PBKM: A Secure Knowledge Management Framework

Shouhuai Xu Weining Zhang

Dept. Computer Science, UT @ San Antonio

Roadmap

- Motivation
- Properties of KMS
- The PBKM Framework
- Instantiations of PBKM Framework
- Related Work
- Challenges and On-going Work

Motivation

- When we had large volumes of data
  - DBS/DBMS was invented to manage data
- When the volume of data was too large
  - Data mining was invented to extract knowledge
- When we get a large amount of knowledge
  - We envision Knowledge Management System

Motivation

- Knowledge Management System (KMS) is an analogy of DataBase Management System (DBMS)
- Why do we need KMS?
  - Sharing of data might be prohibited, but sharing of (the hidden) knowledge is not
  - Knowledge extracted from a joint database is more useful

Roadmap

- Motivation
- Properties of KMS
- The PBKM Framework
- Instantiations of PBKM Framework
- Related Work
- Challenges and On-going Work

Terminology

- Knowledge: knowledge models (e.g., decision trees, association rules, neural networks) extracted from raw data and expressed in a certain knowledge representation language
- Knowledge Management: methodology for systematically extracting and utilizing knowledge
- KMS: enabler of knowledge management

Functionalities of KMS

- KMS is a platform facilitating extraction, storage, retrieval, integration, transformation, visualization, analysis, dissemination, and utilization of knowledge
- Quite similar to a DBMS

Security Requirements on KMS

- Desired security properties
  - Access control
  - Privacy-preservation
  - Breaching-awareness
  - Abuse-accountability
- Quite different from security in DBMS

Security Requirements on KMS

- Access control
  - A certain policy/objective
  - A certain model (MAC, DAC, RBAC)
  - A certain architecture
  - Enforcement mechanisms

Security Requirements on KMS

- Privacy-preserving knowledge extraction
  - Multiple parties jointly extract knowledge from their databases without exposing individual data
  - Extraction is mainly based on mining
  - Gap between crypto approach and database approach

Security Requirements on KMS

- Breaching-aware knowledge dissemination
  - Breaching happens when knowledge owner is different from knowledge consumer
  - Owner holds K: Q $\rightarrow$ R based on its knowledge
  - Consumer queries q $\in$ Q', gets K(q), where Q' $\subset$ Q
  - Consumer knows K(q*) with a high probability, q* $\notin$ Q'

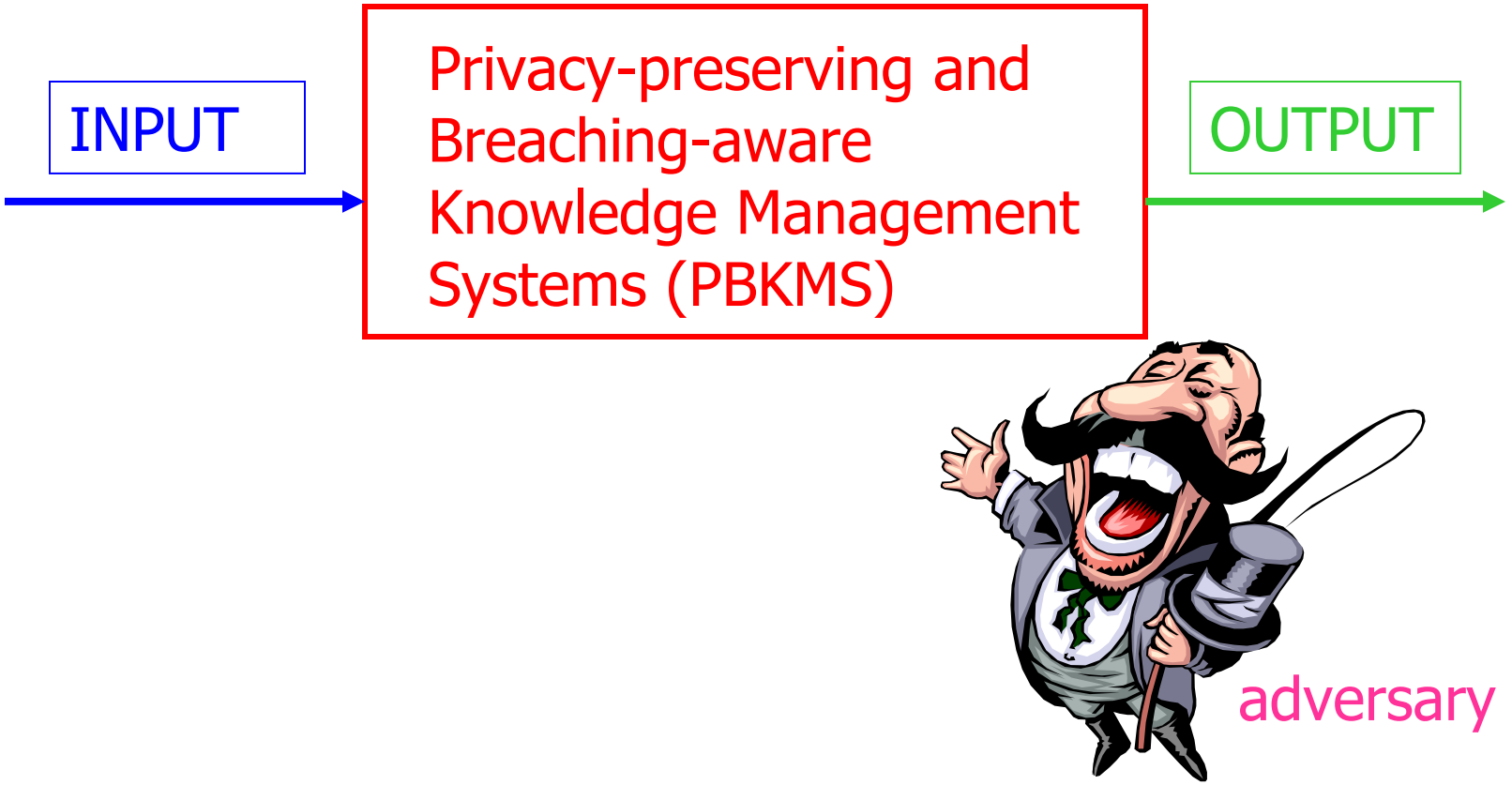
Security Requirements on KMS

- Abuse-accountability
  - Abuse of knowledge (e.g., insider) could result in catastrophic consequences
  - We need to hold abusers accountable
  - More than traditional auditing: automatically correlate incidents, even if data is encrypted

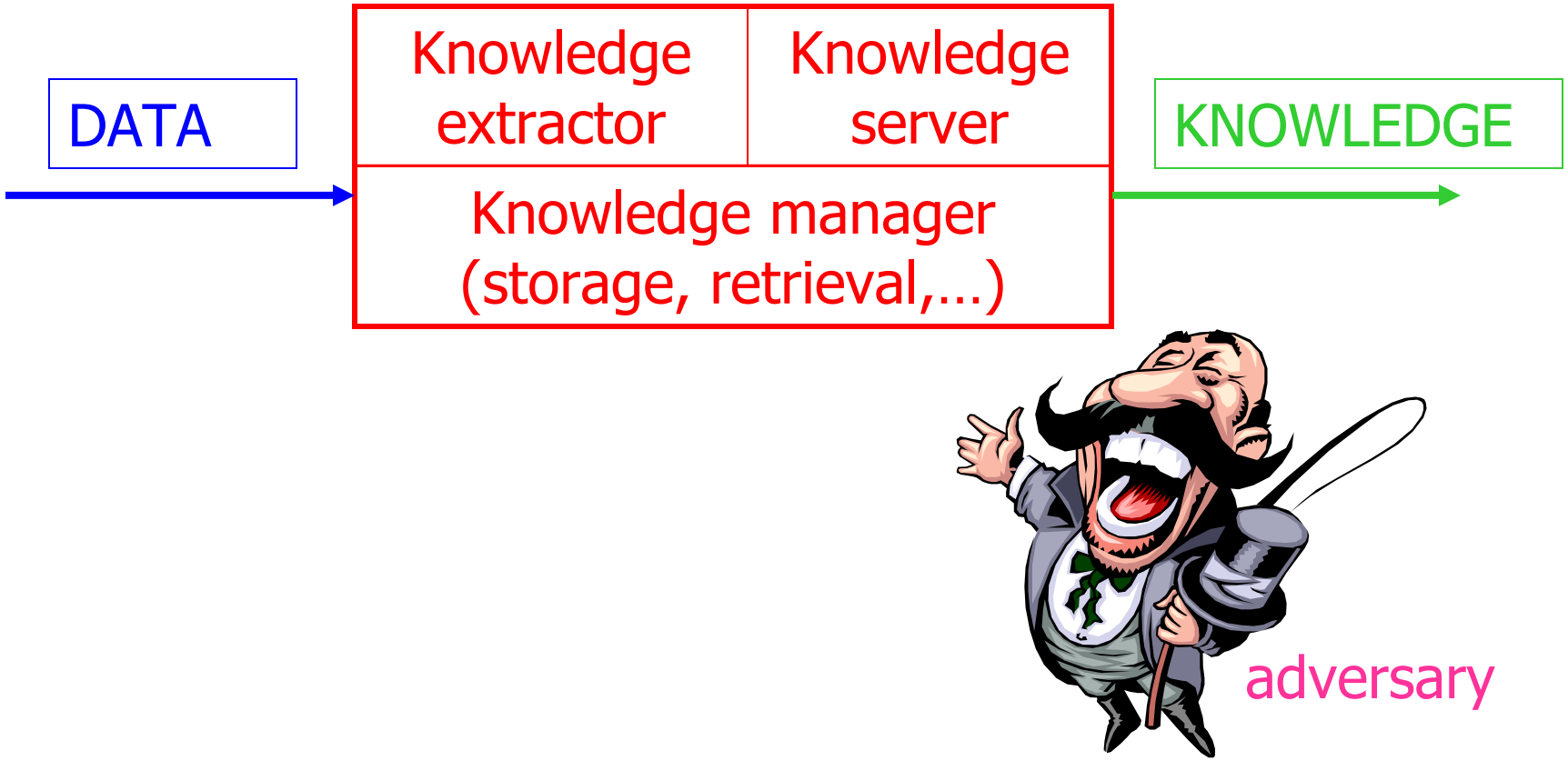
Roadmap

- Motivation
- Properties of KMS
- The PBKM Framework
- Instantiations of PBKM Framework
- Related Work
- Challenges and On-going Work

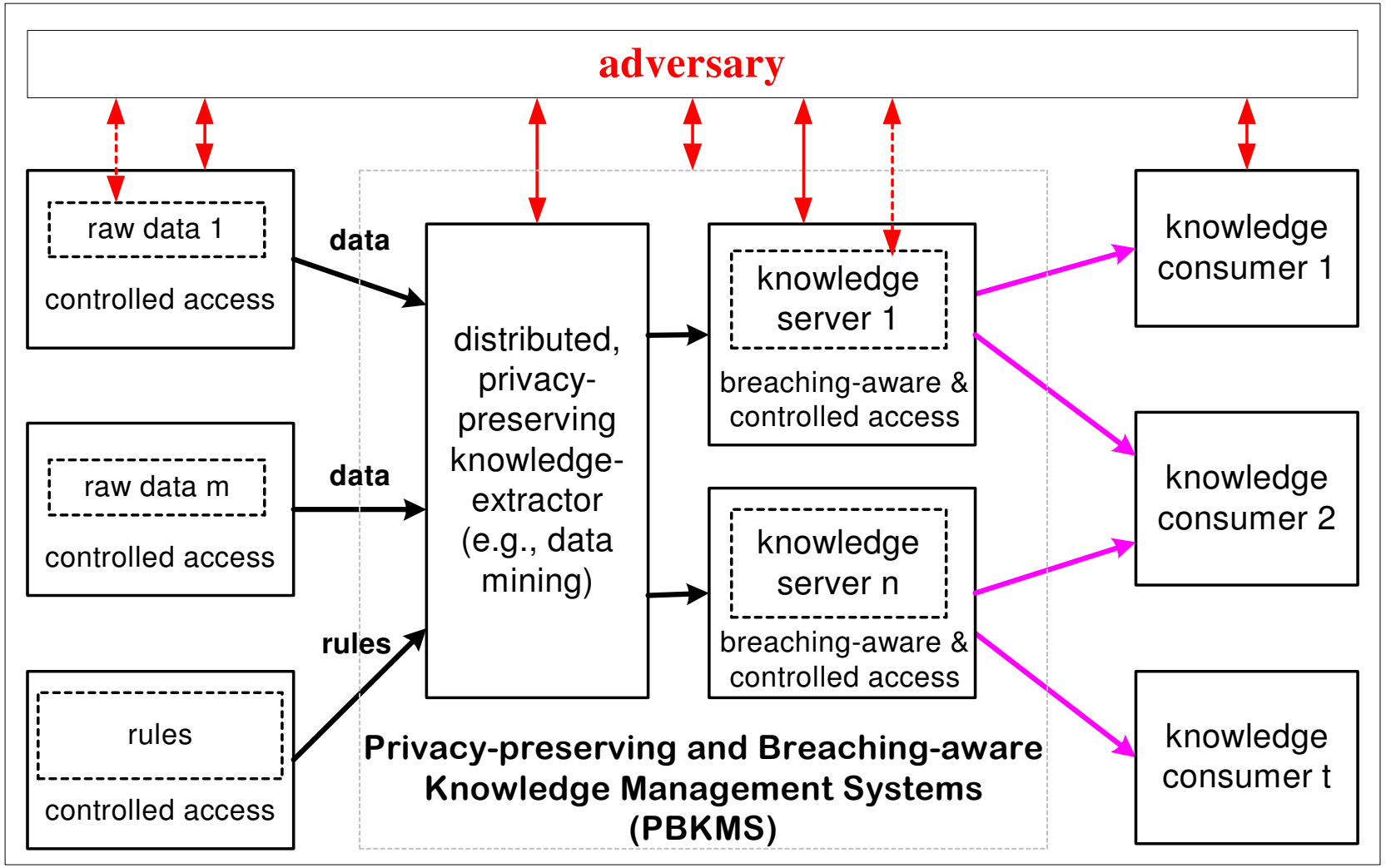
The PBKM Framework: high-level



The PBKM Framework: mid-level



The PBKM Framework: low-level



Roadmap

- Motivation
- Properties of KMS
- The PBKM Framework
- Instantiations of PBKM Framework
- Related Work
- Challenges and On-going Work

A Business Case

- A specific case of the PBKM framework, called "knowledge as a service" is investigated in a separate paper
- Scenario:
  - A life insurance company needs to know the likelihood of a new customer being involved in fatal car accidents
  - The likelihood can be extracted from the databases of the car insurance companies
  - A new venture capital can extract the knowledge from the databases, and sell the knowledge to the life insurance company (e.g., via queries based on individual record)

A Government Case

- Each agency has its own database
- In order to profile terrorists, they jointly run some knowledge extraction algorithm (if they don't want to share data)
- The extracted knowledge is shared among the agencies
- Abuse-accountability is important: leakage of profiling information would make law-enforcement much harder in counter-terror

Roadmap

- Motivation
- Properties of KMS
- The PBKM Framework
- Instantiations of PBKM Framework
- Related Work
- Challenges and On-going Work

Related Work

- An instantiation of PBKM is "knowledge as a service"
  - Application as a service, database as a service
- An instantiation of privacy-preserving knowledge extraction is privacy-preserving data mining
  - Accuracy vs. performance (crypto vs. perturbation)
- Data mining/machine learning for extracting knowledge
  - Can also be used to breach knowledge in knowledge dissemination

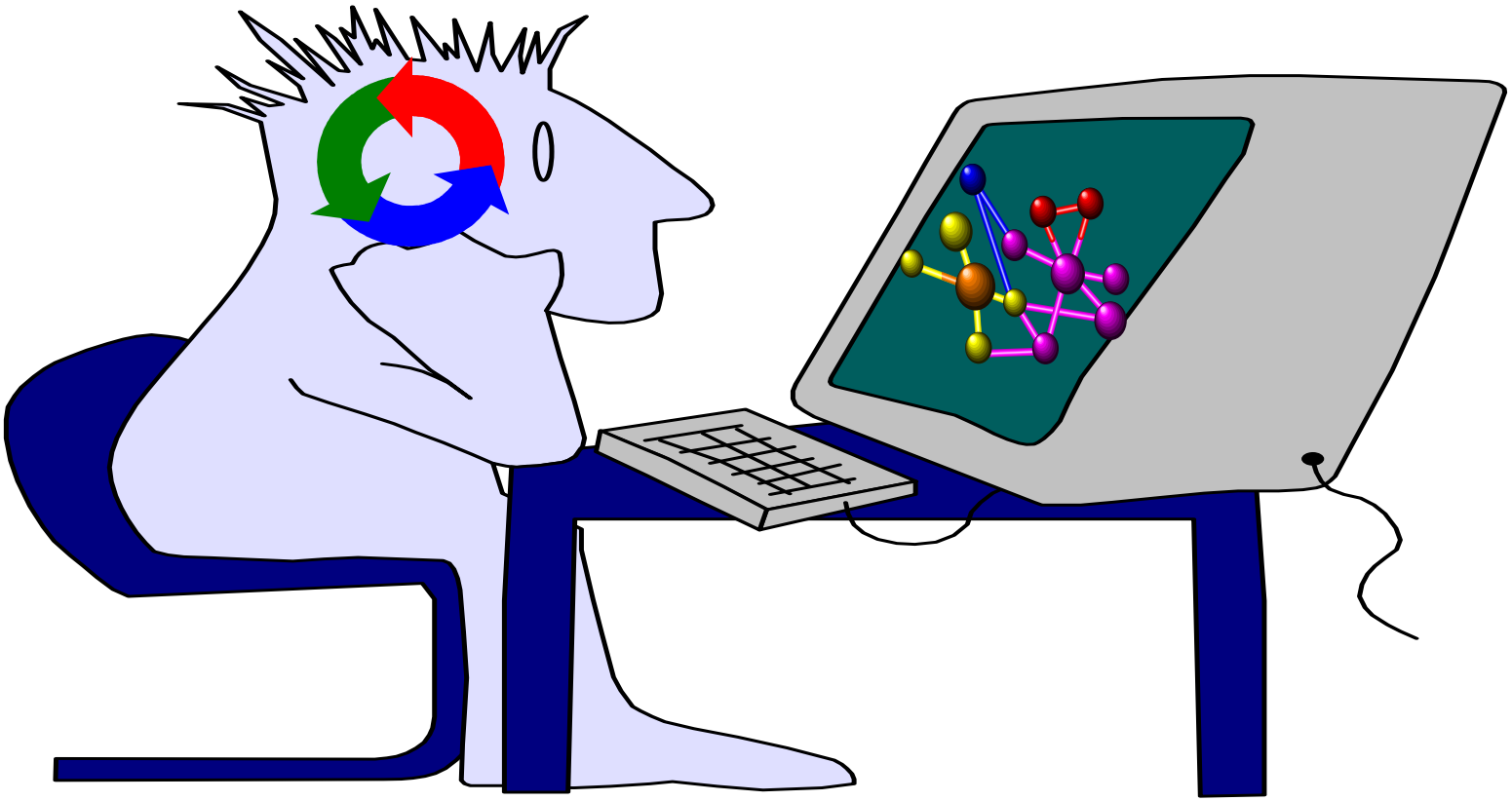
Roadmap

- Motivation
- Properties of KMS
- The PBKM Framework
- Instantiations of PBKM Framework
- Related Work
- Challenges and On-going Work

Challenges and On-going Work

- Develop practical and provably-secure privacy-preserving knowledge extraction techniques
- Knowledge breaching triggers many new questions
  - How the knowledge owners get compensated?
  - What is the foundation of knowledge breaching?
  - What is the metrics of knowledge leakage?
- Ensure abuse-accountability automatically, even if the data is encrypted

Questions?



Thoughts after Two Buffalo Days

- What is knowledge management?
  - Unifier of techniques (DL, semantic web, DB, data mining, etc.)
  - Analogy: I need a car (i.e., sharing of knowledge) not the parts (e.g., engines can be used in many different machines)
- Where does (extended) PBKM stand?

Where does PBKM stand?

Jim Gray's