The Role of Audit Analysis in CyberSecurity

Dr. Tyrone Grandison MBA FHIMSS

Proficiency Labs

Quick Intro

- Over twenty years in computer science.
  - Industry, Academia, Industry Research, Consulting, Startups.
- Professional Activity
  - Over a hundred publications.
  - Over forty-five patents.
  - Three books on either Privacy, Security or Trust.
- Memberships
  - Fellow – British Computer Society; Fellow – Royal Society for the Advancement of Sciences; Fellow – Healthcare Information Management Systems Society; Distinguished Engineer – IEEE; Senior Member, ACM.

More at http://www.tyronegrandison.org

Outline Of This Talk

- Definitions and What-Not
- The Importance of CyberSecurity
- The Current State of Affairs
- The Opportunities
- My Research & Commercialization Focus
- Case Studies
  - Compliance Auditing
  - Exception-Based Access
- Future Work
- Conclusion

Definitions and What-Not

- Perspectives on CyberSecurity
  - Scope of CyberSecurity
- My Definition of CyberSecurity

Perspectives on CyberSecurity

- Very wide-ranging term
- Everyone has a different perspective
- No standard definition
- A socio-technical systems problem

Scope of CyberSecurity

- Threat and Attack analysis and mitigation techniques
- Protection and recovery technologies, processes and procedures for individuals, business and government
- Policies, laws and regulation relevant to the use of computers and the Internet

Cybersecurity

The field that synthesizes multiple disciplines, both technical and non-technical, to create, maintain, and improve a safe environment.

- The environment normally allows for other more technical or tactical security activities to happen, particularly at an industry or national scale.
- Traditionally done in the context of government laws, policies, mandates, and regulations.

The Importance of CyberSecurity

- What Factors Make CyberSecurity Important?
  - Why is it so difficult?

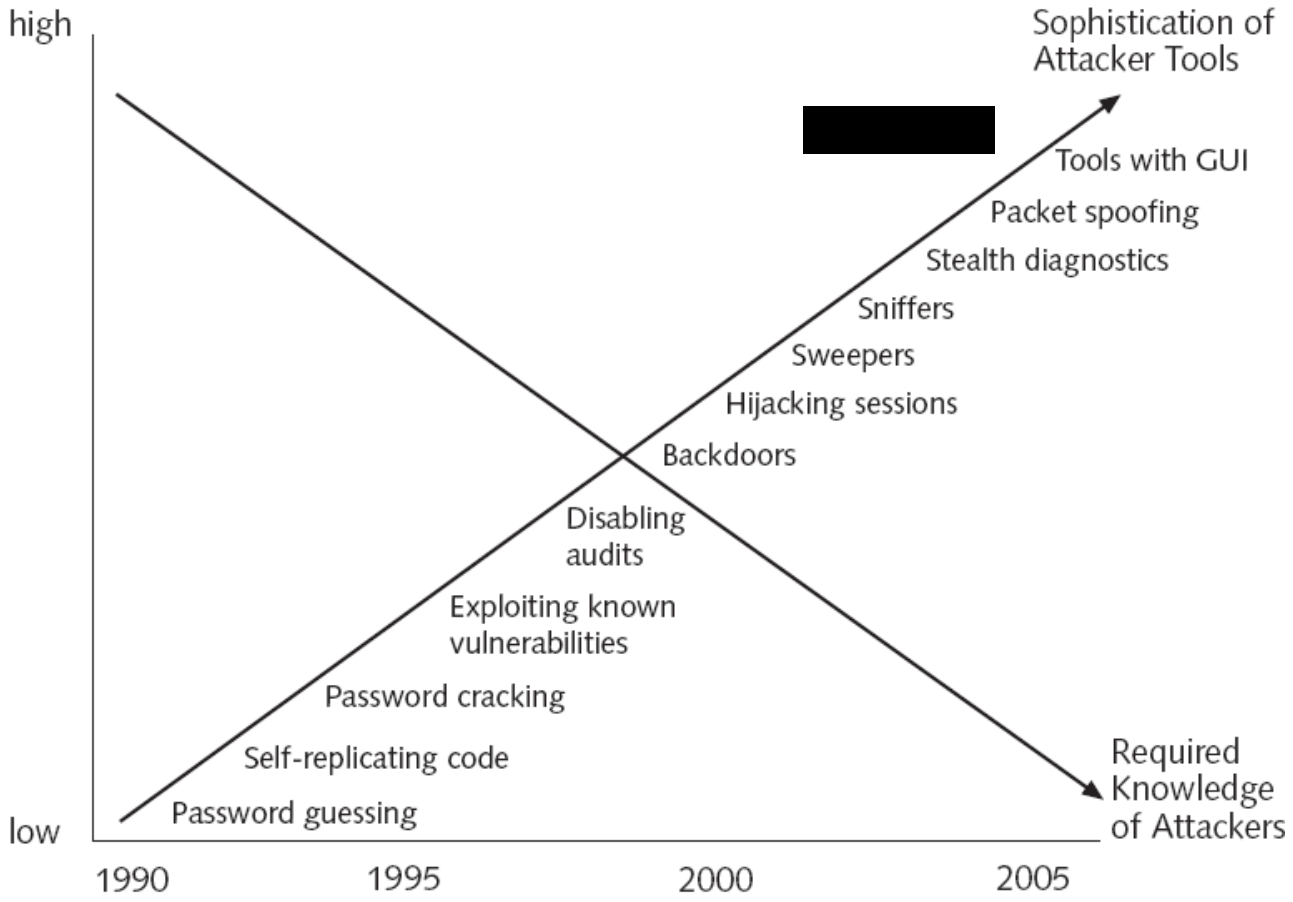
Why Is It Important?

- Heavy Reliance on the Internet
  - Commerce
  - Internet of Things
- Impact of Attack
  - Risk, Harm, Reputation, Brand
- Incentive to Attack
- Increased Difficulty in Defense

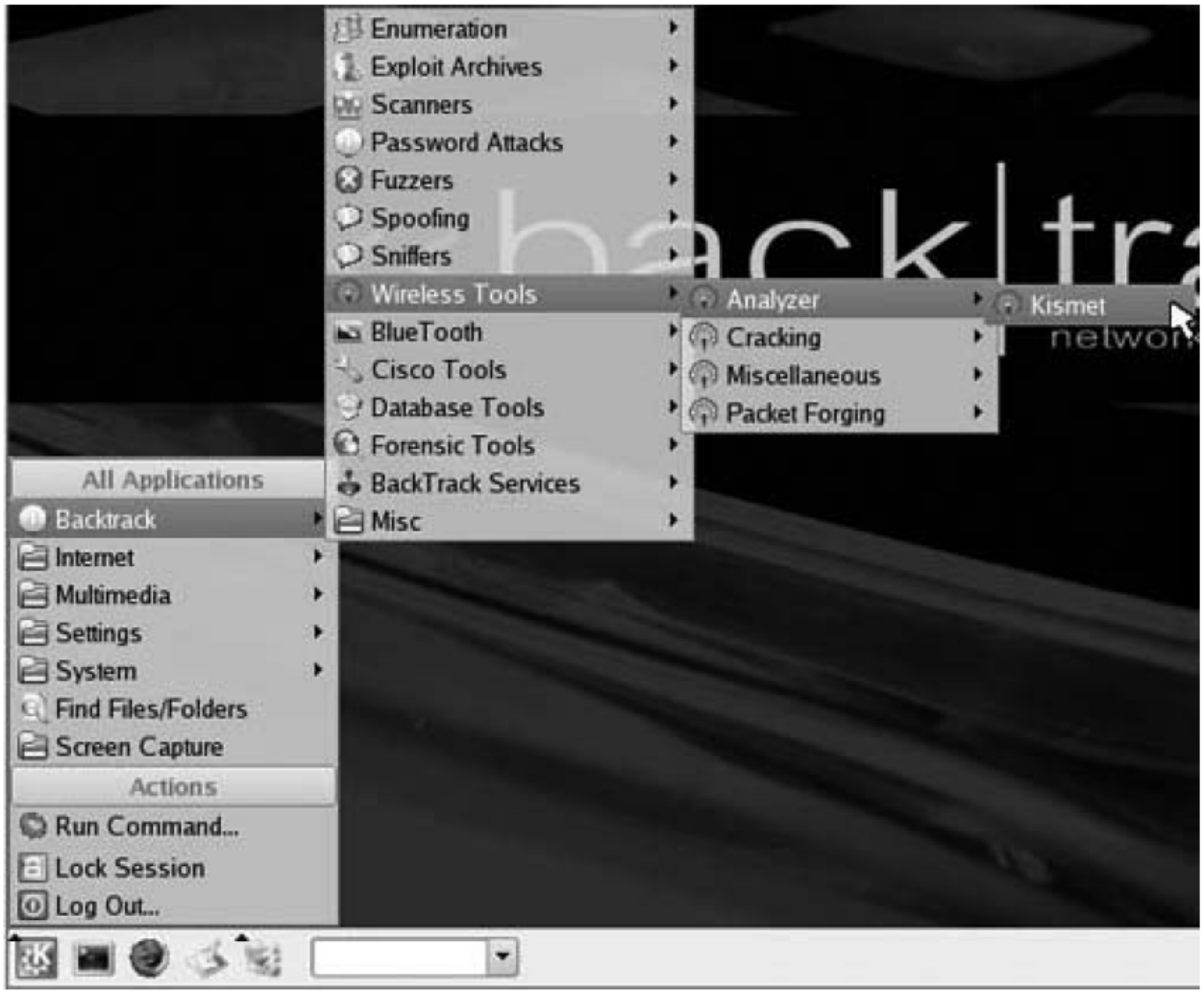
Difficulties in Defending against Attacks

Reason	Description
Speed of attacks	Attackers can launch attacks against millions of computers within minutes.
Greater sophistication of attacks	Attack tools vary their behavior so the same attack appears differently each time.
Simplicity of attack tools	Attacks no longer limited to highly skilled attackers.
Detect vulnerabilities more quickly	Attackers can discover security holes in hardware or software more quickly.
Delay in patching	Vendors are overwhelmed trying to keep pace by updating their products against attacks.
Distributed attacks	Attackers can use thousands of computers in an attack against a single computer or network.
User confusion	Users are required to make difficult security decisions with little or no instruction.

Increased Sophistication of Attack Tools



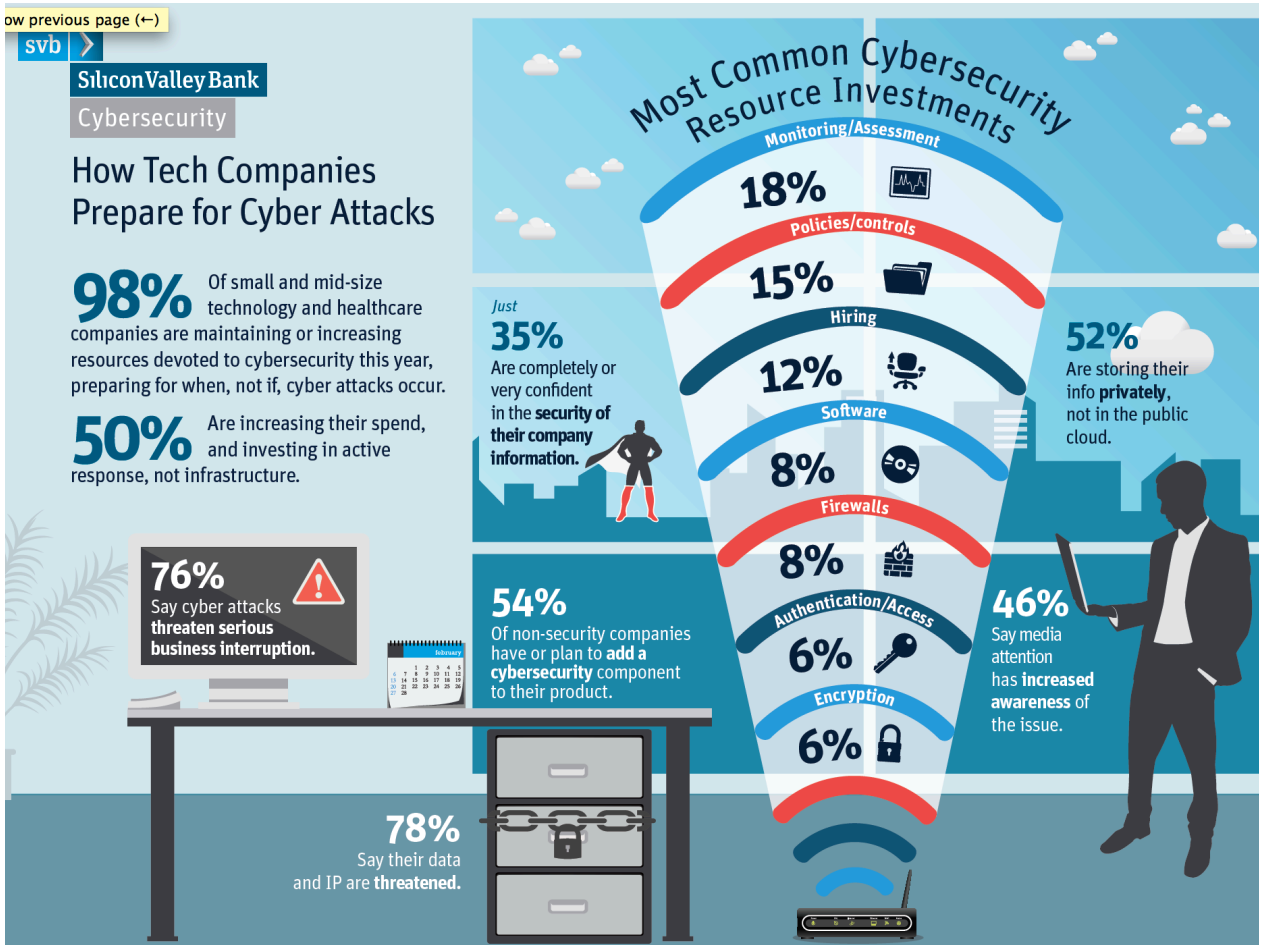
Menu of Attack Tools



The Current State of Affairs

- Corporate US Landscape
- Global Situation
- Current Insight

Corporate US Landscape



Statistics from the results of an SVB survey about cybersecurity completed by 216 C-level executives from US-based technology and life science companies in July 2013

Global Situation

- 47% of companies know they have suffered a cyber attack in the past year
- 70% say they are most vulnerable through their endpoint devices
- 52% rate at "average-to-non-existent" their ability to detect suspicious activity on these devices

2013 Cyber Security Study - What is the Impact of Today's Advanced Cyber Attacks?
- Bit9 and iSMG

Current Insight

- First-Generation Security Solutions Cannot Protect Against Today's Sophisticated Attackers
- There is No Silver Bullet in Security
- There is an Endpoint and Server Blindspot

2013 Cyber Security Study - What is the Impact of Today's Advanced Cyber Attacks?
- Bit9 and iSMG

Where Are The Opportunities?

- What are the Hard Research Problems?
- Where are companies spending their CyberSecurity dollars?

Hard Problems
(Nine Years Ago)

1. Global-Scale Identity Management
2. Insider Threat
3. Availability of Time-Critical Systems
4. Building Scalable Secure Systems
5. Situational Understanding and Attack Attribution
6. Information Provenance
7. Security with Privacy
8. Enterprise-Level Security Metrics

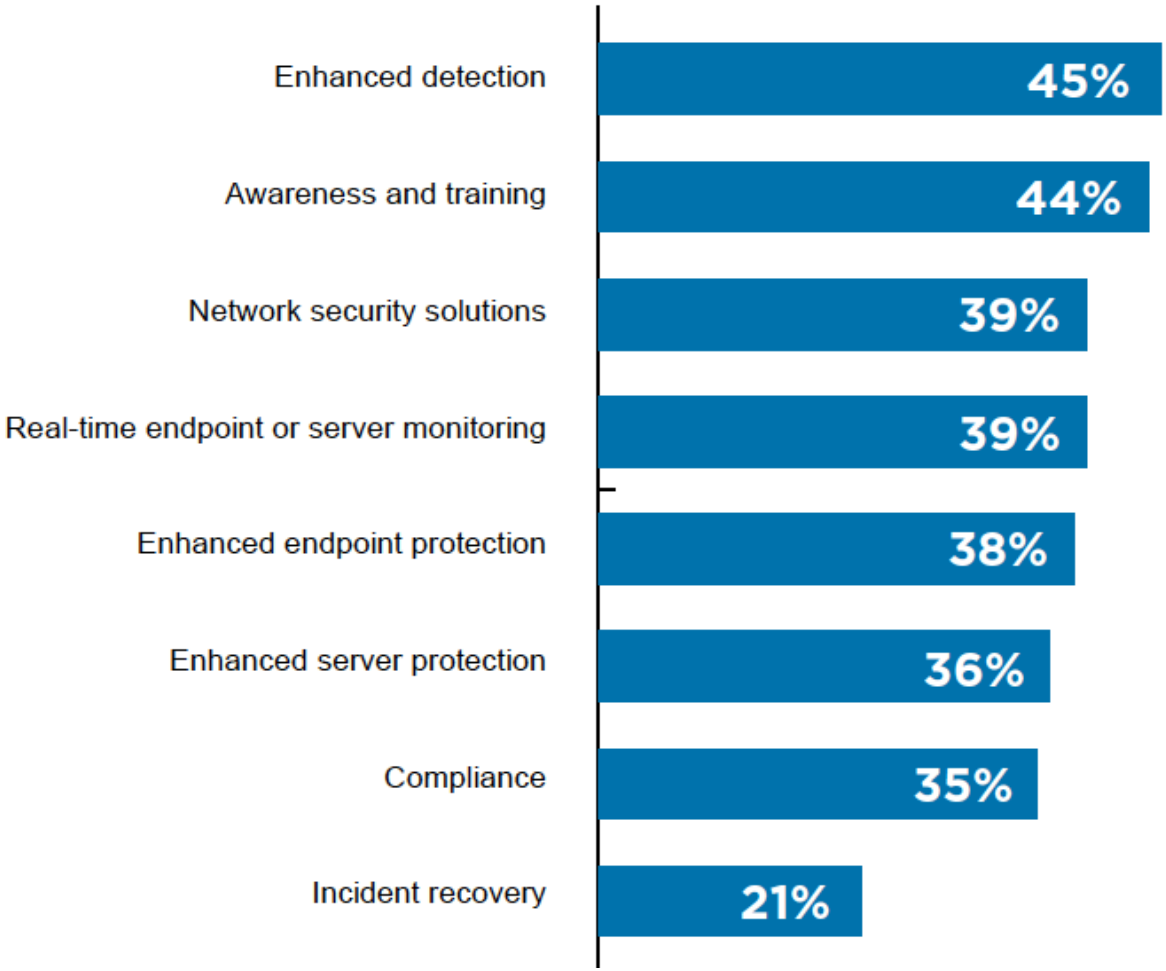
INFOSEC Research Council (2005)

Hard Problems
(Five Years Ago)

1. Global-scale Identity Management
2. Combatting Insider Threats
3. Survivability of Time-critical Systems
4. Scalable Trustworthy Systems
5. Situational Understanding and Attack Attribution
6. Provenance
7. Privacy-aware security
8. Enterprise-level metrics
9. System Evaluation Life Cycle
10. Combatting Malware and Botnets
11. Usable Security

INFOSEC Research Council (2009)

2014 Projected Spending



2013 Cyber Security Study - What is the Impact of Today's Advanced Cyber Attacks? - Bit9 and iSMG

My Research and Commercialization Focus

- Data, Data, Data
- Detection, Detection, Detection

My Focus

- The most valuable asset of the 21st century company - **Data**
- CyberSecurity Realities
  - Proactive, Real-Time Detection impossible
  - Mostly a Losing Game for non-attackers
- My Focus (aka next realistic move):

Proactive, near Real-Time Attack Detection using Audit Logs

Fundamental Challenges

- Audit systems are normally not switched on
  - When on, slows down the production system and degrades the delivery of service.
- Audit systems contain a lot of information
  - Not all of it is useful
- Access to real data
  - Shrouded in mystery due to ramifications

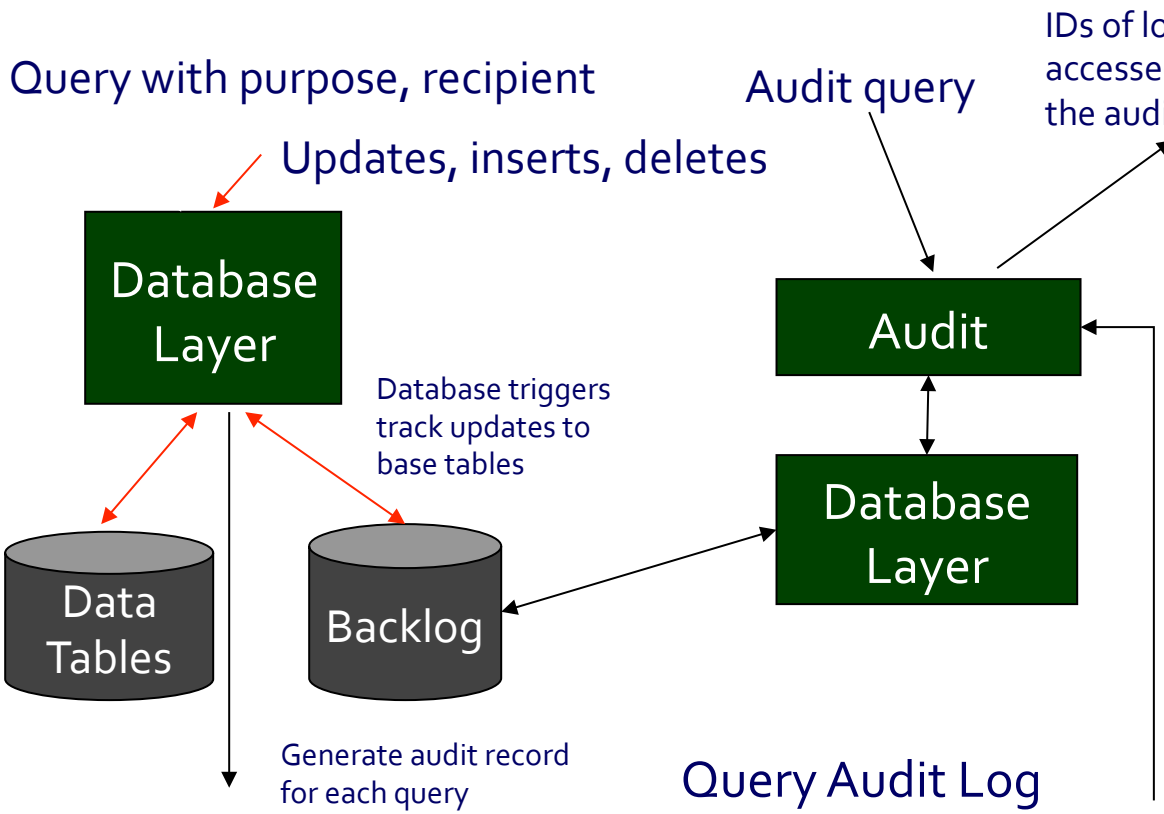
Case Study 1: Compliance Auditing

- Problem
- Solution
- Technical Details

Legal Compliance

- Companies are required to comply with many laws concerning the collection, use and disclosure of sensitive information
  - Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule
  - Compliance with 21 CFR Part 11 auditing requirements
- Compliance with these laws is difficult to implement and monitor
  - Auditing viewed as a nuisance, etc. etc.
- Companies need a way to automate enforcement of these laws and verify compliance

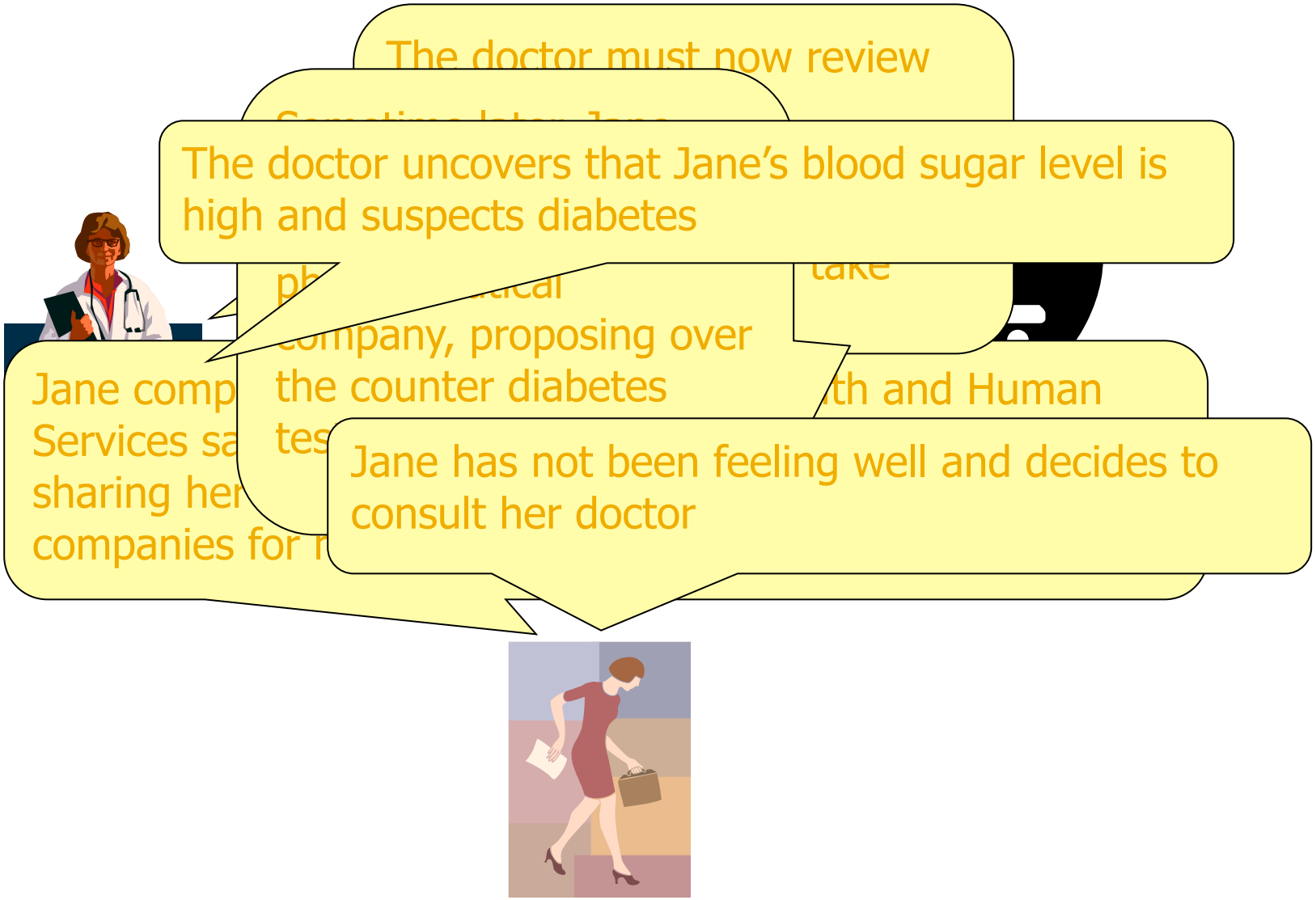
Solution in Action



ID	Timestamp	Query	User	Purpose	Recipient
1	2004-02…	Select …	Jane	Current	Ours
2	2004-02…	Select …	John	Telemarketing	public

- Audits whether particular data has been disclosed in violation of the specified policies
- Audit expression specifies what potential data disclosures need monitoring
- Identifies logged queries that accessed the specified data
- Analyze circumstances of the violation
- Make necessary corrections to procedures, policies, security

Audit Scenario



Audit Expression

Who has accessed Jane's disease information?

audit T.disease

from Customer C, Treatment T

where C.cid=T.pcid **and** C.name = 'Jane'

CS Problem Statement

- Given
  - A log of queries executed over a database
  - An audit expression specifying sensitive data
- Precisely identify
  - Those queries that accessed the data specified by the audit expression

Informal Definitions

- "Candidate" query
  - Logged query that accesses all columns specified by the audit expression
- "Indispensable" tuple (for a query)
  - A tuple whose omission makes a difference to the result of a query
- "Suspicious" query
  - A candidate query that shares an indispensable tuple with the audit expression

Query *Q*: Addresses of people with diabetes
Audit *A*: Jane's diagnosis

Jane's tuple is indispensable for both;
hence query *Q* is "suspicious" with respect to *A*

Suspicious Query

The candidate query *Q* and the audit expression *A* are of the form:

$$Q = \bar{\pi}_{CoQ}(\sigma_{PQ}(T \times R))$$
$$A = \bar{\pi}_{CoA}(\sigma_{PA}(T \times S))$$

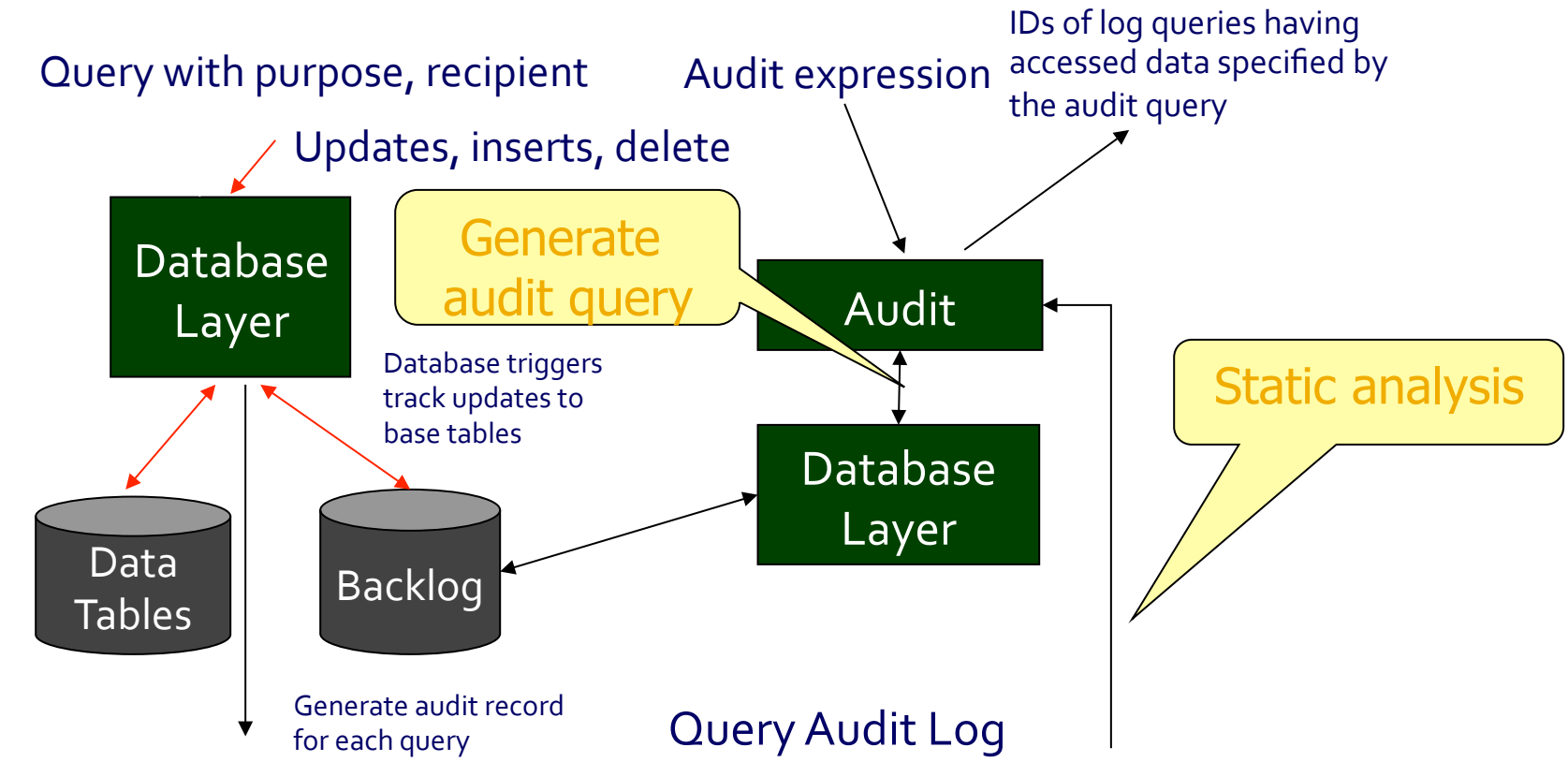
Theorem - A candidate query *Q* is suspicious with respect to an audit expression *A* iff:

$$\sigma_{PA}(\sigma_{PQ}(T \times R \times S)) \neq \varphi$$

Query Graph Modeler (QGM) rewrites *Q* and *A* into:

$$\pi_{"Q_i"}(\sigma_{PA}(\sigma_{PQ}(T \times R) \times S))$$

System (in Progress)

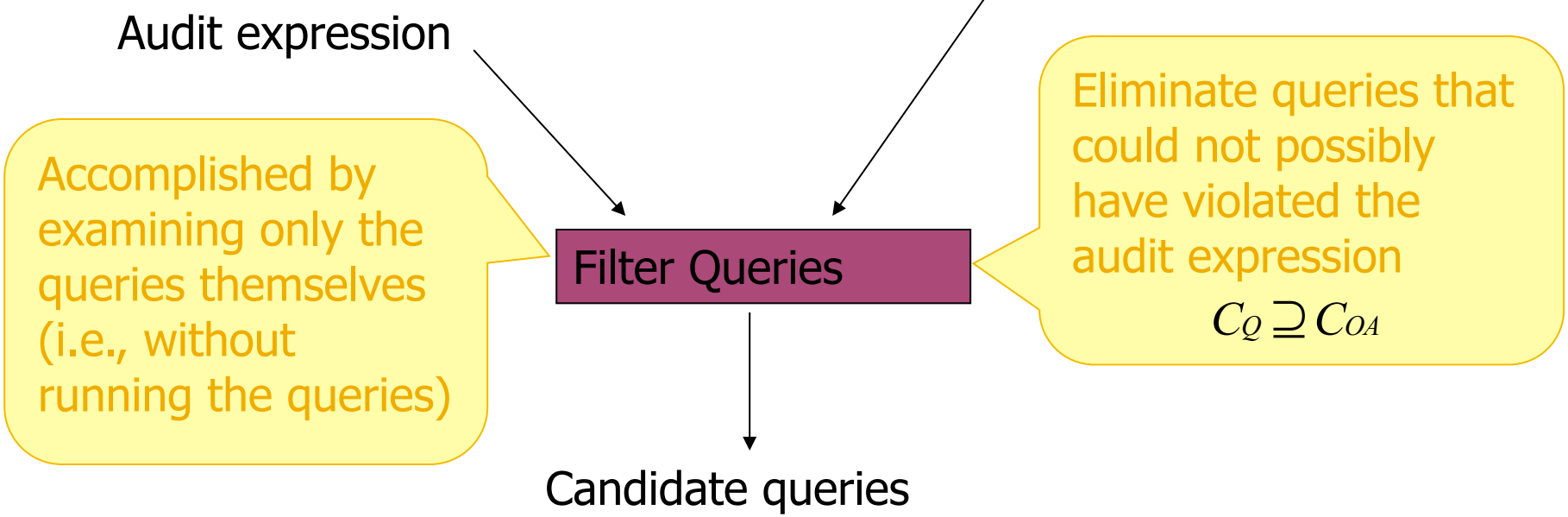


ID	Timestamp	Query	User	Purpose	Recipient
1	2004-02…	Select …	Jane	Current	Ours
2	2004-02…	Select …	John	Telemarketing	public

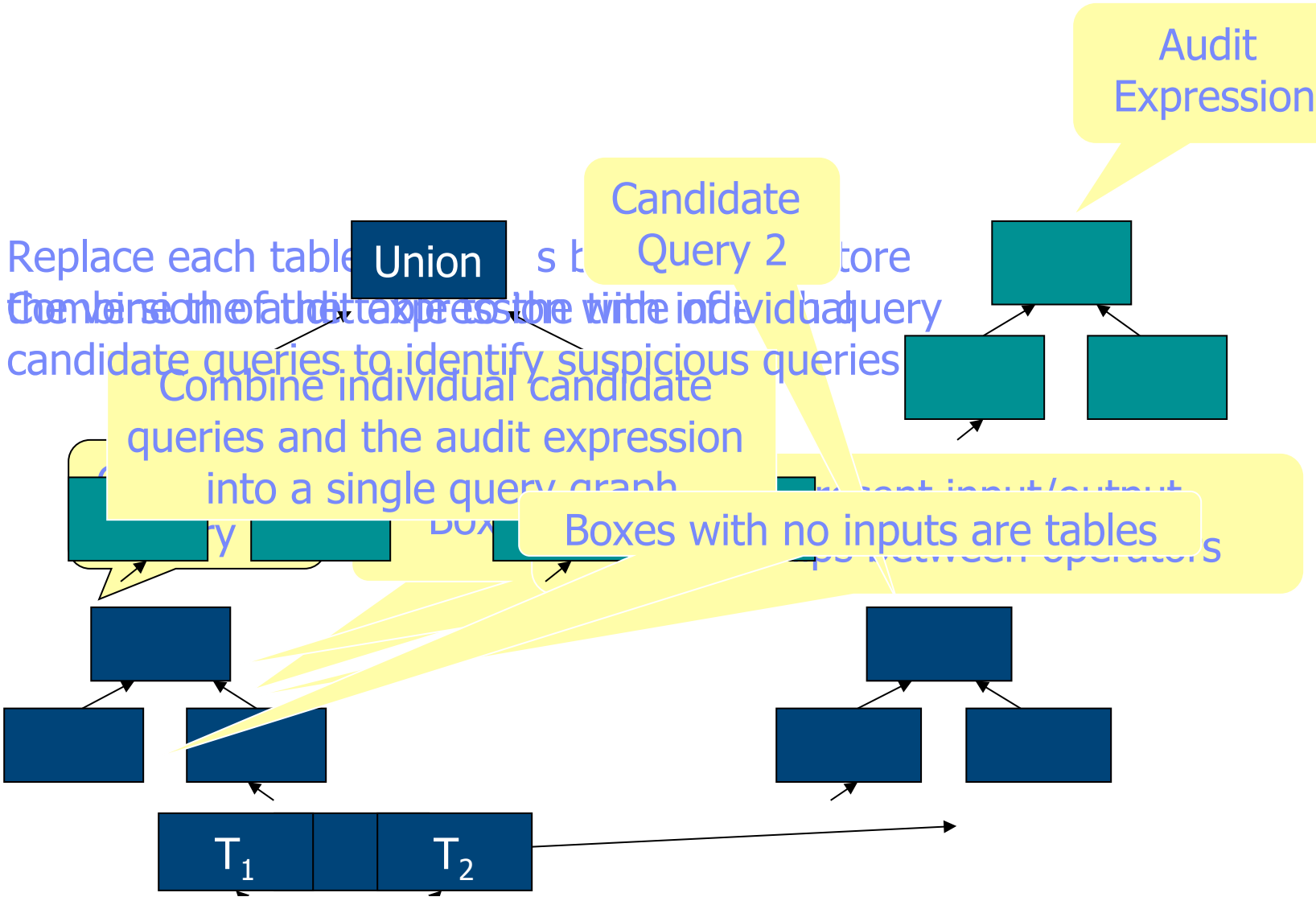
Static Analysis

Query Log

ID	Timestamp	Query	User	Purpose	Recipient
1	2004-02…	Select …	James	Current	Ours
2	2004-02…	Select …	John	Telemarketing	public

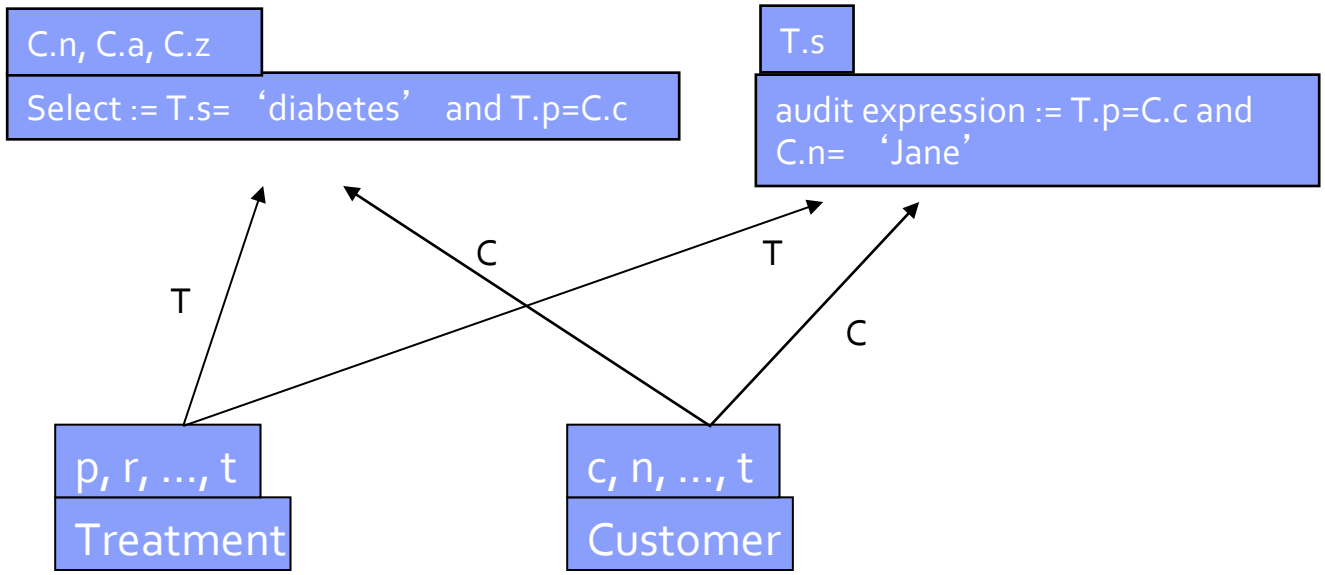


Generating the Audit Query

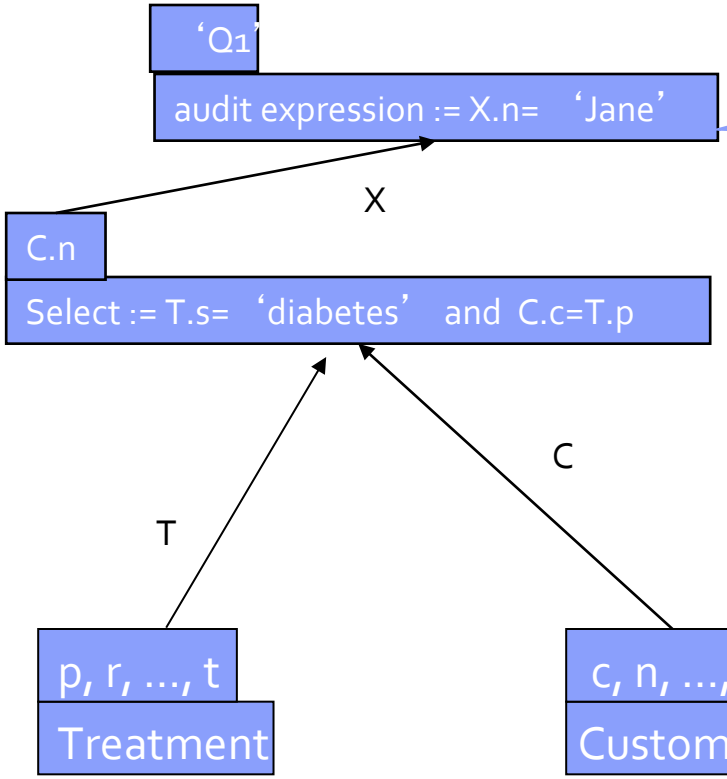


Merge Logged Queries and Audit Expression

Merge logged queries and audit expression into a single query graph



Transform Query Graph into an Audit Query



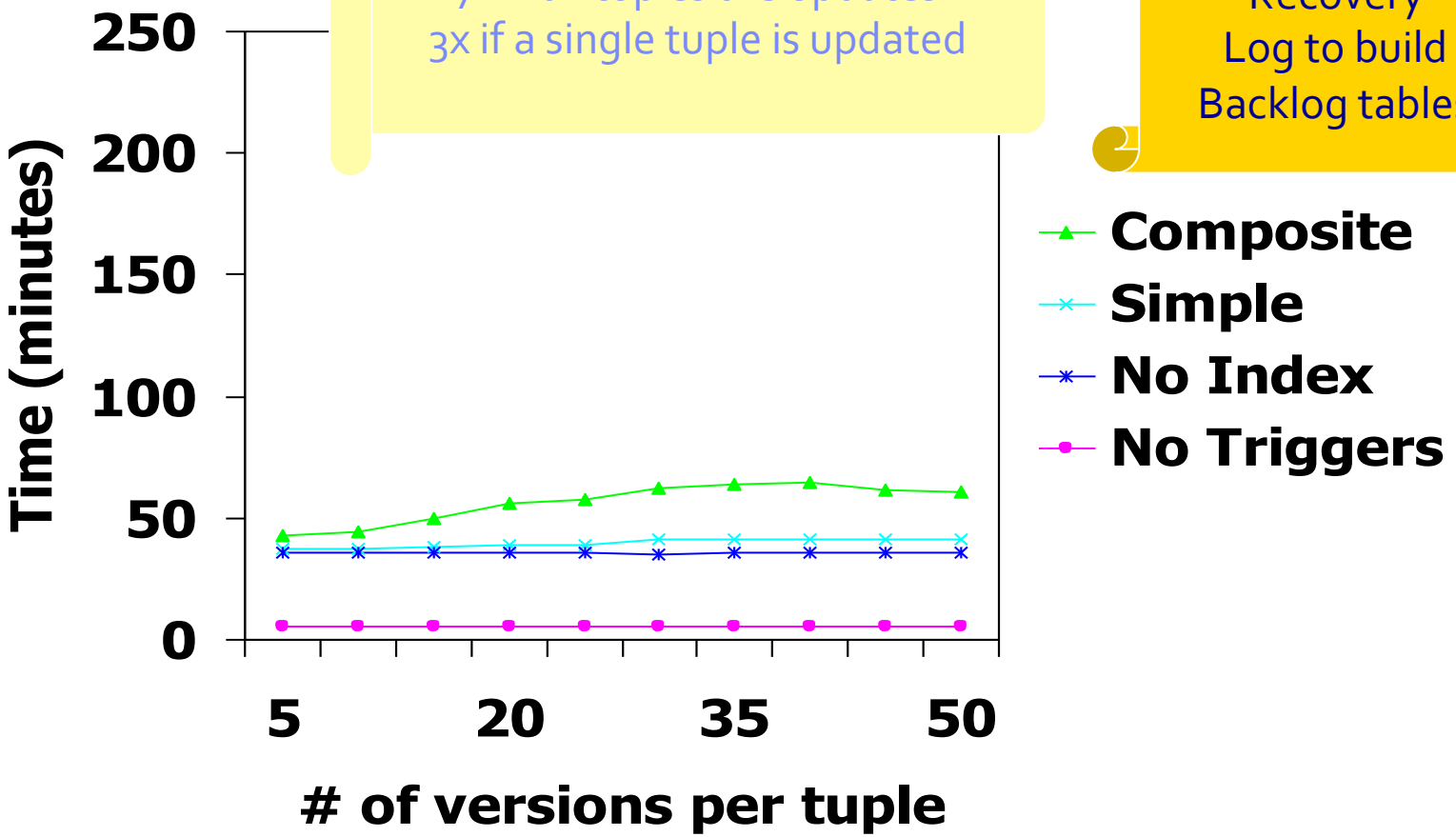
The audit expression now ranges over the logged query. If the logged query is suspicious, the audit query will output the id of the logged query

View of Customer (Treatment) is a temporal view at the time of the query was executed

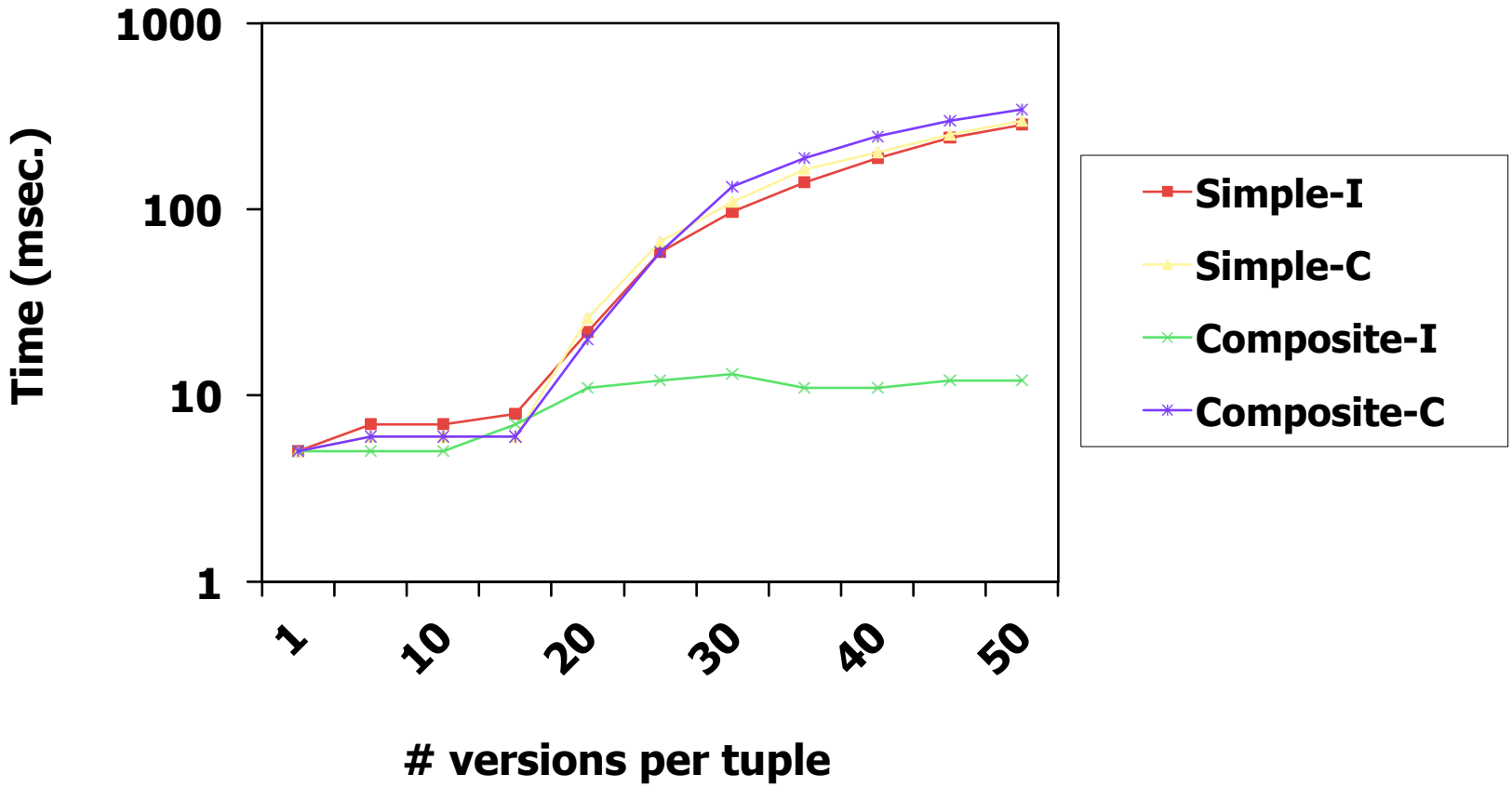
Overhead on Updates

7x if all tuples are updates
3x if a single tuple is updated

Negligible by using Recovery Log to build Backlog tables



Audit Query Execution Time



Solution Features

Access and Disclosure Tracking

- Audit trails provide detailed information about data access, changes, insertions and deletions.
- Facilitates investigations of data access, use, and disclosure.

Compliance Verification

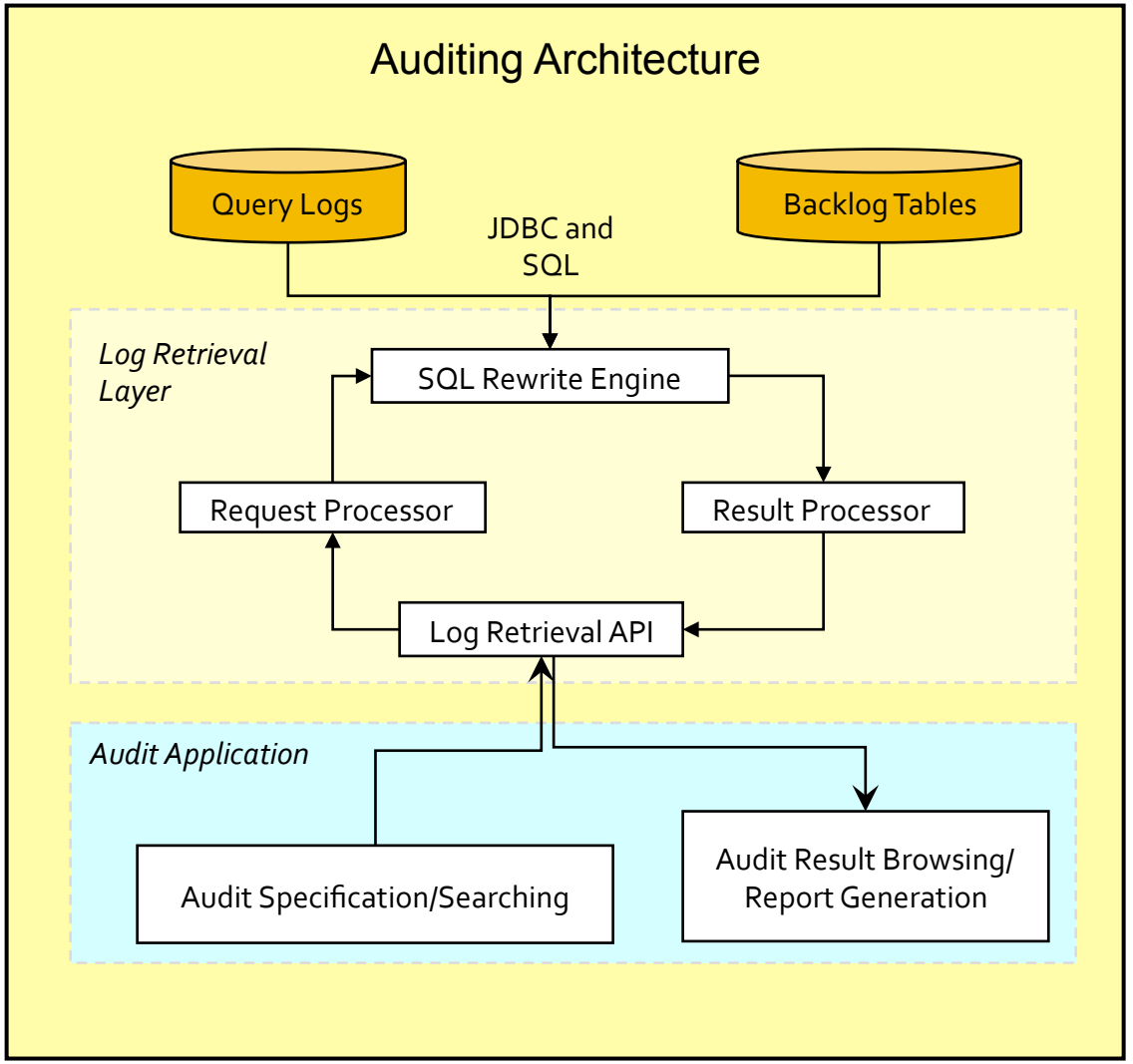
- Determines whether a particular disclosure or transaction was compliant with policies (e.g., legal requirements).

Data Recovery

- Reconstructs the exact state of any cell in the database at a given point in time.

Audit Flags

- Alert companies to suspicious data access and disclosures or policy violations.



Case Study Summary

- The overhead on query processing is small, involving primarily the logging of each query string along with other minor annotations.
- Database triggers are used to capture updates in a backlog database.
- At the time of audit,
  - a static analysis phase selects a subset of logged queries for further analysis.
  - These queries are combined and transformed into an SQL audit query, which when run against the backlog database, identifies the suspicious queries efficiently and precisely.

Case Study 2: Exception Based Access

- Problem
- Solution
- Technical Details

Problem

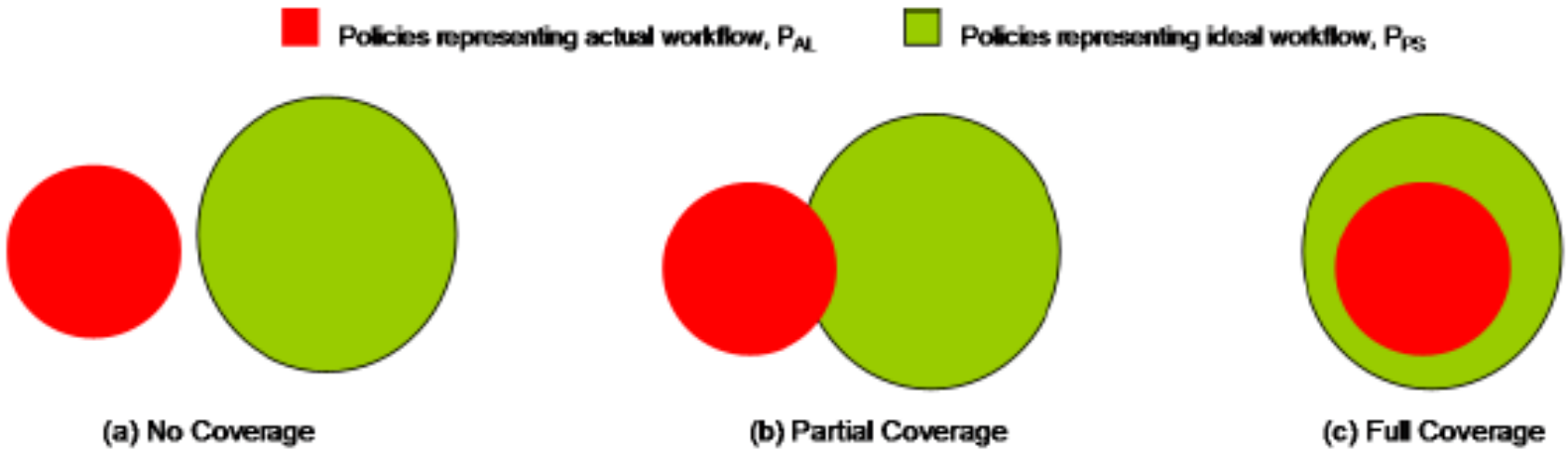
- A lot of the information in audit logs is misleading
  - Audit logs are bypassed by people legitimately doing their jobs 70% of the time.
  - Thus, logs contain legal activity, unformalized activity and security breaches.
- Goals:
  - Reduce wasted effort on differentiating between undocumented legal behavior and a cyber-attack.
  - Enable security/privacy policy to encapsulate rules.

Contribution

- **Formalizing Policy Refinement Process**
  - Introduce the notion of Policy Coverage
  - Design the algorithm for Policy Refinement

- **PRIvacy Management Architecture (PRIMA)**
  - An architecture designed to perform Policy Refinement
  - Leverages data mining and Hippocratic Database (HDB) technology

Coverage & Policy Refinement (Illustrated)



Formal Model

Consider an Organization (HO):

- **Ideal Workflow W_{Ideal}:**
  - HO's policy embodies regulations, legislations, laws. Essentially, what HO would ideally like to follow
- **Real Workflow W_{Real}:**
  - HO's policy as represented by the audit trail of system accesses over a period of time
  - The real workflow of HO, primarily filled with exception-based accesses
- Our Goal is to reduce of gap between real and ideal workflows
- The formal model is used to represent
  - the privacy specification notation, which comprises the W_{Ideal}
  - the artifacts that the system manipulates, which comprises the W_{Real}
  - the mapping from the terms in W_{Ideal} to the corresponding terms in W_{Real}

Core Constructs

- **RuleTerm:**
  - Models the assignment of attributes in a policy rule

Definition 1. *(RuleTerm): A RuleTerm (RT) is a tuple with two literal-valued elements, attr and value. It is written as* $RT = (attr, value)$*. The two elements of RT are accessed as* $RT.attr$ *and* $RT.value$*.* □

- **Rule:**
  - Models a specific combination of attribute assignments

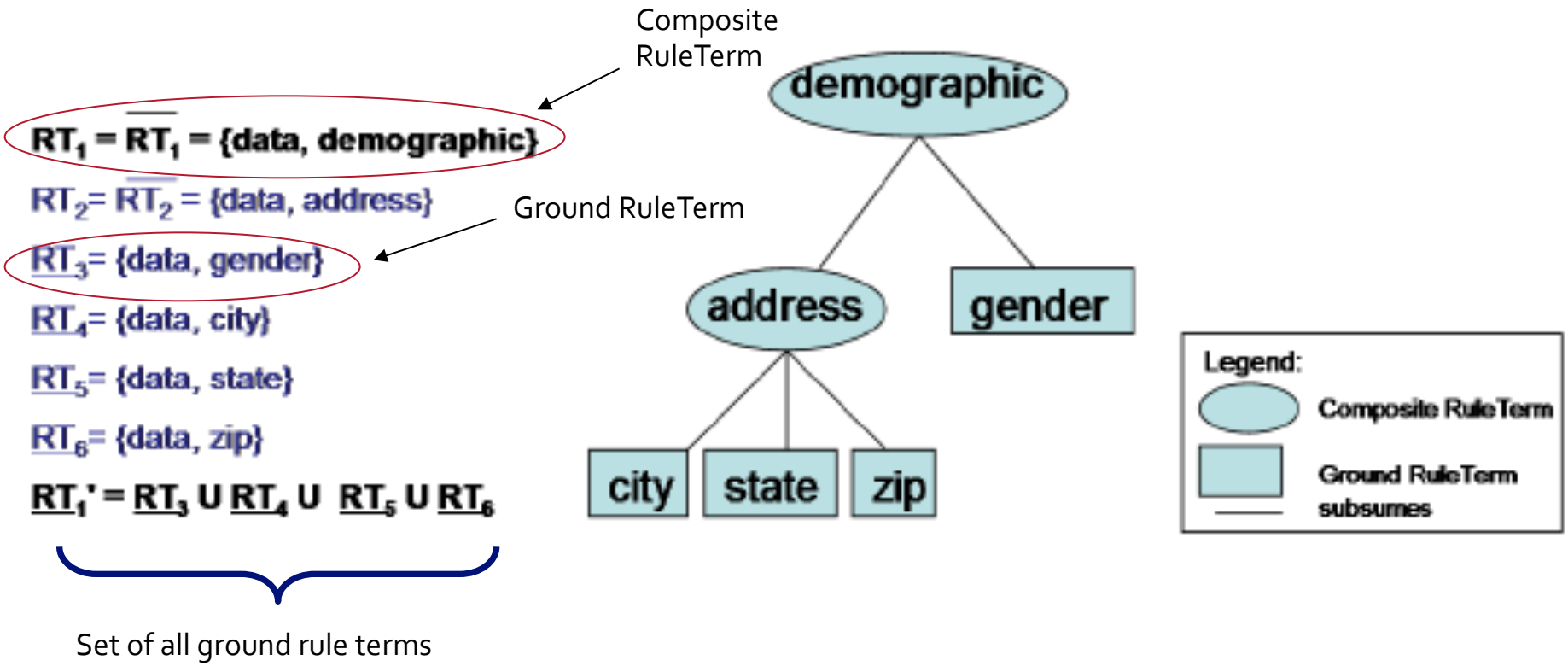
Definition 5. *(Rule): A Rule,* R_i*, is a conjunction of RuleTerms. It is written as* $R_i = \{RT_1 \wedge \ldots \wedge RT_n\}, n \geq 1$*. The number of RuleTerms of a Rule, n, is referred to as the cardinality of the Rule, written as* $\#R$*.* □

Two types of RuleTerms and Rules:

- Ground - If comprises entirely of atomic attributes
- Composite – Otherwise

A policy is ground when represented only in terms of ground rules

Example Policy Vocabulary



Policy Coverage

Coverage is computed by comparing $\mathbf{P_{AL}}$ and $\mathbf{P_{PS}}$

- $\mathbf{P_{AL}}$ is the policy found in the audit logs representing the real state of system
- $\mathbf{P_{PS}}$ is the policy found in the policy store representing ideal state of system

- **Informally:**
  - Coverage is the overlap between $\mathbf{P_{AL}}$ and $\mathbf{P_{PS}}$

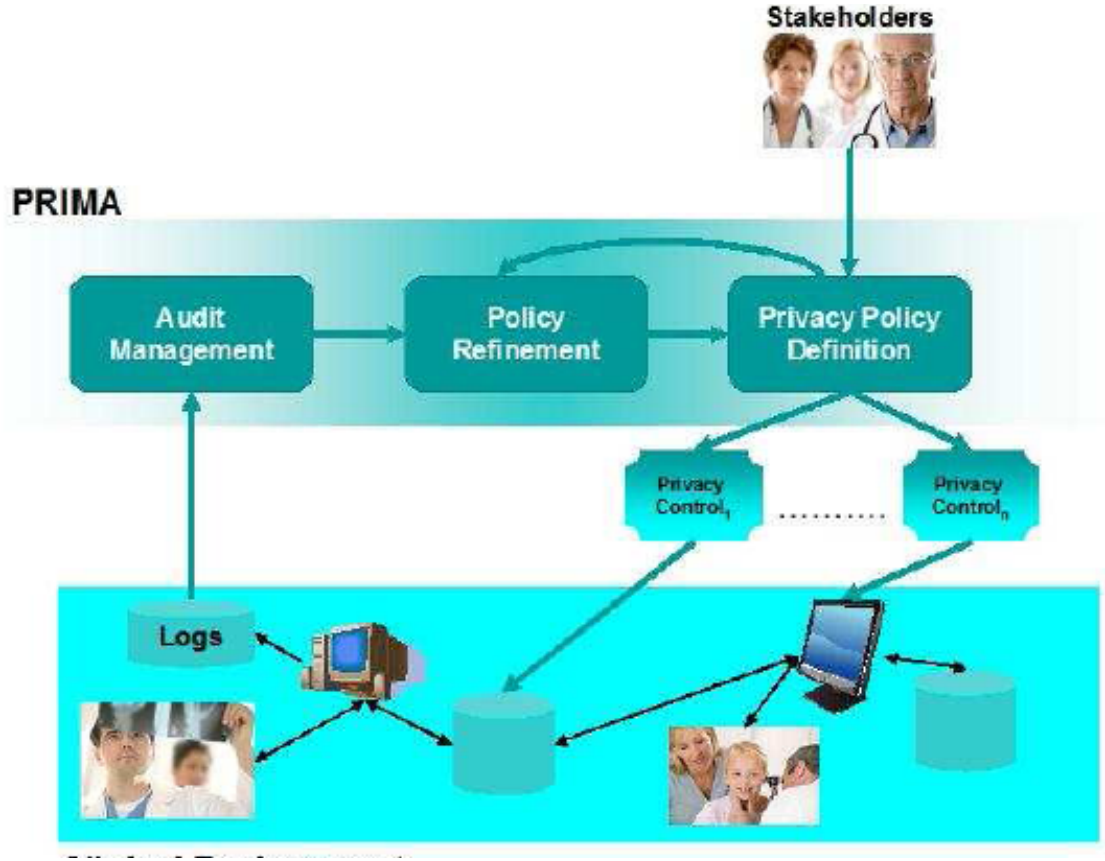
- **Formally:**
  - Given

    $Range_P$ as the set containing all the rules in a ground policy P, and $\# Range_P$ as the cardinality of $Range_P$
  - Coverage of $\mathbf{P_x}$ in relation to $\mathbf{P_y}$ is given by

    $\# (Range_{Px} \cap Range_{Py}) \div \# Range_{Py}$

Goal is to have complete coverage, i.e. $Range_{Px} \cap Range_{Py} = Range_{Py}$

System Overview



Audit Management

- **Use storage efficient, contextually rich logs**
  - Log management and usability is better
- **Use logs in a pro-active process, as opposed to after-the-fact**
  - Consolidate all logs in one place
- **Fix a schema for the log entries**
  - Current schema is

(time,tj), (op,Xj),(user,uj), (data,dj), (purpose,pj), (authorized,aj), (status,sj) where

tj is the entry's timestamp
Xj is either 0 (disallow) or 1 (allow)
uj is the entity that requested access
dj is the data to be accessed
pj is the purpose for which the data is accessed
aj is the authorization category (e.g. role) of the entity that requested access, and *sj* is either 0 (exception-based access) or 1 (regular access).

Policy Refinement

- **Leverage audit logs**
  - Analyze all entries that are regular accesses
  - Define new rules based on analysis
- **Improve the policy coverage**
  - Coverage is the ratio of accesses addressed by the policy to all access recorded by the system
- **Gradually embed policy controls**
  - Essentially, a feedback loop between ideal and real policy

Refinement Algorithm

- **Filter**
  - Flag *exceptions* to distinguish them from regular accesses
    - Analyze only the regular accesses for possible patterns
- **Extract**
  - Find informal clinical patterns from audit logs
  - Apply algorithm to extract candidate patterns
    - Simple matching:
      - Assumes pruned data, looks for term combinations, returns frequency of occurrence
    - Richer data mining:
      - Not only syntactic but also semantics matching
      - Does not assume pruning, considers relationship between artifacts
      - Reduces probability of violations being reported for analysis phase
  - Get usefulness ratings of patterns
- **Prune**
  - Incorporate or discard patterns based on usefulness threshold
  - Assume a training period
    - Set a threshold appropriate to the target environment
    - Act when threshold is reached over a period of time

Example

Policy coverage is 30% (3/10)

Time	**Op** (1:allow)	**User**	**Data** (Category)	**Purpose**	**Authorized** (Role)	**Status** (0: Exception)
t1	1	John	**Prescription**	**Treatment**	**Nurse**	**1**
t2	1	Tim	**Referral**	**Treatment**	**Nurse**	**1**
t3	1	Mark	Referral	Registration	Nurse	0
t4	1	Sarah	Psychiatry	Treatment	Doctor	0
t5	1	Bill	**Address**	**Billing**	**Clerk**	**1**
t6	1	Jason	Prescription	Billing	Clerk	0
t7	1	Mark	Referral	Registration	Nurse	0
t8	1	Tim	Referral	Registration	Nurse	0
t9	1	Bob	Referral	Registration	Nurse	0
t10	1	Mark	Referral	Registration	Nurse	0

Audit trail, $\mathbf{P_{AL}}$, for a system

Audit Log after "Filter"

Time	**Op** (1:allow)	**User**	**Data** (Category)	**Purpose**	**Authorized** (Role)	**Status** (0: Exception)
t3	1	Mark	Referral	Registration	Nurse	0
t4	1	Sarah	Psychiatry	Treatment	Doctor	0
t6	1	Jason	Prescription	Billing	Clerk	0
t7	1	Mark	Referral	Registration	Nurse	0
t8	1	Tim	Referral	Registration	Nurse	0
t9	1	Bob	Referral	Registration	Nurse	0
t10	1	Mark	Referral	Registration	Nurse	0

Mining Rule in "Extract"

```
SELECT A.Data, A.Purpose, A.Authorized
FROM P_AL A
WHERE A.Status = 'o'
GROUP BY A.Data, A.Purpose, A.Authorized
HAVING COUNT(*) > 5 AND
   COUNT(DISTINCT(A.User)) > 1;
```

Output of "Extract"

Time	**Op** (1:allow)	**User**	**Data** (Category)	**Purpose**	**Authorized** (Role)	**Status** (0: Exception)
t3	1	Mark	**Referral**	**Registration**	**Nurse**	0
t4	1	Sarah	Psychiatry	Treatment	Doctor	0
t6	1	Jason	Prescription	Billing	Clerk	0
t7	1	Mark	**Referral**	**Registration**	**Nurse**	0
t8	1	Tim	**Referral**	**Registration**	**Nurse**	0
t9	1	Bob	**Referral**	**Registration**	**Nurse**	0
t10	1	Mark	**Referral**	**Registration**	**Nurse**	0

Pattern found:
Referral: Registration : Nurse
occurred in the log at least 5 times
observed for at least 2 different users

Case Study Conclusion

- Formally introduced the problem of Policy Coverage to help mitigate the issues in privacy management resulting from exception-based accesses
- Defined the notion of Policy Refinement for improving policy coverage through a systematic, non-disruptive, approach that aims to gradually embed privacy controls within the workflow based on actual practices of the organization.

Case Studies: Summary

- Made first steps to solve:
  - Create efficient auditing systems
  - Pruning audit systems
- Implemented in the context of multiple client engagement.
- Applicable to multiple fields.

Future Work

- Fundamental technologies for critical infrastructure protection
- Privacy-preserving and secure solutions for enabling Cloud and Big Data Analytics
- Solutions for Securing Mobile systems

Conclusion

- Cybersecurity is about protecting, repelling and recovering from cyberattacks
- Cyberattacks are a silent norm
- Our greatest near-term impact lies in using efficient audit systems to detect and respond to security incidents.

Thank you
Any Questions?

tgrandison@proficiencylabs.com
http://www.tyronegrandison.org
http://www.proficiencylabs.com

Sources

- *2013 Cyber Security Study - What is the Impact of Today's Advanced Cyber Attacks? - Bit9 and iSMG*
- *INFOSEC Research Council (2005)*
- *INFOSEC Research Council (2009)*
- Claudio A. Ardagna, Sabrina De Capitani di Vimercati, Sara Foresti, Tyrone W. Grandison, Sushil Jajodia, and Pierangela Samarati. " Access control for smarter healthcare using policy spaces." *Computers & Security* 29, no. 8 (2010): 848-858.
- Christopher Johnson, **Tyrone Grandison**, "Compliance with Data Protection Laws Using Hippocratic Database Active Enforcement and Auditing," IBM Systems Journal, Vol. 46, No. 2, April 2007.
- Rakesh Agrawal, Roberto J. Bayardo Jr., Christos Faloutsos, Jerry Kiernan, Ralf Rantzau, Ramakrishnan Srikant: Auditing Compliance with a Hippocratic Database. VLDB 2004: 516-527
- Rafae Bhatti, **Tyrone Grandison**. "Towards Improved Privacy Policy Coverage in Healthcare Using Policy Refinement". The Proceedings of the 4th VLDB Workshop on Secure Data Management 2007. Vienna, Austria, Sept 2007.