Hochschule Zittau/Görlitz
UNIVERSITY OF APPLIED SCIENCES

Enterprise Application Development Group http://www.hszg.de/ead
Faculty of Electrical Engineering and Computer Science

Corporate Privacy Preserving Data Analysis (CoPPDA)

A project for developing a privacy preserving data mining tool.

Jörg Lässig, Daniel Tasche, Jens Heider, Hendrik Kunert

University of Applied Sciences Zittau/Görlitz, Department of Computer Science, Görlitz, Germany

Introduction

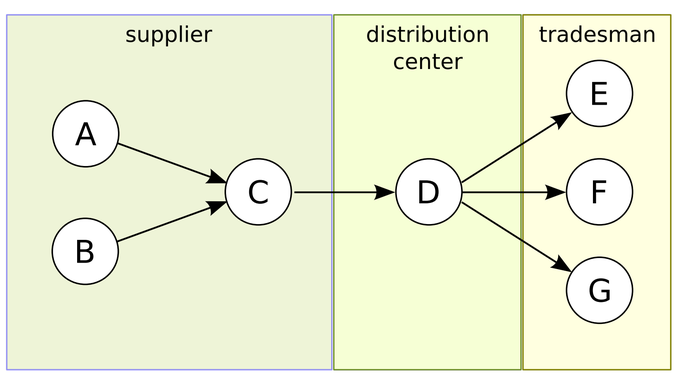
- **companies** mainly use two approaches for raising their efficiency:
  - modern information and communication systems
  - inter organizational networks
- combinational approach of these two strategies is not very common
- great potential for cost and process optimization
- cross company data mining not possible, because of strategic and juristic restraints (e.g. sharing of in-house data, personal information)

⇒ **CoPPDA** as privacy preserving data mining tool.

Goals

- implementation of the following privacy preserving data mining algorithms:
  - ID3 Decision Tree over vertical partitioned data
  - ID3 Decision Tree over horizontal partitioned data
  - Back-Propagation Networks
  - k-means over vertical partitioned data
  - k-means over horizontal partitioned data
- web-service for coordination of communication
- **RapidMiner** plug-in

Example



Classification rules

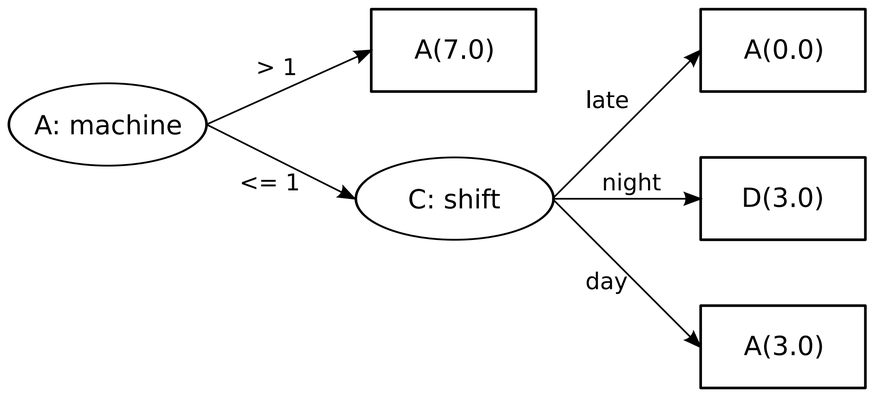
1. products produced by company A on machine 1 are error free
2. products produced by company A on machine 1, that have been processed on day or late shift by company C are error free
3. products produced by company A on machine 1, that have been processed on night shift by company C are faulty

⇒ rules 2 and 3 could only be build with the help of cross company data mining

Preconditions

- A, B and C are component supplier, which are involved in producing some product
- the distribution center D is responsible for classical commissioning tasks
- E, F and G represent tradesman and distributors
- each product has an RFID chip attached

⇒ data mining on measurement data of companies A, B and C could be used to detect cross-company failure causes (e.g. creation of classification tree, like Fig. 2)



Problems

- current solutions depend on data exchange
- data exchange often is not possible, because of strategic and juristic restraints

⇒ **CoPPDA** brings privacy into data mining.

Novelty of the approach

- usage of privacy preserving data mining methods

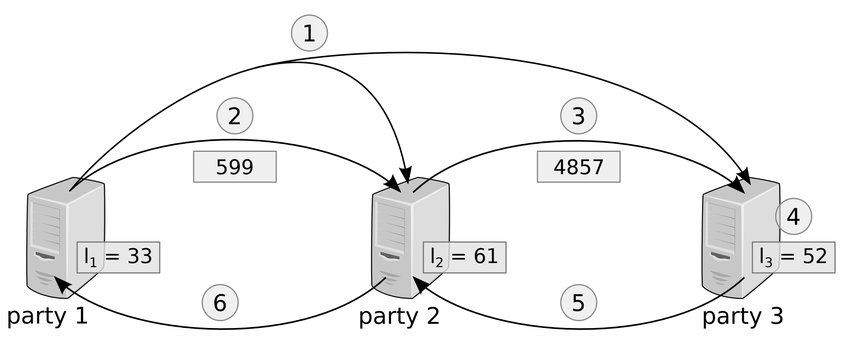
Master Thesis

Privacy Preserving ID3 over horizonal data using Gini index

Paillier Cryptosystem

- additive homomorphic
- essential part of our privacy preserving computations
- Encryption:
  1. plaintext $m < n$
  2. select a random $r < n$
  3. ciphertext $c = g^m \cdot r^n \mod n^2$
- Decryption
  1. ciphertext $c < n^2$
  2. plaintext $m = \frac{\mathsf{L}(c^\lambda \mod n^2)}{\mathsf{L}(g^\lambda \mod n^2)} \mod n$

Example: Addition



- using a additive homomorphic encryption E with the public-key e and the private-key d:

$$\prod_{i=1}^{k} E(x_i, e) = E\left(\sum_{i=1}^{k} x_i, e\right) = E\left(\prod_{i=1}^{k} l_i, e\right) = E\left(l_1, e\right)^{l_2^{\cdot^{\cdot^{l_k}}}} . \quad (1)$$

Goals

- generic communication system for all privacy preserving methods shown above
- implementation of ID3 Decision Tree over horizontal partitioned data
- prototypical RapidMiner plug-in for ID3 over horizontal partitioned data

Partners/Cooperations

Hochschule Zittau/Görlitz
UNIVERSITY OF APPLIED SCIENCES

- Zentrum für Wissens- und Technologietransfer
- Qualitätspakt Lehre - Vielfalt als Stärke

rapidminer

German Research Center for Artificial Intelligence

CARL VON OSSIETZKY universität OLDENBURG

Fraunhofer IAIS

arc SOLUTIONS

Bundesministerium für Bildung und Forschung