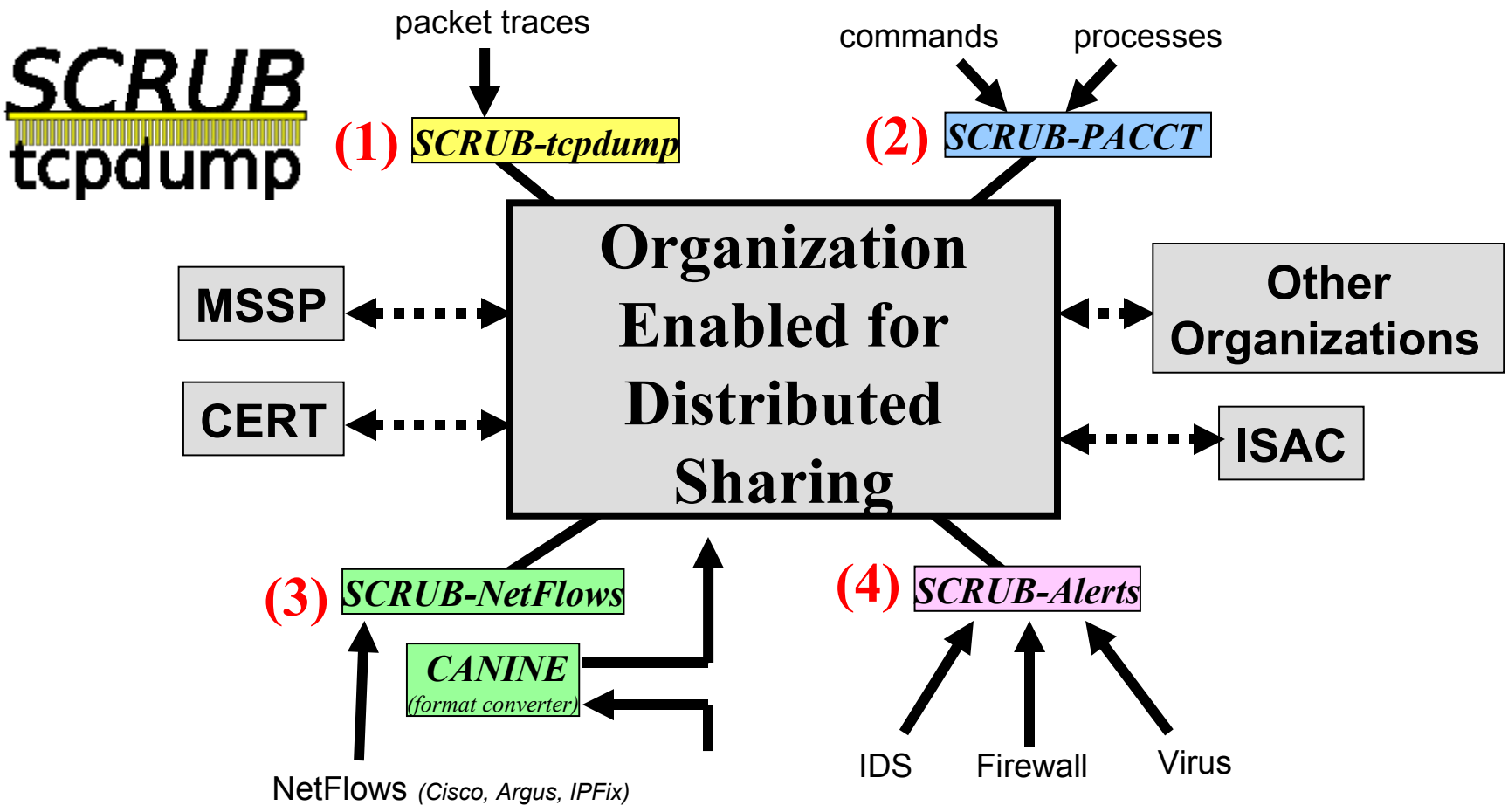
Safely Sharing Data Between CSIRTs:
The SCRUB* Security Anonymization Tool Infrastructure

William Yurcik*
<byurcik@gmail.com>
Clay Woolam, Greg Hellings, Latifur Khan, Bhavani Thuraisingham

University of Texas at Dallas

20th Annual FIRST Conference
Vancouver Canada
June 2008

The SCRUB* Architecture



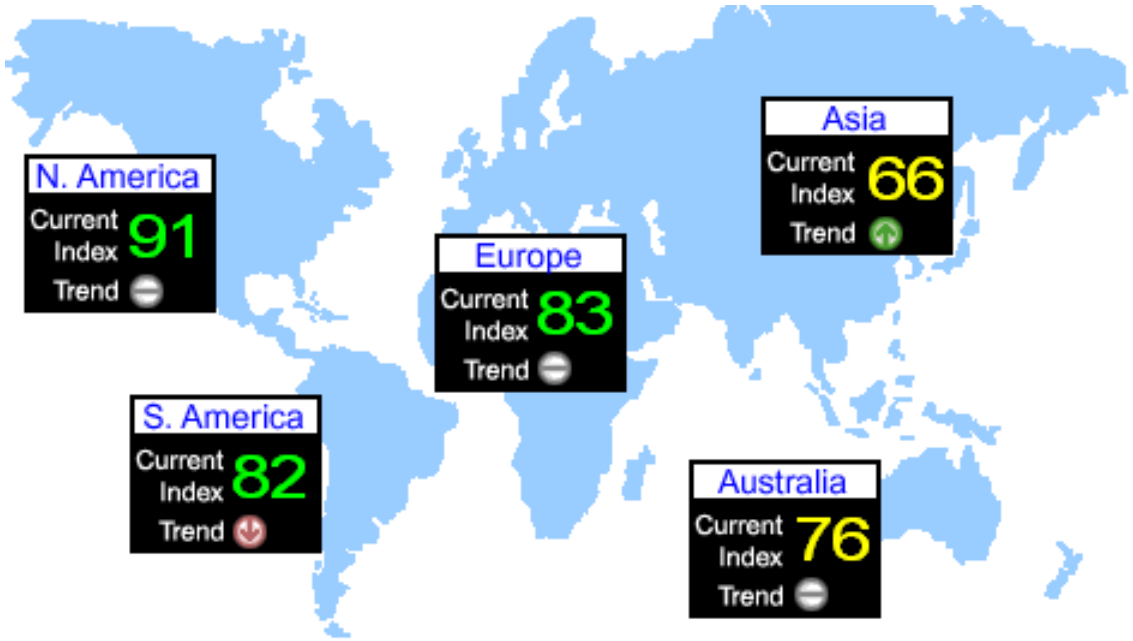
*SCRUB** Motivation
Why Should We Share Security Data?

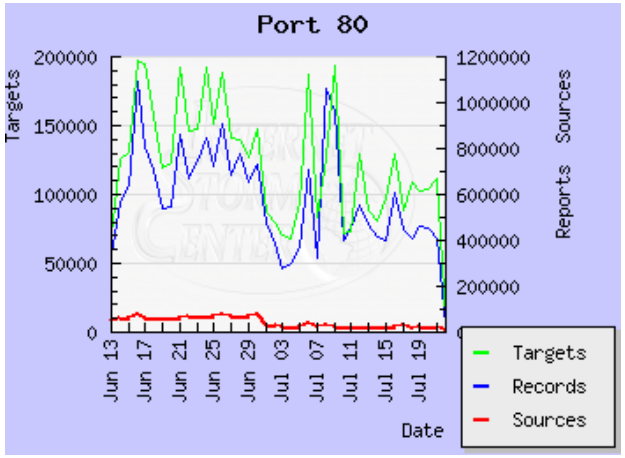
- **Event correlation across administrative domains is needed based on shared data**
  - **We cannot continue to stop attacks at organizational borders, we need to cooperate with law enforcement and each other.**
  - **Chasing attackers away to other organizations does not improve security**
- **Need to share security data between organizations in order to**
  - **Detect attacks**
  - **Blacklist attackers and attacker techniques**
  - **Distinguishing normal versus suspicious network traffic patterns**

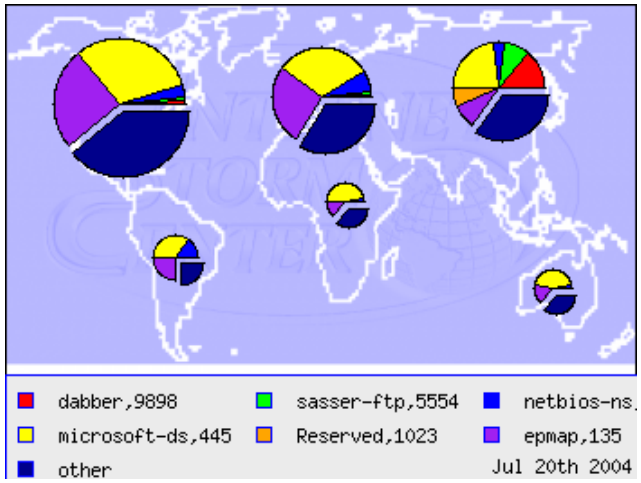
SANS

INTERNET TRAFFIC REPORT

Last update (MST):
7/21/2004 20:20
Global Index 85
Trend

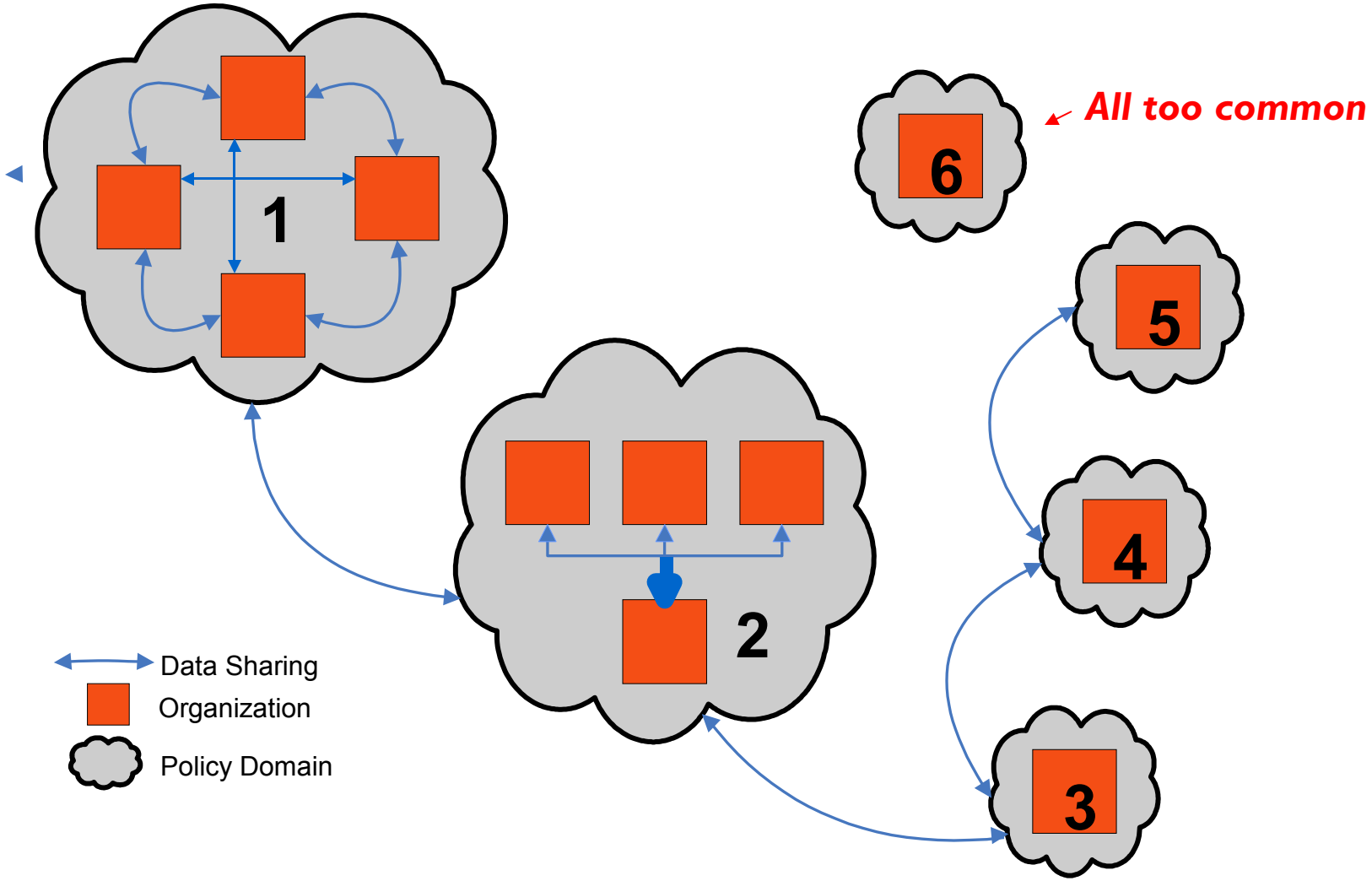






UTD

State-of-the-Art in Security Data Sharing



For Safe Data Sharing: Two Types of Data To Protect

- **Private Data**
  - User-identifiable information
    - user content (Email messages, URLs)
    - user behavior (access patterns, application usage)
  - Machine/Interface addresses
    - IP and MAC addresses
- **Sensitive Data**
  - System configurations (services, topology, routing)
  - Traffic patterns (connections, mix, volume)
  - Security defenses (firewalls, IDS, routers)
  - Attack impacts

*SCRUB** TOOL 1:

- Anonymizes packet traces
  - packet traces can contain the most private/sensitive data
  - packet traces are the authoritative raw security source
- Leverage a popular existing tool – *tcpdump*
- Anonymizes any/all packet fields (12)
- Each field has multiple anonymization options
  - none/low/medium/high levels of protection for protecting the same data field

*SCRUB** TOOL 2: *SCRUB-PACCT*

- Anonymizes proccess accounting logs
  - process accounting records contain user IDs and user command behavior
  - process accounting records contain precise timing information for event correlation between systems
- Anonymizes any/all process accounting fields (16)
- Each field has multiple anonymization options
  - none/low/medium/high levels of protection for protecting the same data field

*SCRUB** TOOL 3: *SCRUB-NetFlows*

- Anonymizes NetFlow logs
  - NetFlows logs efficiently aggregate packet traffic by connections
  - Most commonly shared security data
- Anonymizes any/all NetFlow fields (5)
- Each field has multiple anonymization options
    - none/low/medium/high levels of protection for protecting the same data field

*SCRUB** Fields of Interest Between Data Sources

1. **Transport Protocol Number**
   - data sources: packet, NetFlows, alerts
2. **IP Address**
   - data sources: packet, NetFlows, alerts
3. **Ports**
   - data sources: packet, NetFlows, alerts
4. **Payload**
   - data sources: packet, alerts
5. **Timestamp**
   - data sources: packet, process accounting, NetFlows, alerts

Multi-Level Anonymization Options

- Black Marker (filtering/deletion)
- Pure Randomization (replacement)
- Keyed Randomization (replacement)
- Annihilation/Truncation (time, accuracy reduction)
- Prefix-Preserving Pseudonymization (IP address)
- Grouping (accuracy reduction)
  - Bilateral Classification
- Enumeration (time, adding noise)
- Time Shift (time, adding noise)

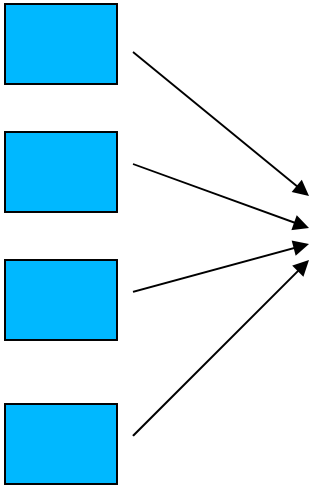
A Problem with Anonymization for Sharing: Privacy vs. Analysis Tradeoffs

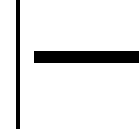
while anonymization protects against information leakage it also destroys data needed for security analysis

- Zero-Sum? (more privacy <> less analysis & vice versa)
- to date, no quantitative measurements of how useful anonymized data is for security analysis

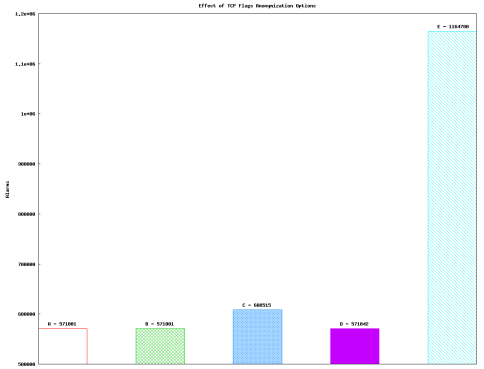
Empirically Measuring Anonymization Privacy/Analysis Tradeoffs

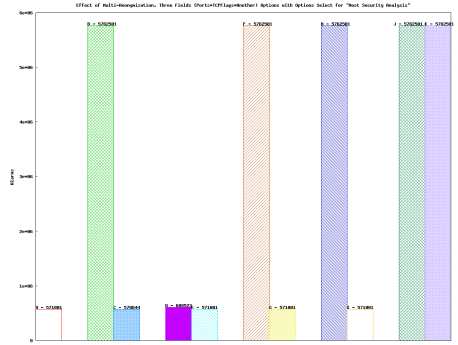
- Series of experiments to test effects of different anonymizations options
- Use snort IDS alarms as a metric for security analysis





alarms





Data Set

Summary

- There is a critical need for security data sharing between organizations
- Anonymization can provide safe data sharing
  - Multi-Field: prevent information leakage
  - Multi-Level: no one-size-fits-all anonymization solution
- A practical data sharing infrastructure is needed which supports multiple data sources
  - *SCRUB** tool suite for packet traces, process accounting, NetFlows, alerts
- Privacy/analysis anonymization tradeoffs can be characterized
  - Zero-Sum tradeoff? (not always, more complex than this)
  - Multi-Level anonymization options can/should be tailored to requirements of sharing parties to optimize tradeoffs
  - More tradeoff measurements are in progress

References

Background on Using Anonymization to Safely Share Security Data

A.J. Slagell and W. Yurcik, "Sharing Computer Network Logs for Security and Privacy: A Motivation for New Methodologies of Anonymization," *1st IEEE Intl. Workshop on the Value of Security through Collab. (SECOVAL),* 2005.

A.J. Slagell and W. Yurcik, "Sharing Network Logs for Security and Privacy: A Motivation for New Methodologies of Anonymization," *ACM Computing Research Repository (CoRR) Technical Report cs.CR/0409005*, September 2004.

X. Yin, K. Lakkaraju, Y. Li, and W. Yurcik, "Selecting Log Data Sources to Correlate Attack Traces For Computer Network Security: Preliminary Results," *11th Intl. Conf. on Telecomunications,* 2003.

W. Yurcik, James Barlow, Yuanyuan Zhou, Hrishikesh Raje, Yifan Li, Xiaoxin Yin, Mike Haberman, Dora Cai, and Duane Searsmith, "Scalable Data Management Alternatives to Support Data Mining Heterogeneous Logs for Computer Network Security," *SIAM Workshop on Data Mining for Counter Terrorism and Security*, 2003.

J. Zhang, N. Borisov, and W. Yurcik, "Outsourcing Security Analysis with Anonymized Logs," *2nd IEEE Intl. Workshop on the Value of Security through Collab. (SECOVAL),* 2006.

J. Zhang, N. Borisov, W. Yurcik, A.J. Slagell, and Matthew Smith, "Future Internet Security Services Enabled by Sharing of Anonymized Logs," *Workshop on Security and Privacy in Future Business Services held in conjunction with International Conference on Emerging Trends in Information and Communication Security (ETRICS),* University of Freiburg Germany, 2006.

SCRUB* Tool (1) SCRUB-tcpdump < http://scrub-tcpdump.sourceforge.net/>

W. Yurcik, C. Woolam, G. Hellings, L. Khan, and B. Thuraisingham, *"SCRUB-tcpdump*: A Multi-Level Packet Anonymizer Demonstrating Privacy/Analysis Tradeoffs," *3rd IEEE Intl. Workshop on the Value of Security through Collab. (SECOVAL),* 2007.

SCRUB* Tool (2) SCRUB-PACCT <http://security.ncsa.uiuc.edu/distribution/Scrub-PADownLoad.html>

C. Ermopoulos and W. Yurcik, "NVision-PA: A Process Accounting Analysis Tool with a Security Focus on Masquerade Detection in HPC Clusters," *IEEE Intl. Conf. on Cluster Computing (Cluster),* 2006.

K. Luo, Y. Li, C. Ermopoulos, W. Yurcik, and A.J. Slagell, "*SCRUB-PA*: A Multi-Level Multi-Dimensional Anonymization Tool for Process Accounting," *ACM Computing Research Repository (CoRR) Technical Report cs.CR/0601079,* January 2006.

W. Yurcik and C. Liu, "A First Step Toward Detecting SSH Identity Theft in HPC Cluster Environments, Discriminating Masqueraders Based on Command Behavior," *1st Intl. Workshop on Cluster Security (Cluster-Sec)* in conjunction with *5th IEEE Intl. Symposium on Cluster Computing and the Grid (CCGrid),* 2005.

SCRUB* Tool (3) SCRUB-NetFlows < http://scrub-netflows.sourceforge.net/> >

Y. Li, A.J. Slagell, K. Luo, and W. Yurcik, "CANINE: A Combined Converter and Anonymizer Tool for Processing NetFlows for Security," *13th Intl. Conf. on Telecomunications Systems*, 2005.

K. Luo, Y. Li, A.J. Slagell, and W. Yurcik, "*CANINE:* A NetFlows Converter/Anonymizer Tool for Format Interoperability and Secure Sharing," *FLOCON – Network Analysis Workshop (Network Flow Analysis for Security Situational Awareness),* 2005.

A.J. Slagell, J. Wang, and W. Yurcik, "Network Anonymization: The Application of *Crypto-PAn* to Cisco NetFlows," IEEE/*NSF/AFRL Workshop on Secure Knowledge Management (SKM),* 2004.

UTD