PURDUE
UNIVERSITY

CS490D:
Introduction to Data Mining
Prof. Chris Clifton

April 14, 2004

Fraud and Misuse Detection

What is Fraud Detection?

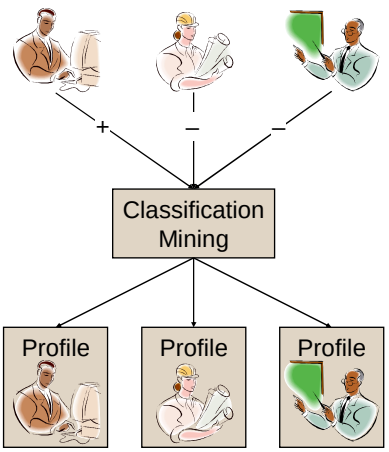
- Identify wrongful actions
  - Is right and wrong universal?
  - If so, why not just prevent wrong actions
- Identify actions by the wrong people
- Identify *suspect* actions
  - Legal
  - But probably not right

In Data Mining terms…

- Classification?
  - Classify into fraudulent and non-fraudulent behavior
  - What do we need to do this?
- Outlier Detection
  - Assume non-fraudulent behavior is normal
  - Find the exceptions
- Problems?

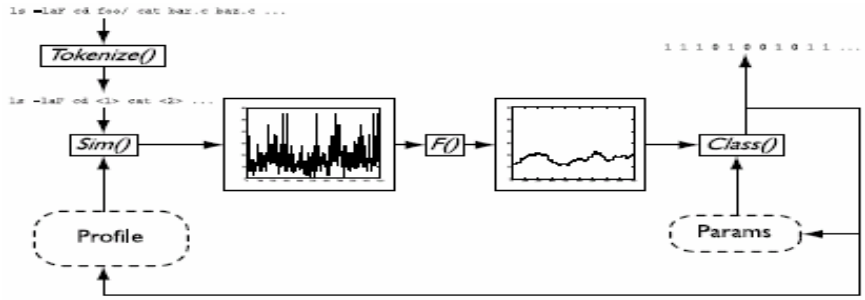
Solution: Differential Profiling

- Determine individual behavior
  - What is normal for the individual
  - What separates one individual from another
- Gives profile of individual behavior
- How do we do this?



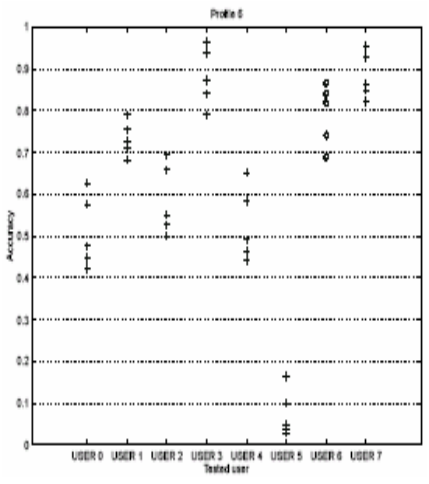
Has this been done?
Intrusion Detection *(Lane&Brodley)*

- Profiled computer users based on command sequences
  - Command
  - Some (but not all) argument information
  - Sequence information

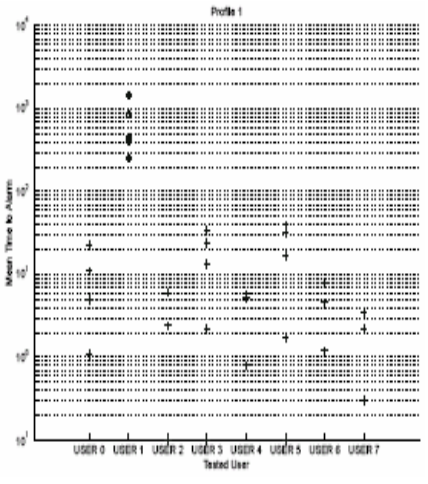


Results

Accuracy



Time to Alarm



Scaling Issues

- What happens with millions of users?
  - Credit card
  - Cell phone
- What about new users?
- Ideas?

Multi-user profiles

- Cluster users
- Develop profiles for clusters
  - E.g., differential profiling
- Old customers: Do they match profile for their cluster?
  - Allows wider range of acceptable behavior
- New customer: Do they match *any* profile?