

Detecting Undesirable Insider Behavior

Joseph A. Calandrino*
Princeton University

Steven J. McKinney*
North Carolina State University

Frederick T. Sheldon
Oak Ridge National Laboratory

May 15, 2007

*This research was performed during an internship at ORNL

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

An Example

"[A bank] employee gained control over data after performing several preparatory actions, such as eliminating some monitoring ... or convincing ... personnel to take deliberately corrupted data from his ... computer instead [of] from the official Reuters terminal."

-From Anderson et al., 2004

- Enters false data
- Receives promotions, bonuses, control
- Financial impact: **\$500 million**

Undesirable Insider Behavior

- **Results in ~29% of attacks against organizations**
(US Secret Service et al., 2004)
- **Can devastate an organization**
- **Fundamentally differs previously addressed threats**
- **Comes from trusted individuals**
- **Is a fuzzy threat**
- **Is tedious to identify**

Roadmap



- **Existing work**
- **Our contributions**
- **Mining approach**
- **Evaluation**
- **Discussion**
- **Conclusion and Future Work**

Existing Work

- Insider threat characterization
- Intrusion detection
- Machine learning / data mining
 - Statistical deviation (Anderson et al., 1995)
 - Real-time IDS (Lee et al., 1998; Lee et al., 2001)
 - MINDS (Ertöz et al., 2004; Chandola et al., 2006)

Our Contributions

- Designed a system to:
 - Monitor insider system and network activity
 - Perform rule-based (or other static) analysis
 - **Mine compiled behavior for anomalies**
- Implemented the system

Data Mining Component Role

- **Periodically extracts aggregate data**
- **Analyzes data to isolate points of interest**
- **Identifies novel threats**
- **Generates new rules (future work)**

Characteristic Derivation

- **Daily data analysis**
- **Per-user data**
- **System-level events alone... for now**
- **Seven characteristics (file accesses, hosts, logins)**
- **Normalization using historical SD**

Data Analysis

- **Similar to MINDS (implementation differs):**
 - Map data to vector space
 - Compute local outlier factor (Breunig et al., 2000)
 - “Neighborhood” outlier metric
 - Euclidean distance – not mandatory
 - Derive additional hints for administrators

Local Outlier Factor

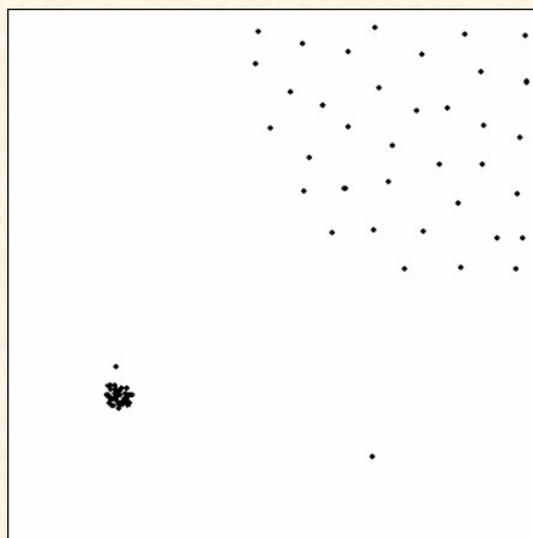


Image Reproduced from Breunig et al. (2000)

Local Outlier Factor

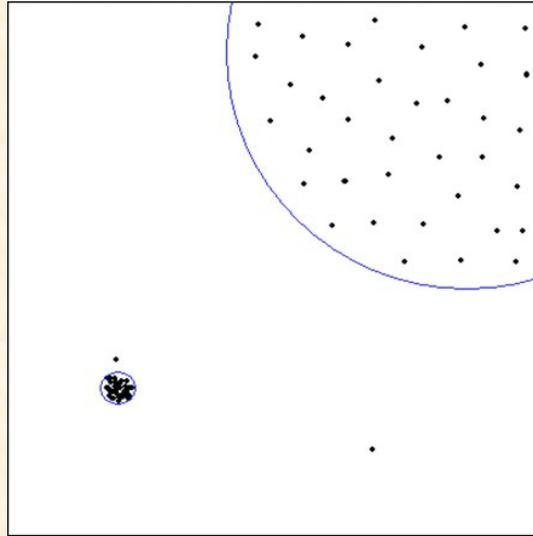


Image Reproduced from Breunig et al. (2000)

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY



LOF Scores

- **Could report to administrator**
- **Supervisors may be preferable**
 - Supervisors are most familiar with day-to-day tasks
 - Supervisors may be less technically literate
- **Consider impact of characteristics on outlier factor**



OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

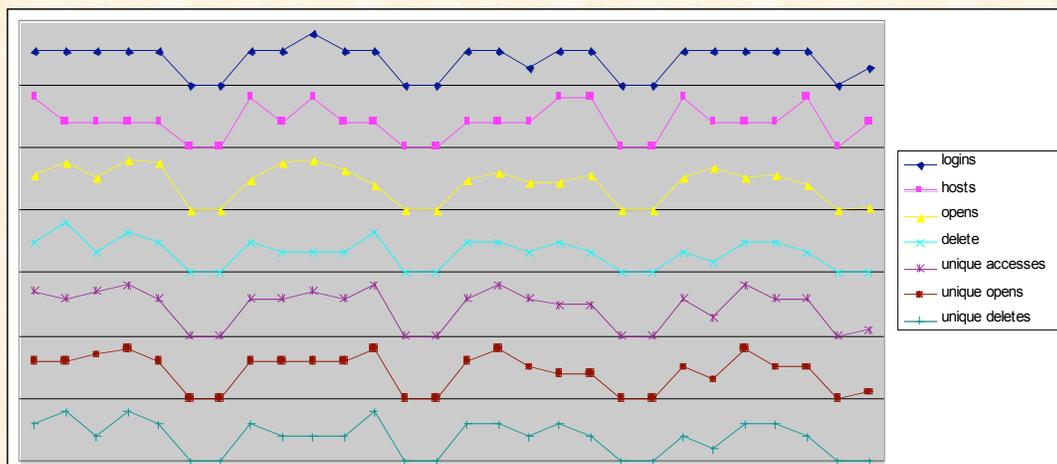


Evaluation

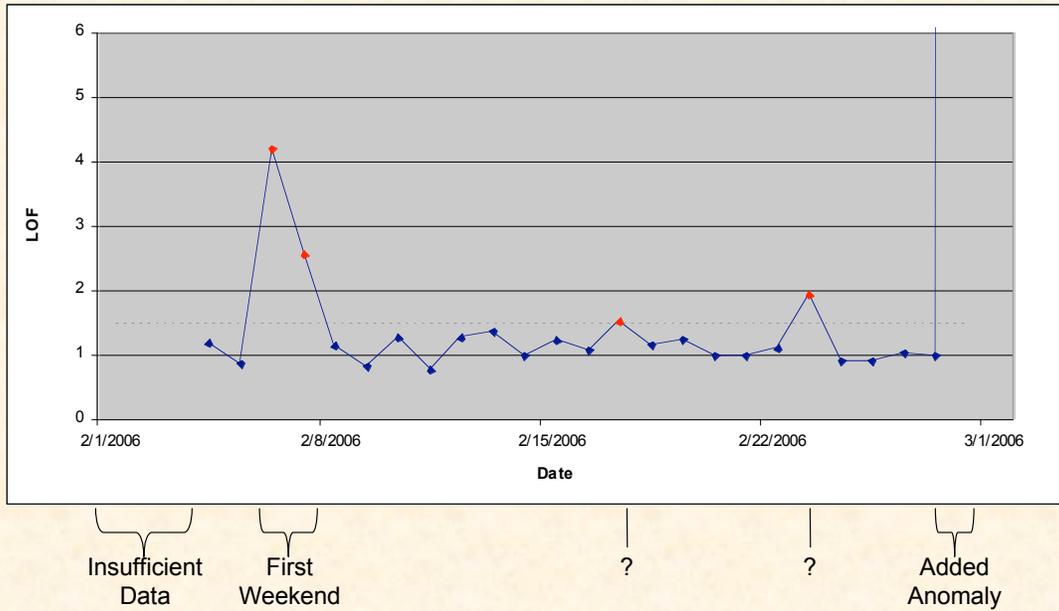
- **Generated 28 Days of artificial data**
 - Presume patterns
 - No activity on weekends until final weekend day
 - Activity comes from distribution on weekdays
 - Is real behavior like this?

Test Data

- **Graphically (scaled):**



Results

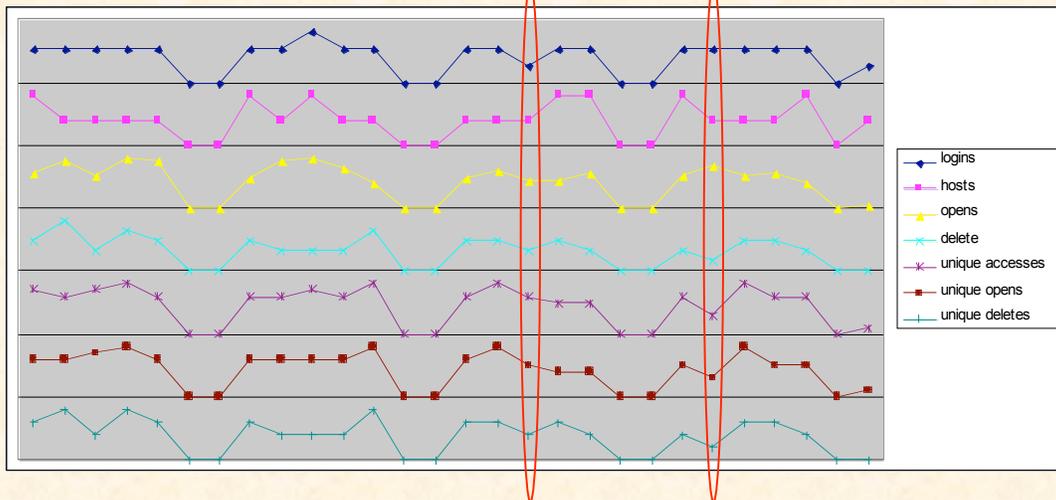


OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY



Results

• What's wrong?

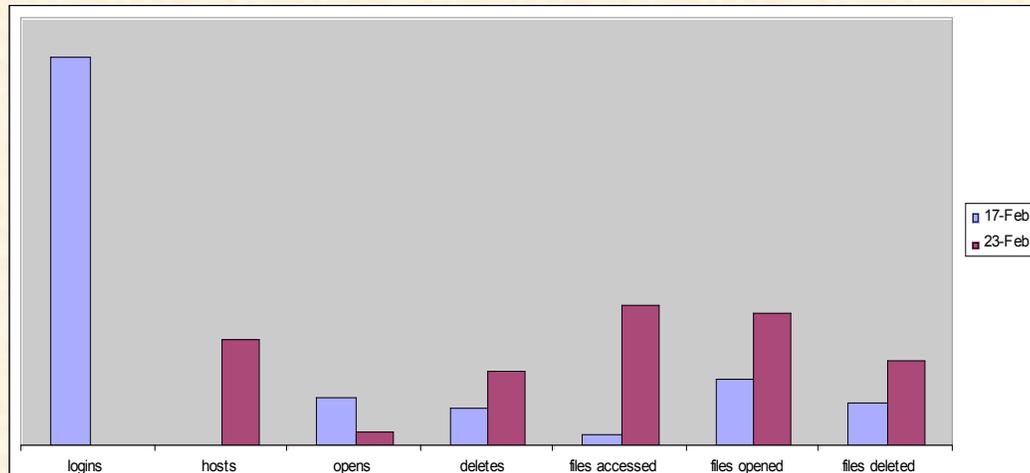


OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY



Results

- **Why guess? Look at hints for 2/17, 2/23...**



Discussion

- **Caught added anomaly**
- **What about the four others?**
 - All were “anomalous”
 - Hints allowed rapid analysis
 - Depend on parameters, administrator focus

Conclusions

- **Insider threat – important problem**
- **Data mining – helpful technique**
- **New tool – promising results**
- **TODO list – long**

Future Work

- **Additional aspects and characteristics**
- **Issues: drift, deja vu**
- **Better test data**
- **Rule extraction**

Thank You

- **Questions?**

Calandrino and McKinney performed this research while under appointment to the Department of Homeland Security (DHS) Scholarship and Fellowship Program, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-AC05-06OR23100. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of DHS, DOE, or ORAU/ORISE.

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

