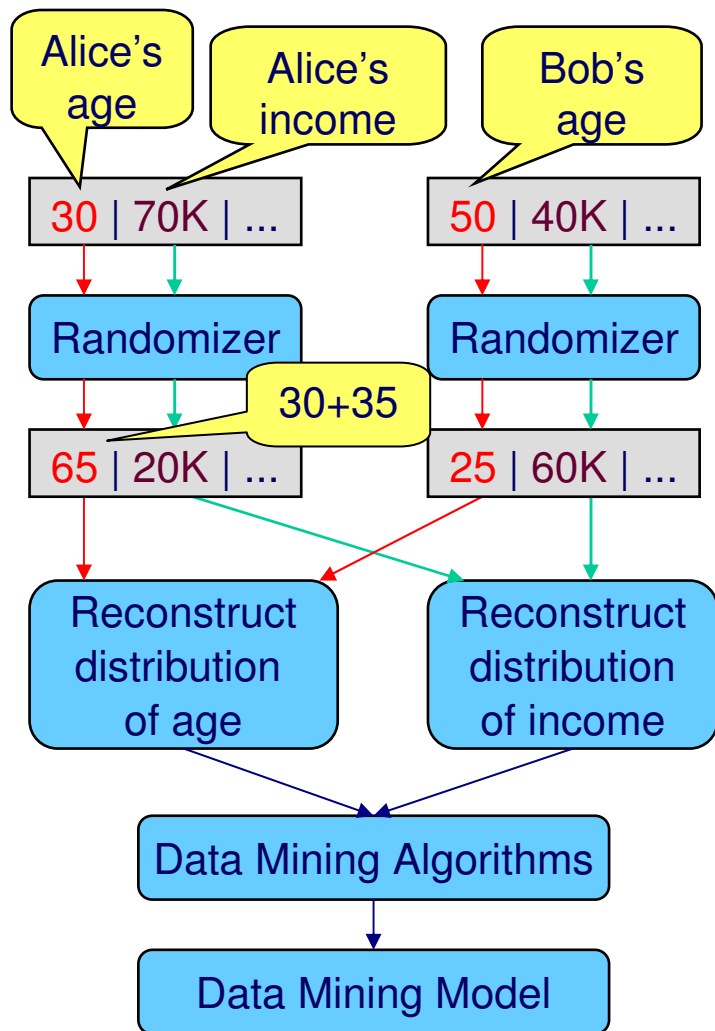
# Hippocratic Database Active Enforcement

- **Privacy Policy:** Organizations define a set of policies describing who may access data (users or roles), for what purposes data may be accessed (purposes) and to whom data may be disclosed (recipients).
- **Consent:** Data subjects are given control, through opt-in and opt-out choices, over who may see their data and under what circumstances
- **Active Enforcement:** Intercepts and rewrites incoming queries to comply with policies, subject choices, and context.
- **Efficiency:** Rewritten queries benefit from all of the optimizations and performance enhancements provided by the underlying engine (e.g. parallelism).
- **Advantages:**
  - Cell-level access and disclosure control.
  - Application modification not required.
  - Database agnostic; does not require changes to the database engine.

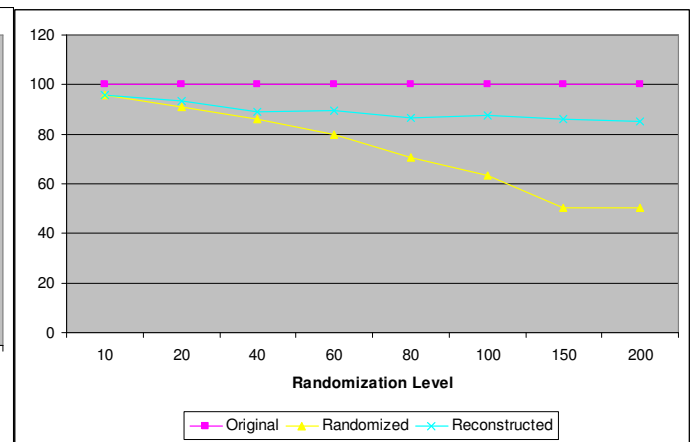
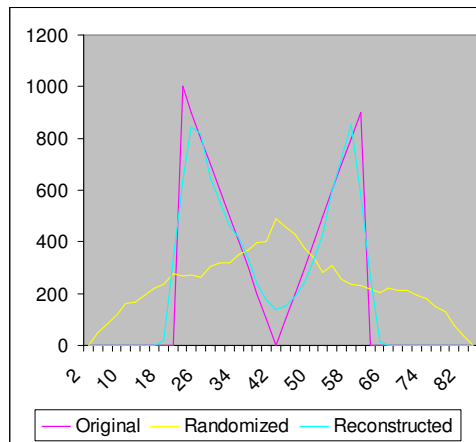


#	Name	Age	Phone
1	Adam	25	(111) 111-1111
3	Bob	-	(333) 333-3333
4	Daniel	40	-

# Privacy-Preserving Data Mining

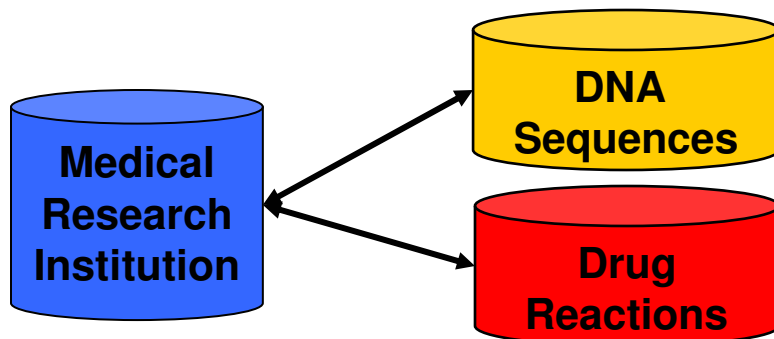


- Preserves privacy at the individual level, but allows accurate data mining models to be constructed at the aggregate level.
- Adds random noise to individual values to protect data subject privacy.
- EM algorithm estimates original distribution of values given randomized values + randomization function.
- Algorithms for building classification models and discovering association rules on top of privacy-preserved data with only small loss of accuracy.



# Sovereign Information Integration

- Autonomous databases for competitive, statutory, or security reasons.
  - Provides selective, minimal sharing on need-to-know basis.
- Example: Which DNA expressions correlate with reactions to certain drugs?
- Algorithms for computing secure joins and join counts without revealing any additional information among the databases.



## Minimal Necessary Sharing

R	
a	
u	
v	
x	

S	
b	
u	
v	
y	

$R \bowtie S$

- R must not know that S has b & y
- S must not know that R has a & x

$R \bowtie S$	
u	
v	

**Count ( $R \bowtie S$ )**

- R & S do not learn anything except that the result is 2.

# Optimal $k$ -Anonymization

- **Optimal  $k$ -Anonymization** (Bayardo, Agrawal, 2005)
  - Algorithm finds optimal  $k$ -anonymizations under two representative cost measures and variations of  $k$ .
- **Advantages of optimal  $k$ -anonymization:**
  - **Truthful** - Unlike other disclosure protection techniques that use data scrambling, swapping, or adding noise, all information within a  $k$ -anonymized dataset is truthful.
  - **Secure** - More secure than other de-identification methods, which may inadvertently reveal confidential information.

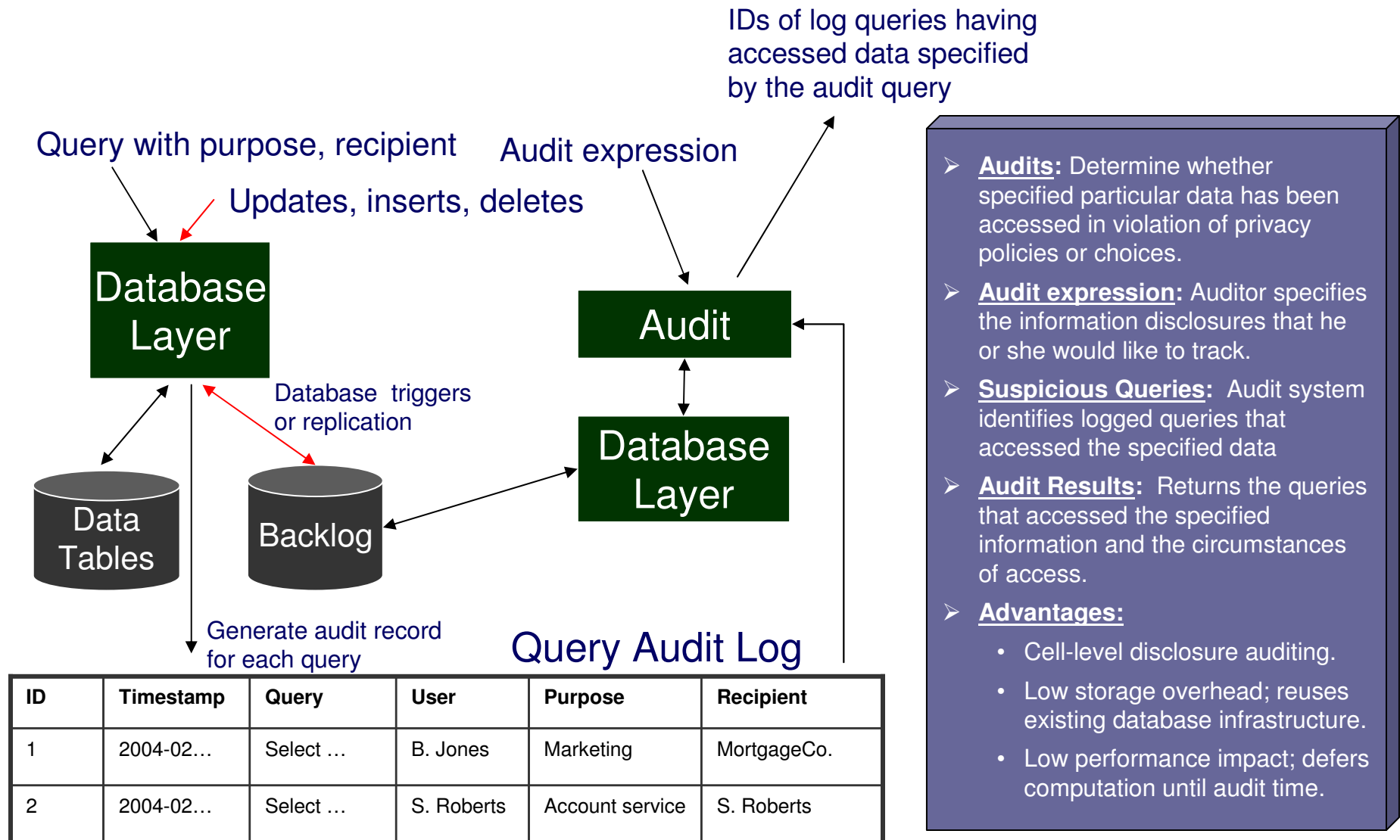
Name	Address	City	Age	Income
Erica	19 Main Street	San Jose	26	\$42,000
Paul	130 Harry Road	San Jose	42	\$88,000
Mark	4800 17th Street	San Jose	47	\$120,000
Henry	210 Almaden Pkwy	San Jose	28	\$50,000

( $k=2$ , on name,  
address, age)



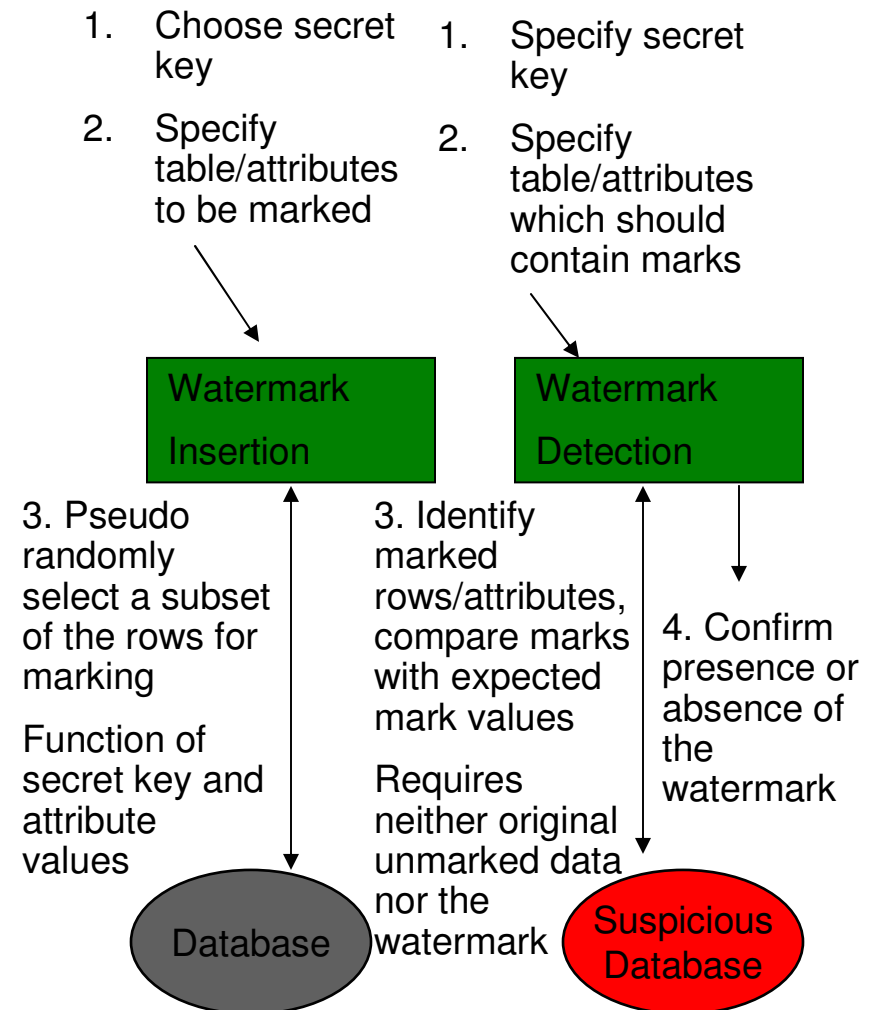
Name	Address	City	Age	Income
*	95131	San Jose	20-29	\$42,000
*	95120	San Jose	40-49	\$88,000
*	95120	San Jose	40-49	\$120,000
*	95131	San Jose	20-29	\$50,000

# Compliance Auditing



# Watermarking Databases

- Deters data theft and asserts ownership of pirated copies by intentionally introduced pattern in the data.
  - Very unlikely to occur by chance.
  - Hard to find => hard to destroy (robust against malicious attacks).
- Existing watermarking techniques developed for multimedia are not applicable to database tables.
  - Rows in a table are unordered.
  - Rows can be inserted, updated, deleted.
  - Attributes can be added, dropped.
- New algorithm for watermarking database tables.
  - Watermark can be detected using only a subset of the rows and attributes of a table.
  - Robust against updates, incrementally updatable.



## Conclusion

- Technology controls for security and privacy must be used in conjunction with legal policy, organizational requirements and social awareness programs in order to address the current and future problems in medical informatics systems.
- Controls must be moved to the data level in order to:
  - Reduce the complexity in current system.
  - Provide a unified protection framework.
  - Allow the resolution of conflicts at the data level.
  - Scale to future technology without infrastructure modification.
- There is a current set of enablers that would avert breaches and integrate seamlessly into current systems.

# Thank You



## Selected References

- Rakesh Agrawal, Tyrone Grandison, Christopher Johnson, Jerry Kiernan, "Enabling the 21st Century Healthcare Information Technology Revolution," Communications of the ACM, Vol. 50, No. 2, February 2007.
- Tyrone Grandison, Ranjit Ganta, Uri Braun, Jamie Kaufman, "Protecting Privacy while Sharing Medical Data Between Regional Healthcare Entities". To appear in Medinfo 2007 Congress. August 2007. Brisbane, Australia.
- Rakesh Agrawal, Christopher Johnson, "Securing Electronic Health Records without Impeding the Flow of Information," International Journal of Medical Informatics, January 2007, doi:10.1016/j.ijmedinf.2006.09.015.

<http://www.almaden.ibm.com/cs/projects/iis/hdb/publications.shtml>

Slides available at

[http://www.almaden.ibm.com/cs/people/tgrandison/AMIA\\_Spring2007.pdf](http://www.almaden.ibm.com/cs/people/tgrandison/AMIA_Spring2007.pdf)