IMSCAN: A Framework for Instant Messenger User Analysis

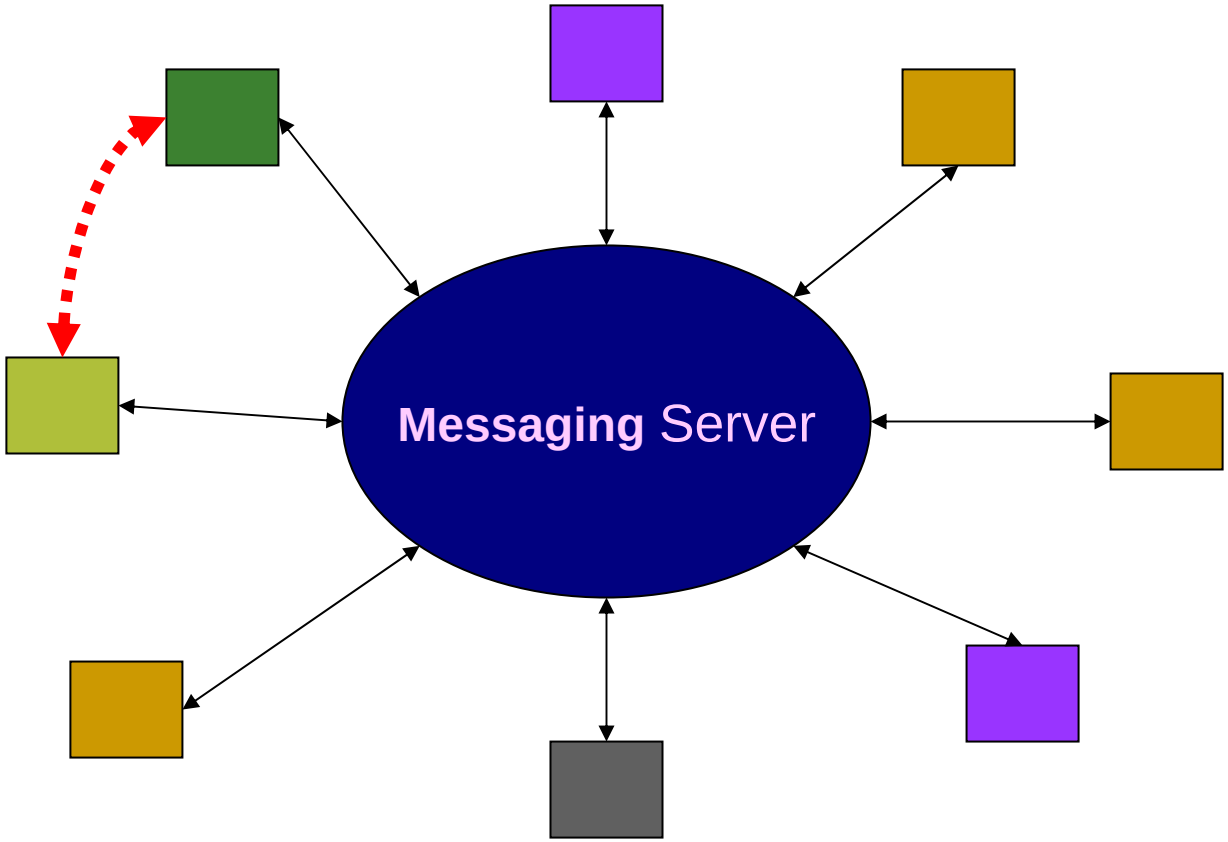
John Resig and
Ankur Teredesai
Assistant Professor,
Dept. of CS.
Associate Director,
Center for Discovery Informatics
Laboratory for Applied Computing
Rochester Institute of Technology (RIT)

DMRG Data Mining Research Group
Rochester Institute of Technology

Instant Messaging (IM): Resources for Data Miners

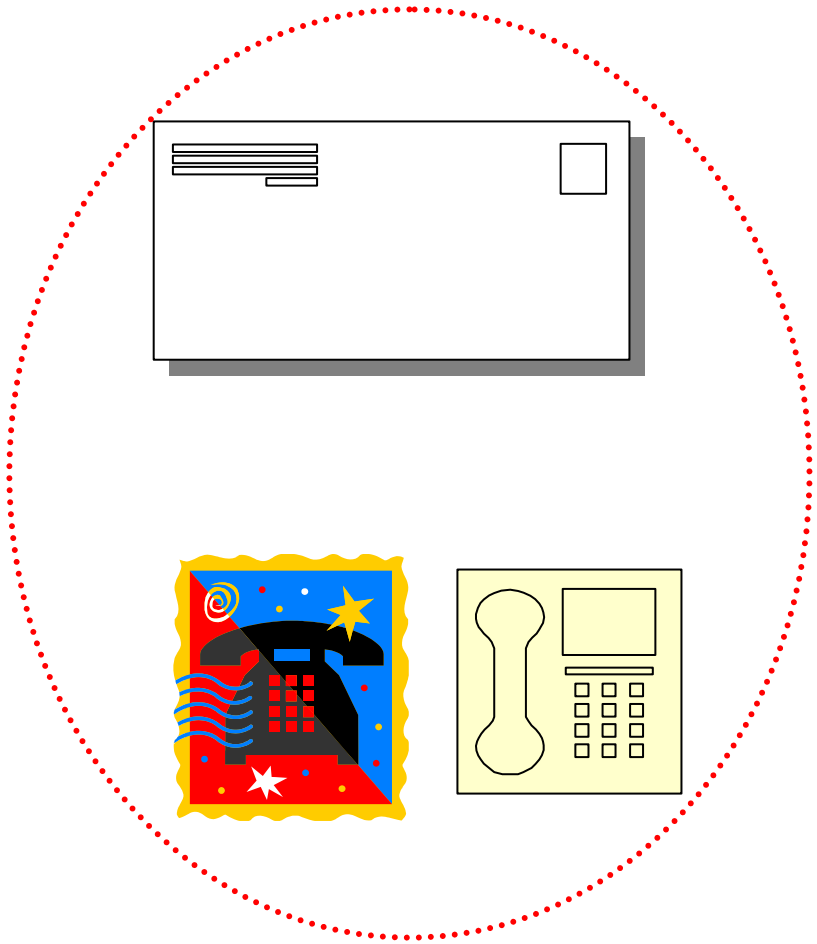
- User Status –
  - {Online, Offline, Idle, Away}
  - Duration
- Social Networks –
  - Buddy Lists
  - 3rd Party Information Mongering
- Textual Data –
  - Personal user profiles
  - Away messages
  - Public chat room discussions

IM Network Overview



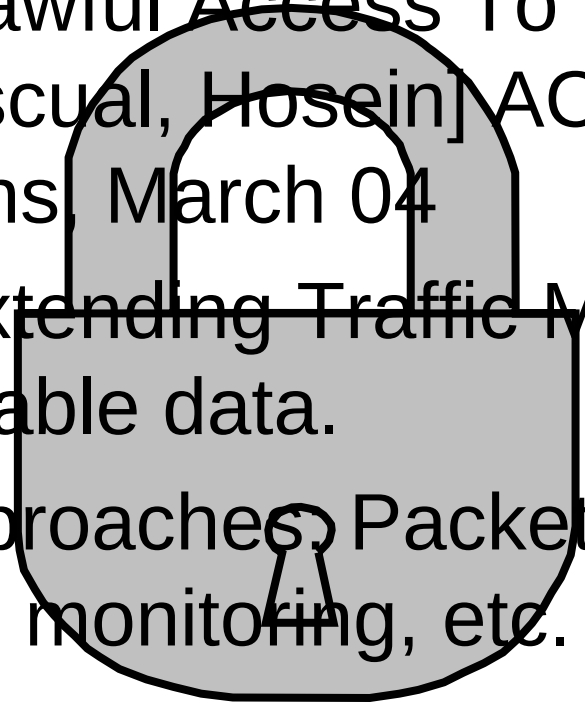
- Current Instant Messaging networks follow a strict client-server architecture

A Word on Traffic/Content Monitoring

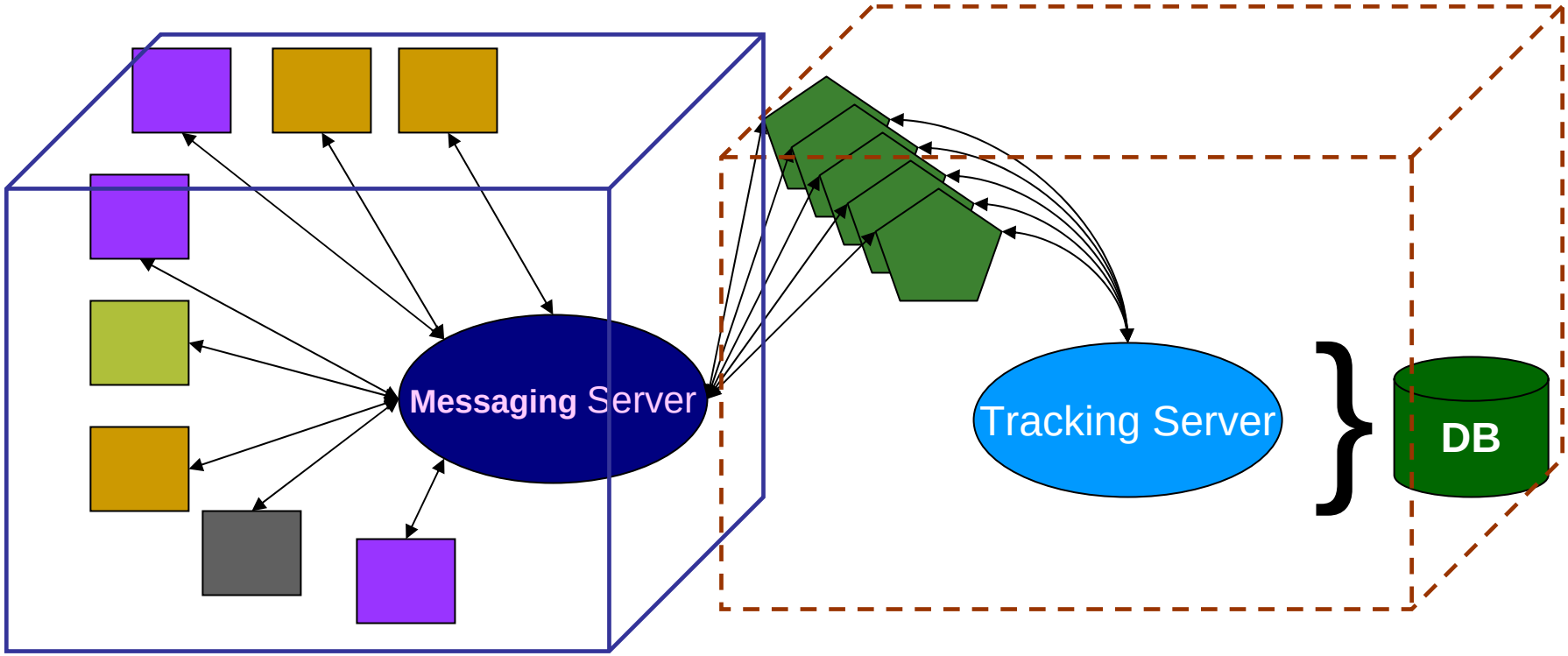


Policy

- Questioning Lawful Access To Traffic Data – [Escudero-Pascual, Hosein] ACM communications, March 04
- Patriot Act : Extending Traffic Monitoring to Internet and Cable data.
- Traditional Approaches: Packet Sniffing, Telephone call monitoring, etc.



IMSCAN: Framework Overview



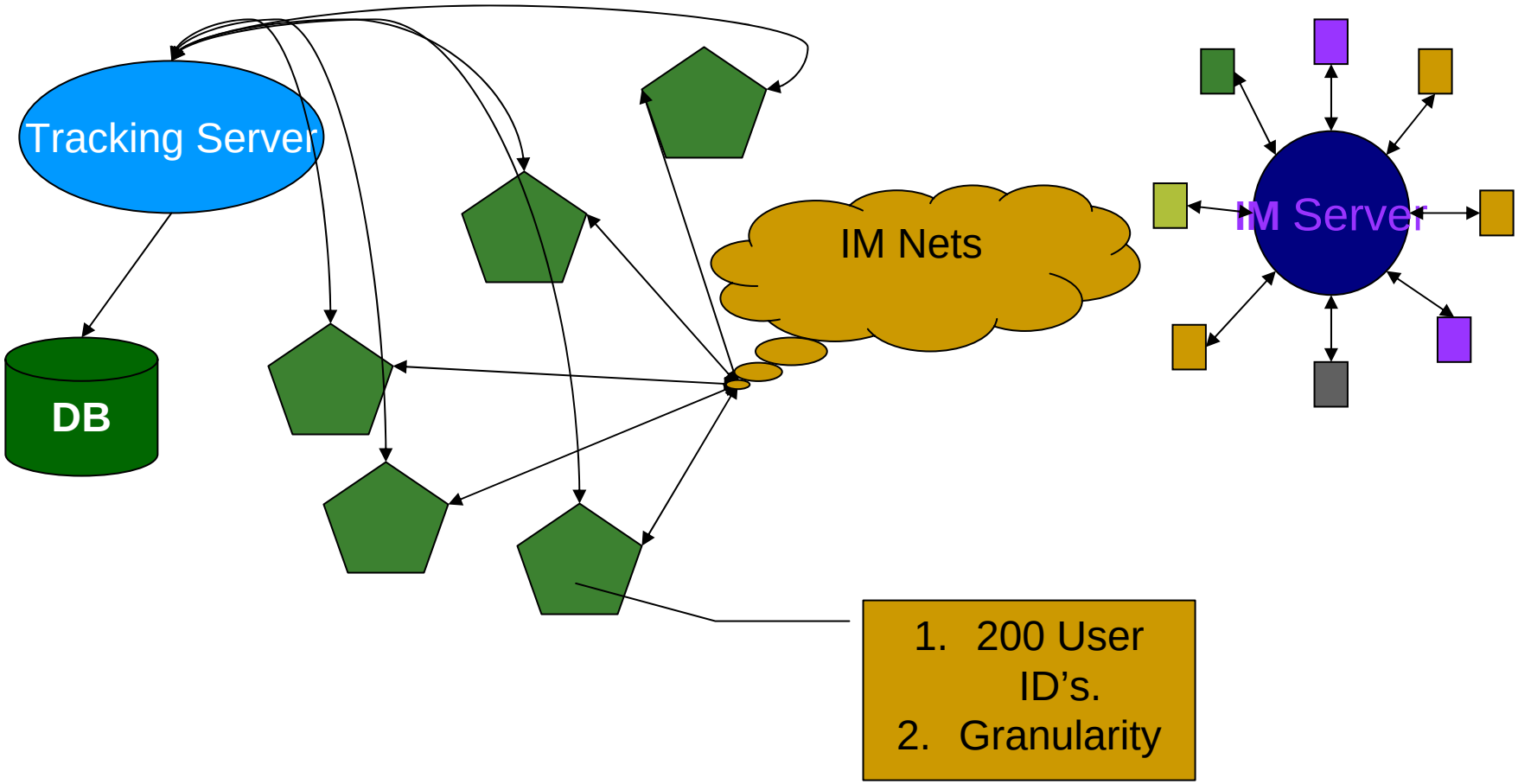
- Designed to work with the existing IM network infrastructure, providing “Status” information of users across networks and protocols.

User Status: Collection Format

- This paper: session lasted 67 days and tracked 207 users.
- Current Session: tracking approx. 5000 users.

Time	User	Status
4337217	151	Offline
4337239	151	Away
4337354	137	Idle
4337385	158	Idle
4337388	158	Away
4337399	119	Idle
4337419	43	Online
4337482	42	Offline

Scale of Tracking

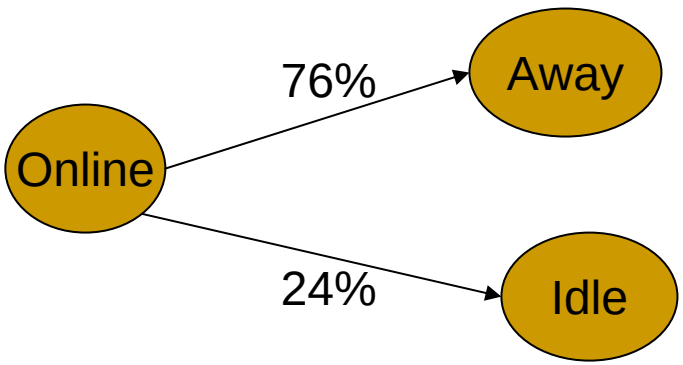


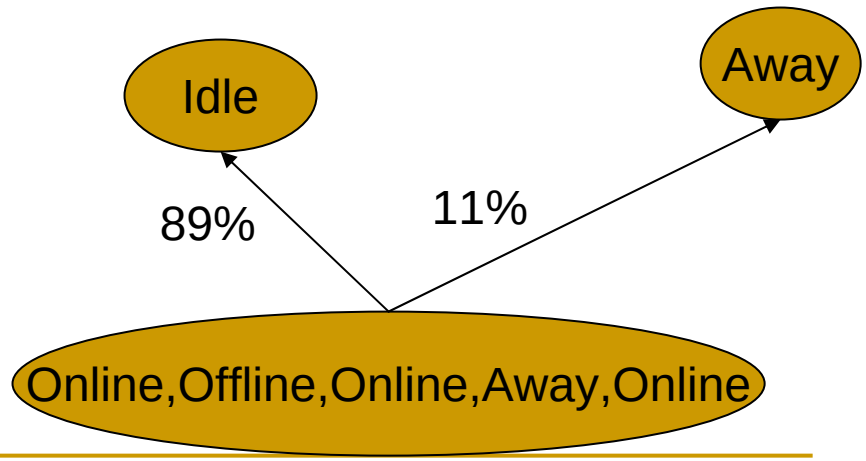
User Status: Analysis (cont.)

- Issues:
  - First: Predict a user's next status given their past N status changes.
  - Second: Predict when a user will switch from his current status to his next status.

User State: Prediction (Markov-Based)

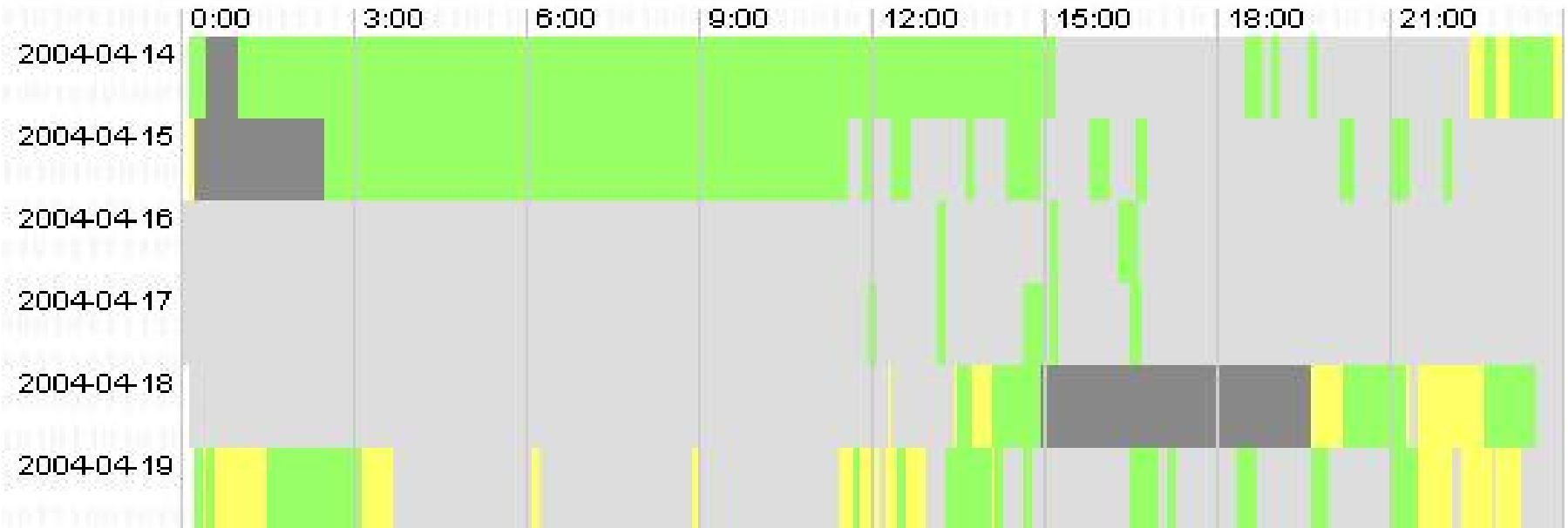
- Global and Personal First-Order Markov Models were constructed for the various user state sequences.
- State sequences from 1 to 25 actions long in length were tested, with varying results.





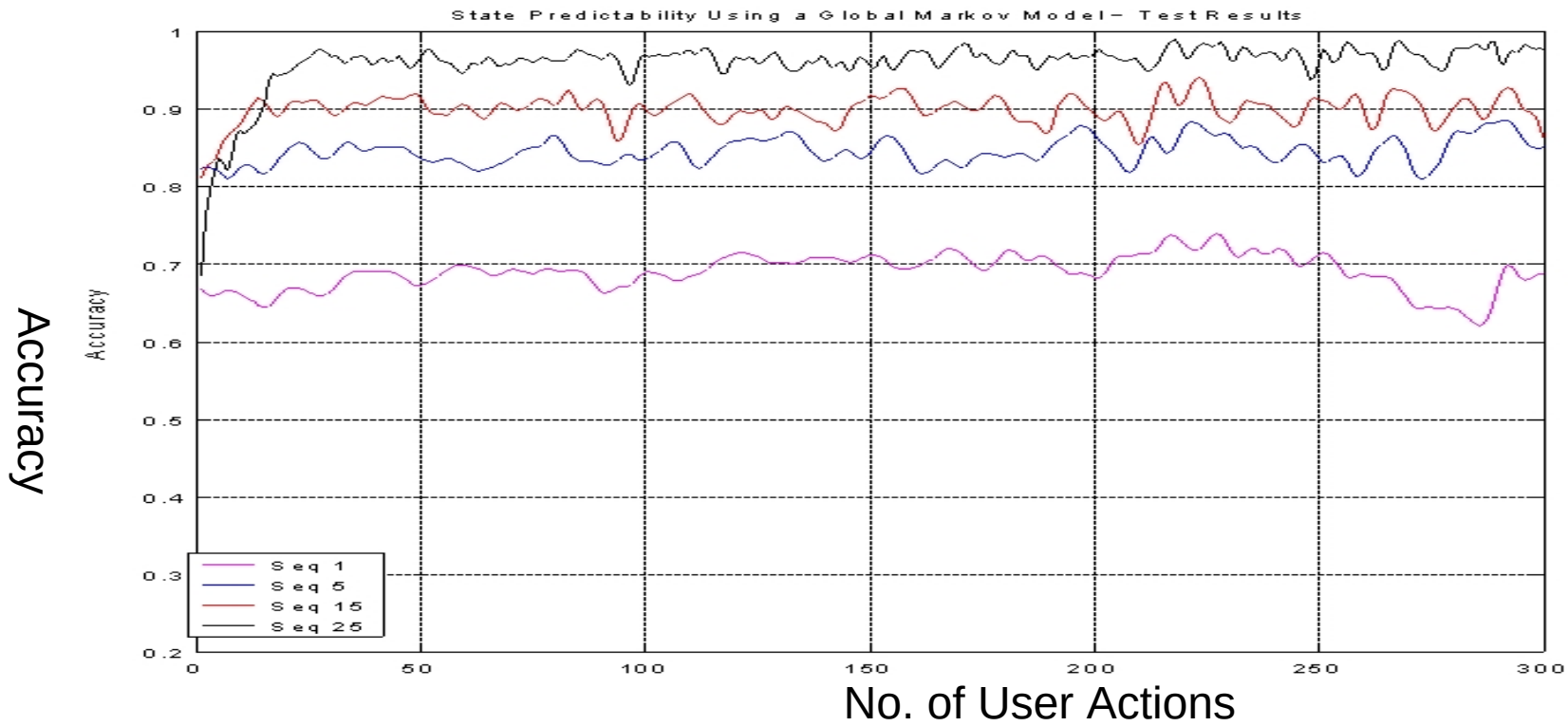
User Status: Analysis

- Tools: To determine when their friends are most likely to perform their next status change. (Under Development)



Green = Online Black Grey = Offline White Grey = Idle Yellow = Away

User State: Prediction (cont.)



- Using the global model shows state prediction accuracies upwards of 95%. While the personal models range from 80%-85% accuracy.

Pattern Analysis

- Usage models are constructed for each state, for each user.
- Using the results of the Markov Models, a user's next state is predicted.
- Using the model for the current state, and the model for the next state, a time for state change is predicted based on historical data.

IM as a Social Network

- A large aspect of Instant Messaging is the impressive amount of social network information discernable.
- All users of Instant Messaging networks have the ability to specify ‘friends’, or people with whom they communicate frequently.
- Using this friend information, one can glean additional information concerning how users interact with each other on a day-to-day basis.

Future Work

- Optimization of Markov Model creation – attempting to find the optimal settings under which a model can be quickly constructed.
- Correlating social network data and user status information.
  - Develop a user comparison metric and cluster users based upon their state information.
  - Find relationships between these clusters and the communities of users that exist in the physical social network.

The Troops

Students:

- John Resig
- Vineet Chaoji
- Willem Ave

Faculty:

- Christopher Homan
- Roger Gaborski

Questions?

- For more information, please feel free to contact us at:
  - John Resig jer5513@cs.rit.edu
  - Ankur Teredesai amt@cs.rit.edu
- Or visit our web page at:
  - http://www.cs.rit.edu/~dmrg/CDI/