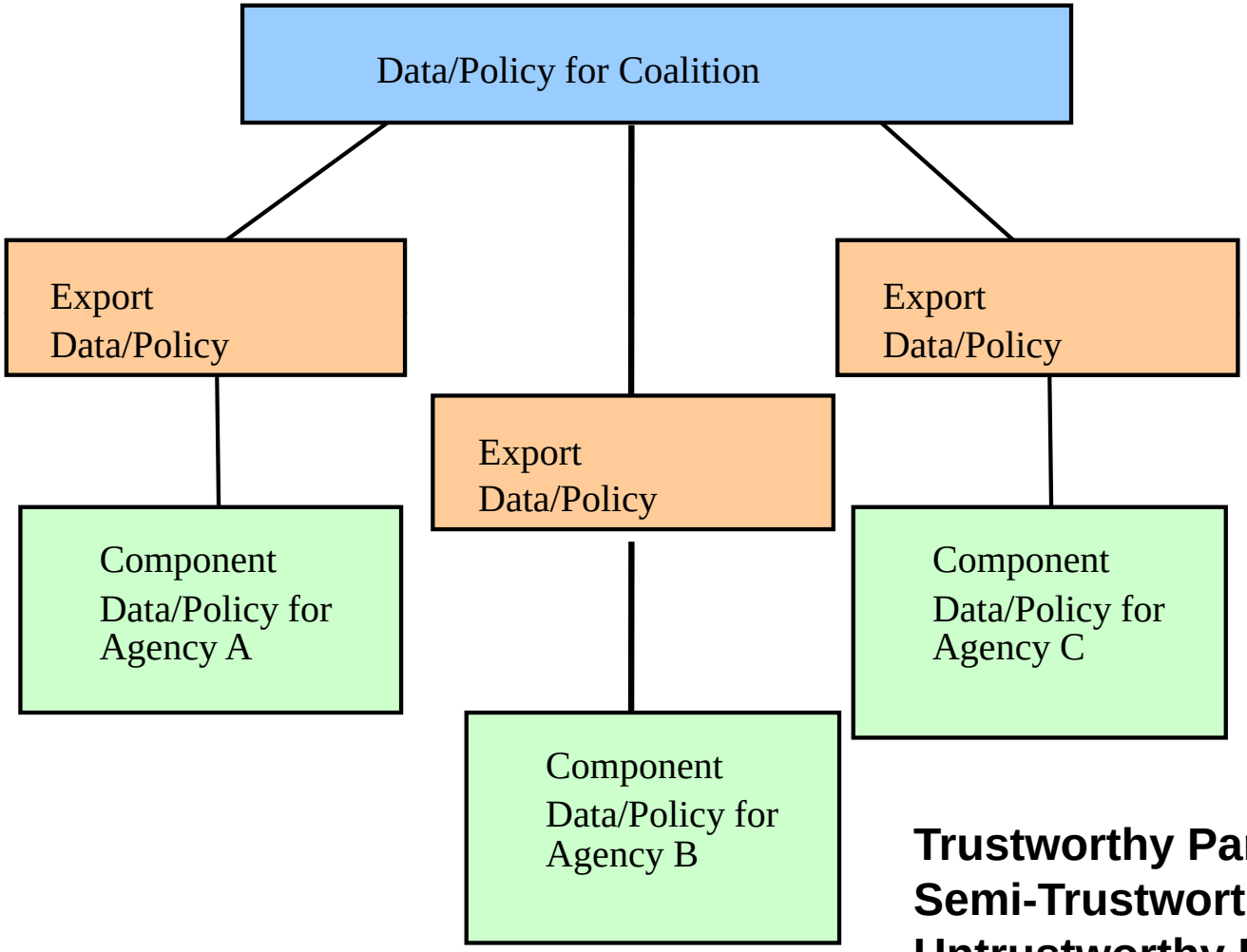

# Information Operation across Infospheres

Prof. Bhavani Thuraisingham
and Prof. Latifur Khan
The University of Texas at Dallas

Prof. Ravi Sandhu
George Mason University
(UTSA as of 6/4/2007)

June 2007

# Architecture




**Trustworthy Partners**
**Semi-Trustworthy Partners**
**Untrustworthy Partners**

# Our Approach

- Integrate the Medicaid claims data and mine the data; next enforce policies and determine how much information has been lost (Trustworthy partners); Prototype system
- Apply game theory and probing to extract information from semi-trustworthy partners
- Trust for Peer to Peer Networks
- Conduct information operations (defensive and offensive) and determine the actions of an untrustworthy partner.
- Examine RBAC and UCON for coalitions (George Mason University)
- Funding: AFOSR 300K; Texas Enterprise Funds 150K for students; 60K+ for faculty summer support; 45K+ for postdoc

# Accomplishments to date

- **FY06: Presented at 2006 AFOSR Meeting**

  - Investigated the amount of information loss by enforcing policies – Considered release factor

  - Preliminary research on RBAC/UCON; Game theory approach, Defensive operations

- **FY07: Presented at 2007 AFOSR Meeting**

  - Completion of Prototype

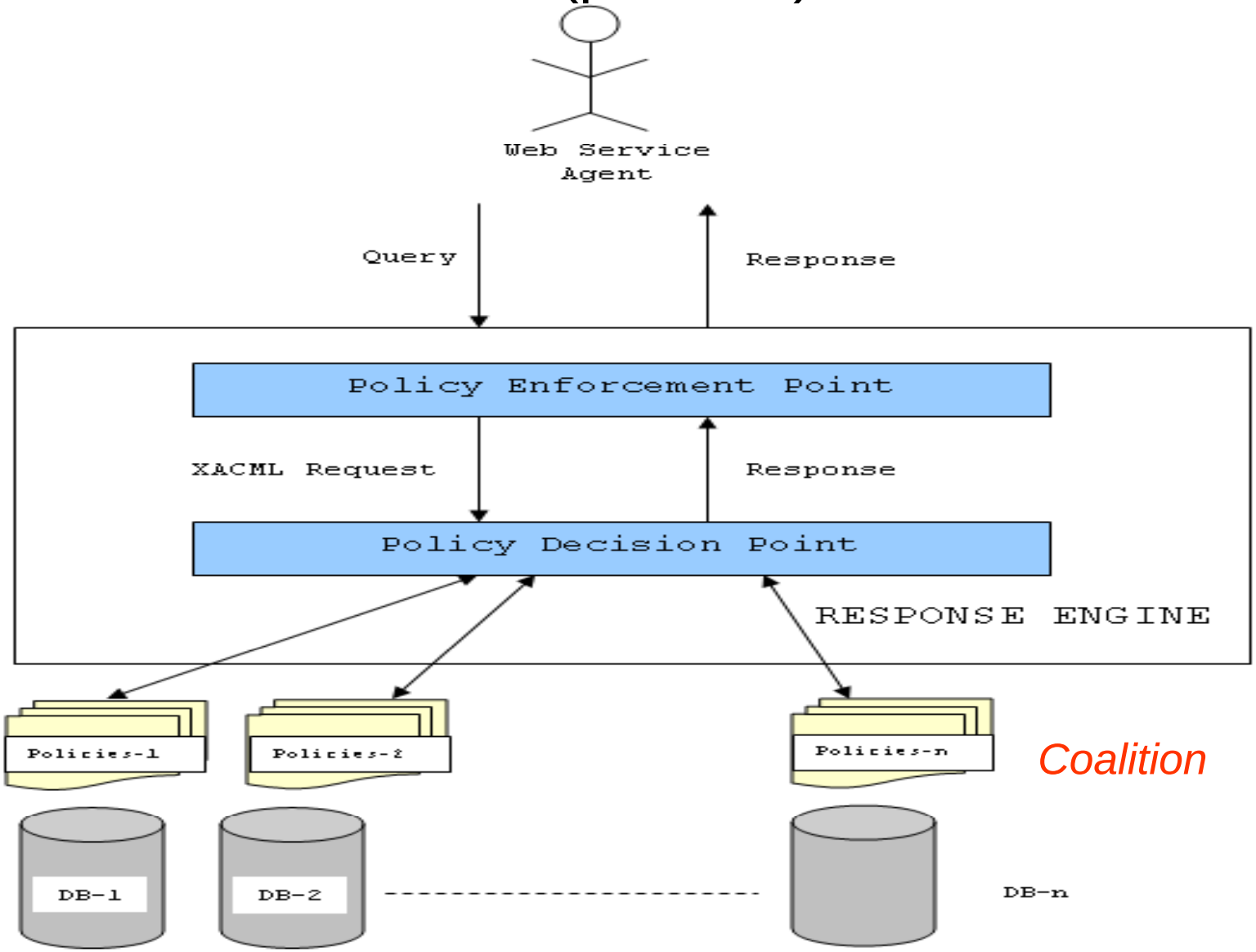
  - Solutions using game theory, Penny for P2P Trust, Data mining for Code blocker and Botnet, RBAC/UCON

- **FY08 Plans: To be presented 2008 AFOSR Meeting**

  - Offensive Operations, Near operational prototype integrated system

# Policy Enforcement Prototype

Dr. Mamoun Awad (postdoc)  and students



# Architectural Elements of the Prototype

•***Policy Enforcement Point (PEP)***:
•Enforces policies on requests sent by the Web Service.
•Translates this request into an XACML request; sends it to the PDP.

•***Policy Decision Point (PDP):***
•Makes decisions regarding the request made by the web service.
•Conveys the XACML request to the PEP.

***Policy Files:***

- Policy Files are written in XACML policy language. Policy Files specify rules for "Targets". Each target is composed of 3 components: Subject, Resource and Action; each target is identified uniquely by its components taken together. The XACML request generated by the PEP contains the target. The PDP's decision making capability lies in matching the target in the request file with the target in the policy file. These policy files are supplied by the owner of the databases (Entities in the coalition).

***Databases:***
- The entities participating in the coalition provide access to their databases.

# Semi-Trustworthy Partners Enforcing Honesty (Prof. Murat Kantarcioglu, Ryan Layfield)

- Everyone has a choice:
  - Tell the truth
  - Lie
- Unless we can afford to have a neutral 3rd party that everyone can agree on, we need some way of enforcing 'good' behavior
- However, there is a third option: *refuse to participate*
  - Usually not researched
  - Drastic measure that only makes sense if we can influence behavior
- Our modeling suggests that, with proper use of refusal, we can ultimately enforce helpful behavior without a managing agent

# Evolutionary Strategy

- Every 200 rounds, we create a new generation of agents, using the most successful strategies available
- The fitness *f()* of a given agent is a function of how well they have performed during interaction with other agents
  - More successful agents have a higher probability of being a part of the next generation
- Our mathematical models suggest that, assuming we punish by cutting off communication, the equilibrium is to always tell the truth
- Therefore, using an evolutionary environment, we have placed our particular rationality amongst a heterogeneous pool of competing ideologies
  - **Tit-For-Tat**: A famous algorithm that simply mirrors the last move an opponent made
  - **Random**: An agent that selects it's strategy with a 50/50 chance
  - **Casual Liar**: lies with a 10% probability
  - **Subtle Liar**: chooses to lie when it perceives the piece being traded is of significant value
  - **Truthful-punishment:** Says the truth; punishes lies by cutting off communication
- With equal parts given to each agent, which one will emerge victorious?
- - Truthful-punishment performs the best
- Next steps: Assume that the communication is not secure; cannot verify every piece of data shared

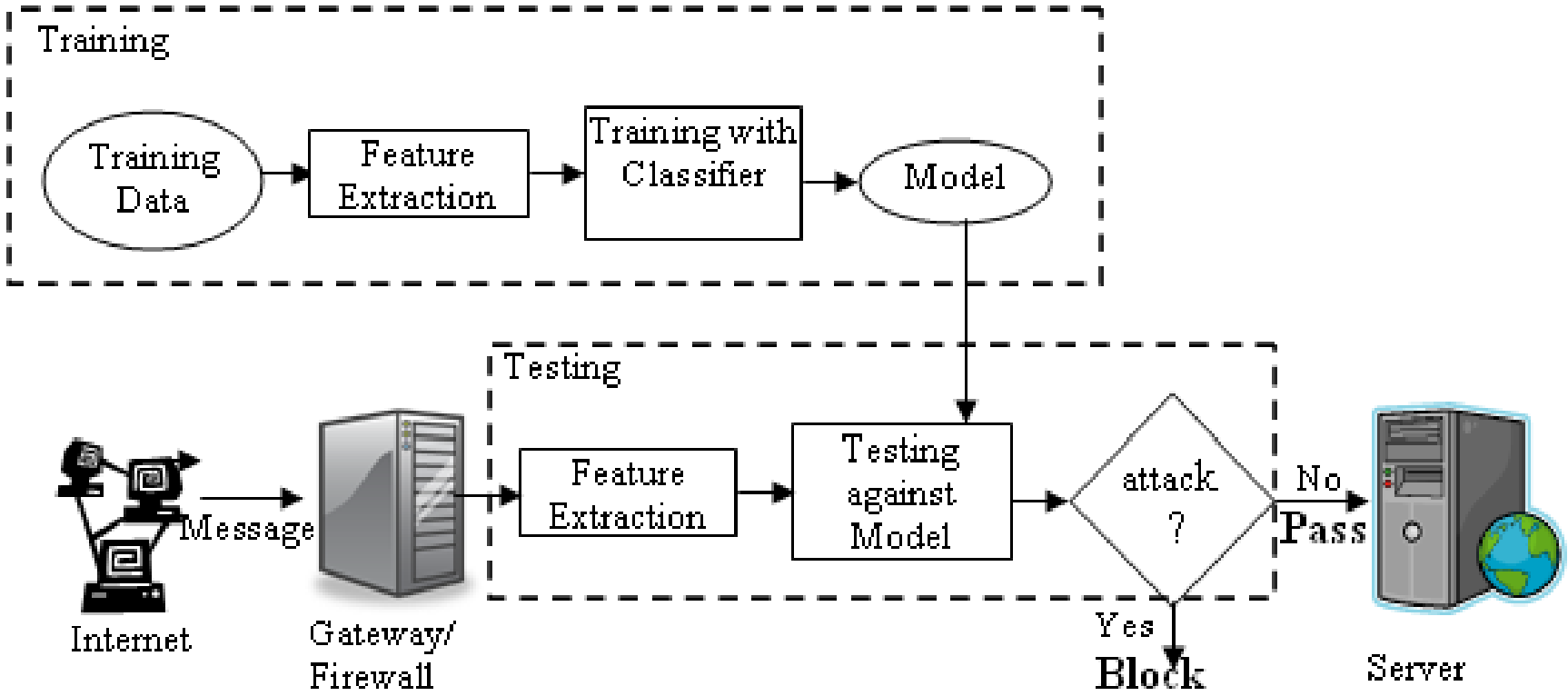
# Penny: Trust in P2P Network

## Prof. Kevin Hamlen and Nathalie Tsublinik

- **A P2P Network** that addresses the following types of attacks:
  - Spread of corrupt or incorrect data
  - Attaching incorrect labels to data
  - Discovering which peers own particular data
  - Generating a list of all peers who own particular data
- P2P Network that supports shared data labeling of:
  - Confidentiality
  - Integrity
- Peers can share data without revealing which data object they own
- Security labels are global but do not require a centralized server
- P2P Network uses reputation-based trust management system
  - Store/retrieve labels
  - Despite malicious peer existence
- Maintain efficiency of network operations
- O(log N)

# Untrustworthy Partners CodeBlocker (Our approach)

## Prof. Latifur Khan and Mehdy Masud



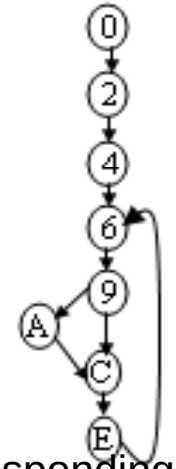

•Based on the Observation: ***Attack messages usually contain code while normal messages contain data;***
Check whether message contains code
Problem to solve: Distinguishing code from data

# Feature extraction

- Features are extracted using
  - – *N*-gram analysis
  - – Control flow analysis
- *N-gram analysis*

| address | instruction |
|---|---|
| 00 | xor eax, eax |
| 02 | mov ebx, eax |
| 04 | mov ecx, ebx +3F |
| 06 | cmp ebx, [ds: eax + A1F] |
| 09 | jbe short 00C |
| 0A | xor [ds:esi], ebx |
| 0C | add ebx, ecx |
| 0E | loopd short 006 |

Assembly program




Corresponding IFG

What is an *n*-gram?
-Sequence of *n* instructions

Traditional approach:
-Flow of control is ignored

**2-grams are:**

**02, 24, 46, 69, 9A, AC, CE**

# Experiments and Results

- Training data
  - Real instances of attack and normal messages; 10 different polymorphic attacks
  - 6,000 normal and 6,000 attack messages
- Testing data
  - 6,000 normal and 6,000 attack messages
  - All different from the training data
- Test bed
  - Windows XP; Intel P-IV 1.7GHz; 512MB

**Table : Comparing performances among different features and SigFree**

| Method | | CodeBlocker | | SigFree |
|---|---|---|---|---|
| Feature | CFBn | CFF | Combined | |
| Acc% | 96.4 | 88.6 | 96.4 | 58.3 |
| FP% | 1.2 | 3.5 | 1.2 | 0.2 |
| FN% | 5.9 | 19.3 | 5.9 | 82.7 |

# Botnet Detection Proposed Method

- Consists of three phases:
  - Phase I:
    Identifying Zombie machines
    [In-Progress]
  - Phase II:
    Identifying the Command & Control (C&C) traffic between zombie and botmaster [Future Work]
    Phase III:
    Preventing future infection/attack by blocking all C&C traffic into/out of the local network [Future Work]
- Experimental Setup
- The machines to be tested are connected to a gateway
- The gateway is connected to the Internet
- All traffic 'log's are collected at the Gateway

# Data Collection and Results

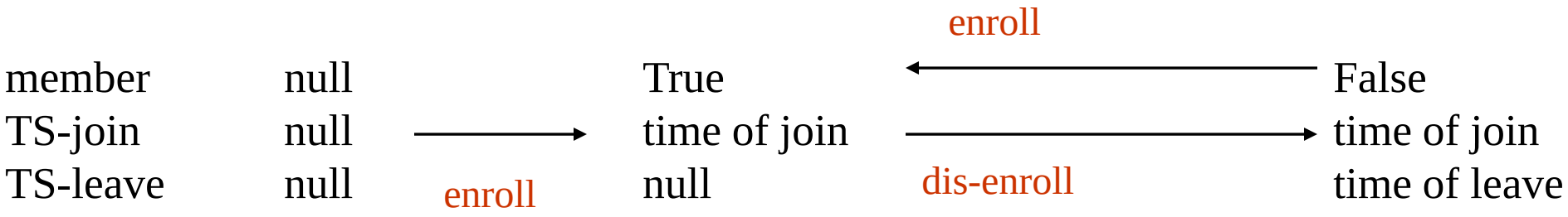
- We run 'clean' machines collecting traffic logs generated at the gateway
- We have collected about 130 malicious bots from Rajab Rajab, M. A., Zarfoss, J., Monrose, F. and Terzis, A. (JHU). "*A multifaceted approach to understanding the botnet phenomenon*." In Proceedings of the 6th ACM SIGCOMM on Internet Measurement Conference (IMC), 2006."
- We run each bot in a clean machine collecting traffic logs
- We analyze the logs and extract several features (data mining techniques)

| | **Accuracy (%)** | **False Positive (%)** | **False Negative (%)** |
|---|---|---|---|
| **SVM** | 91.6 | 11.7 | 4.5 |
| **NB** | 92.6 | 11.7 | 2.2 |
| **J48** | 98.9 | 0.0 | 2.2 |

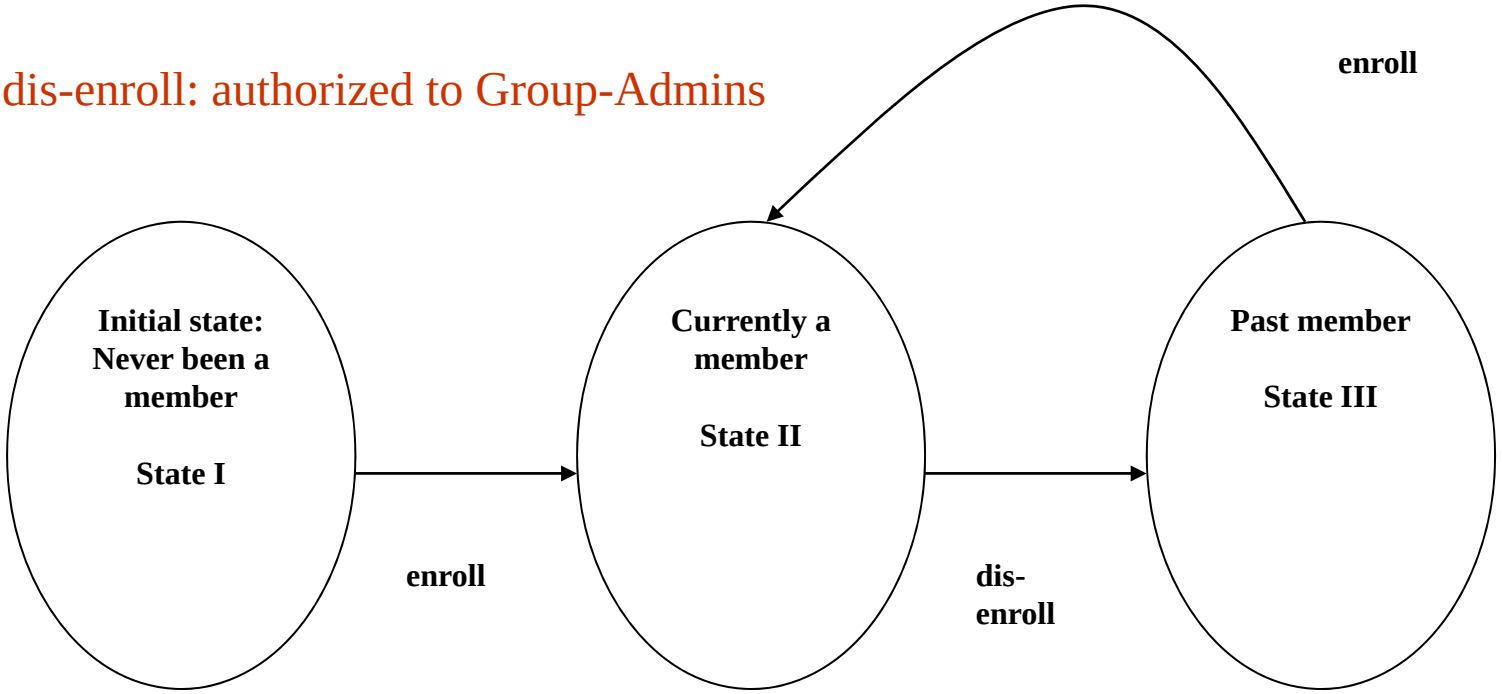
# UCON Policy Model (Prof. Ravi Sandu, X. Min)

- Operations that we need to model:
  - Document read by a member.
  - Adding/removing a member to/from the group
  - Adding/removing a document to/from the group
- Member attributes
  - Member: boolean
  - TS-join: join time
  - TS-leave: leave time
- Document attributes
  - D-Member: boolean
  - D-TS-join: join time
  - D-TS-leave: leave time

# Policy model: member enroll/dis-enroll

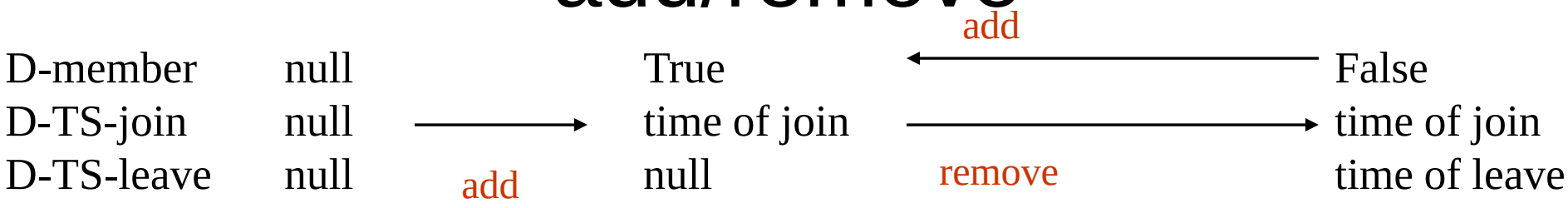



enroll, dis-enroll: authorized to Group-Admins

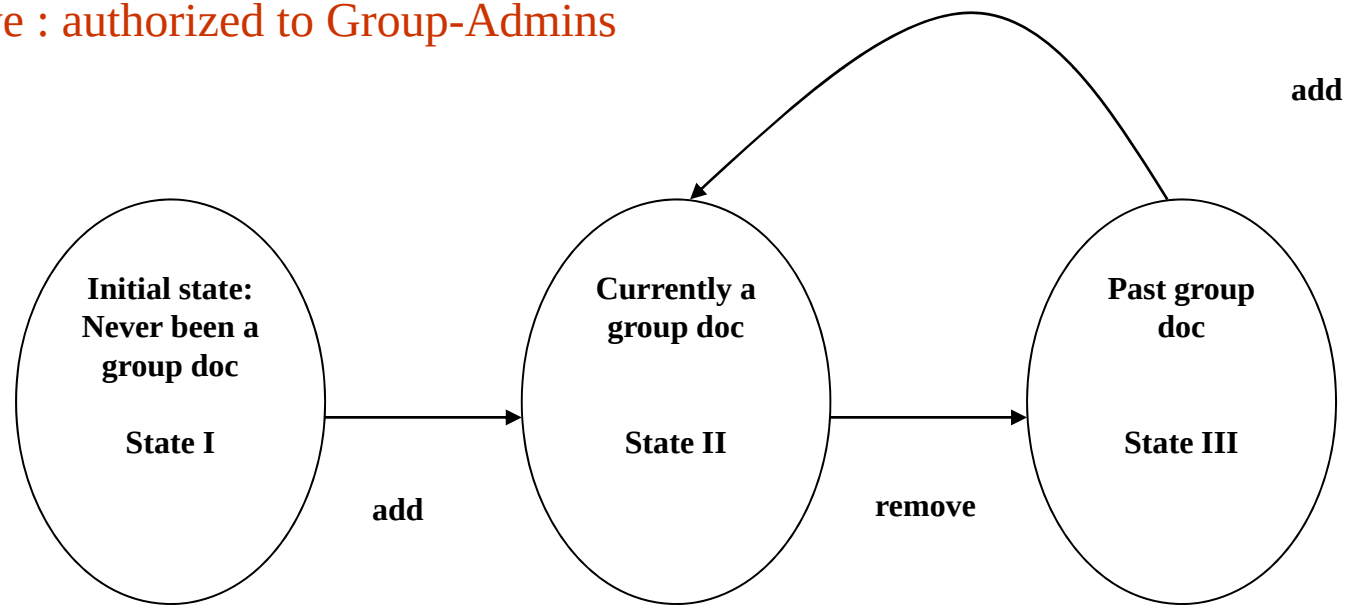



UCON elements:
Pre-Authorization, attribute predicates, attribute mutability

# Policy model: document add/remove




add, remove : authorized to Group-Admins




UCON elements:
Pre-Authorization, attribute predicates, attribute mutability

# Publications and Plans

- Some Recent Publications:
- Assured Information Sharing: Book Chapter on Intelligence and Security Informatics, Springer, 2007
- Simulation of Trust Management in a Coalition Environment, Proceedings IEEE FTDCS, March 2007
- Data Mining for Malicious Code Detection, Journal of Information Security and Privacy, Accepted 2007
- Enforcing Honesty in Assured Information Sharing within a Distributed System, Proceedings IFIP Data Security Conference, July 2007
- Confidentiality, Privacy and Trust Policy Management for Data Sharing, IEEE POLICY, Keynote address, June 2007 (Proceedings)
- Centralized Reputation in Decentralized P2P Networks, Submitted to ACSAC 2007
- Also units on assured information sharing on courses we teach at AFCEA
- Plans:
- Offensive Operations – find out what our untrustworthy partners are doing
- Integrated prototype – partners will change trust levels
- Scenario developments for prototype demonstration
- Technology Transfer to commercial products (data mining tools); operational programs (forming collaboration with Raytheon – Prime contact for AF DGCS – Distributed Common Ground System)