Measuring and Modeling the Group Membership in the Internet

Jun-Hong Cui
University of Connecticut

Joint work with: Michalis Faloutsos (UC Riverside), Dario Maggiorini (Univ. Milan) , Mario Gerla, Khaled Boussetta (UCLA)

The Problem: Multicast Modeling

- The locations of the group members
  - Given a graph, where should we place them?
- Current assumptions: uniform random model (unproven)
  - All members uniformly distributed
  - Not realistic for many applications

Group Modeling is Critical

- Some studies have shown the locations of members have significant effects on
  - Scaling properties of multicast trees [Phil99, Chal03]
  - Aggregatability of multicast state [Thal00]
  - Performance of state reduction schemes [Wong00...]
- Realistic group models
  - Improve effectiveness of simulation
  - Guide the design of protocols

Our Contributions

- Measure real group membership properties
  - MBONE (IETF/NASA) and Netgames (Quake)
- Design a model to generate realistic membership
  - GEneralized Membership Model (GEM)
  - Use Maximum Enthropy: an excellent statistical method

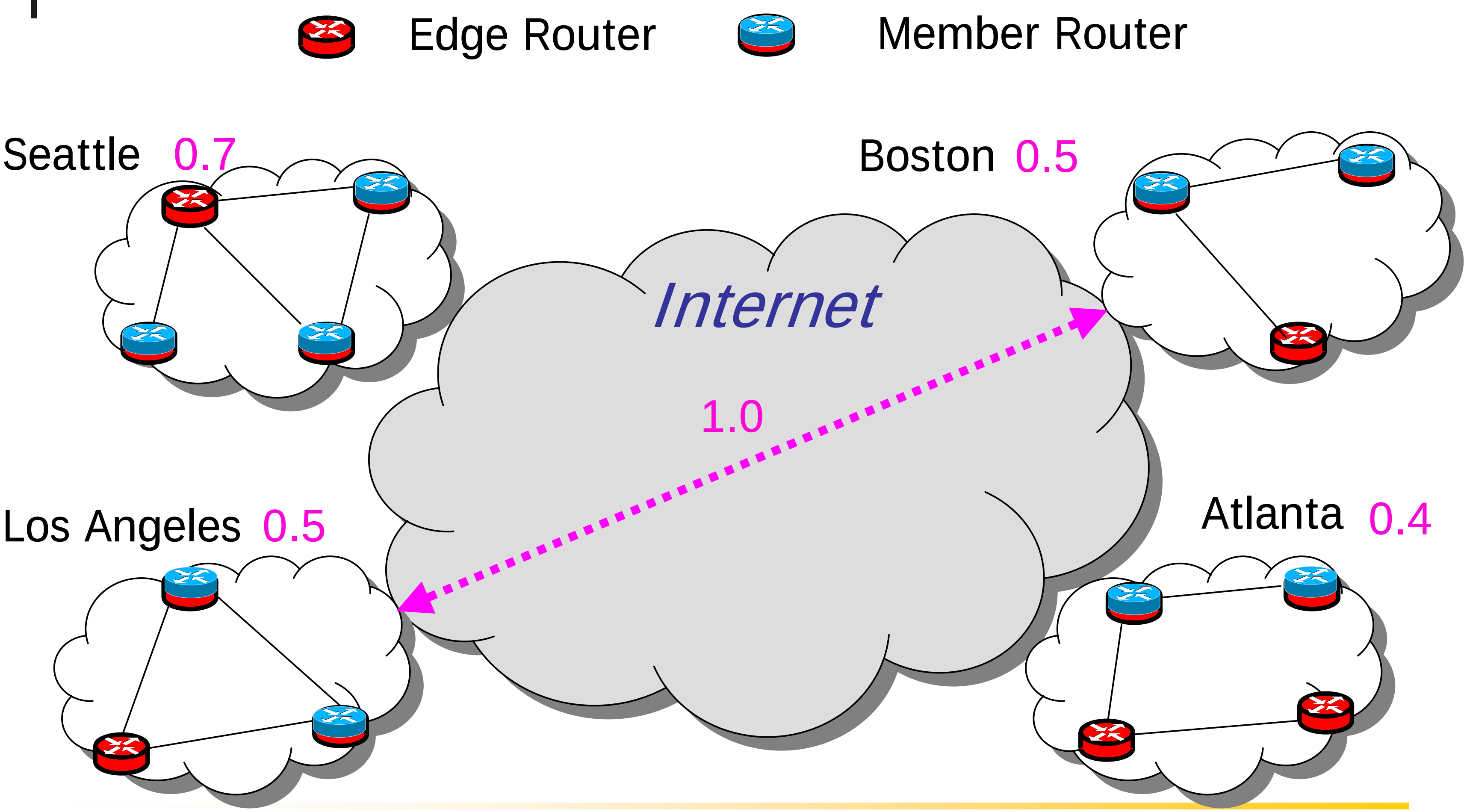
Roadmap

- Introduction
- Membership Characteristics
- Measurement and Analysis Results
- Model Design and Validation
- Conclusions and future work

Beyond Uniform Random Model

- How close are the members in a group?
- Are all the members same in group participation?
- What are the correlations between members in group participation?

An Illustration (Teleconference)



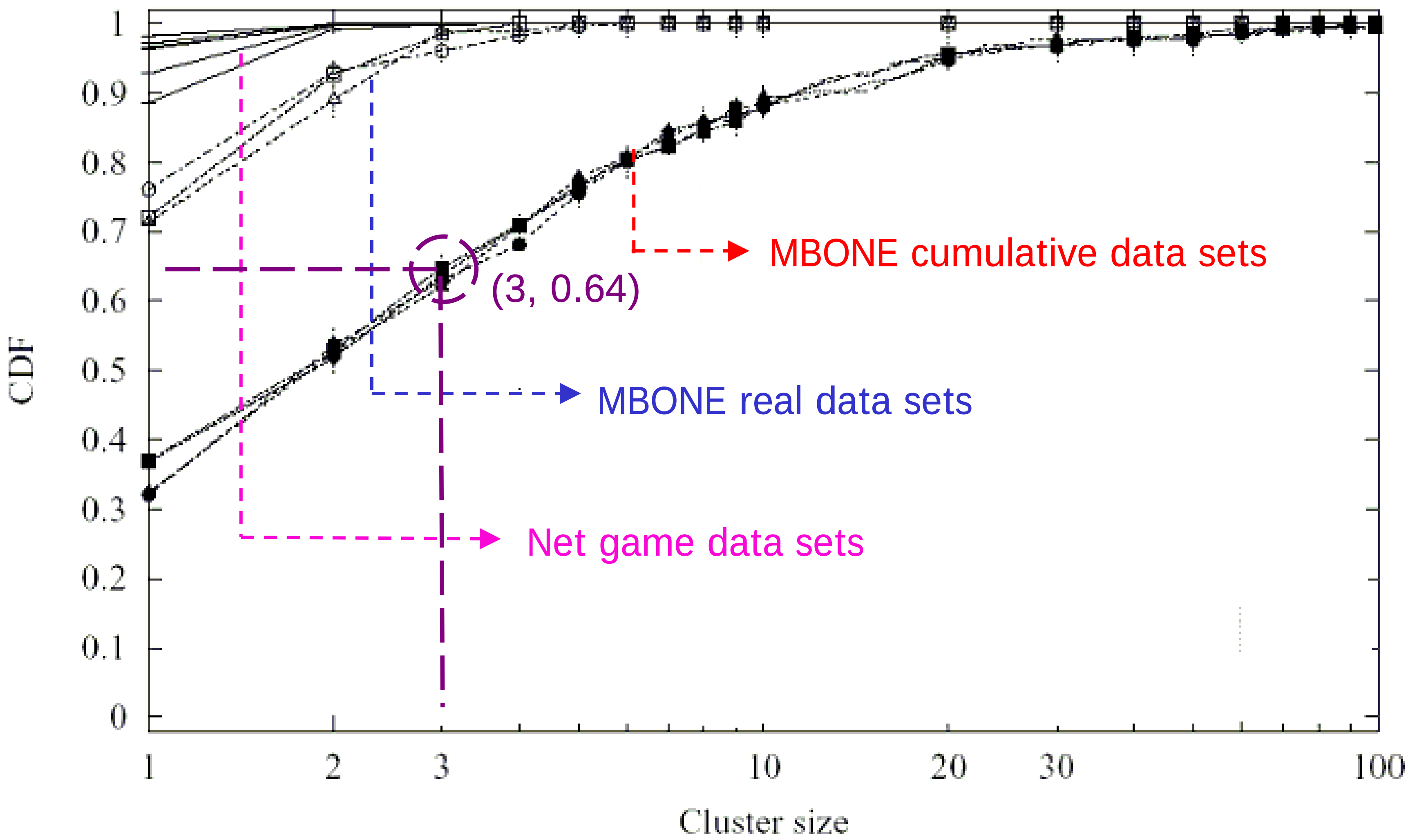
Membership Characteristics

- Member clustering
  - Capture proximity of group members
  - Use network-aware clustering method [Krish00]
- Group participation probability
  - Show difference among members/clusters
- Pairwise correlation in group participation
  - Capture joint probability of two members/clusters
  - Use correlation coefficient (normalized covariance)

Measure Membership Properties

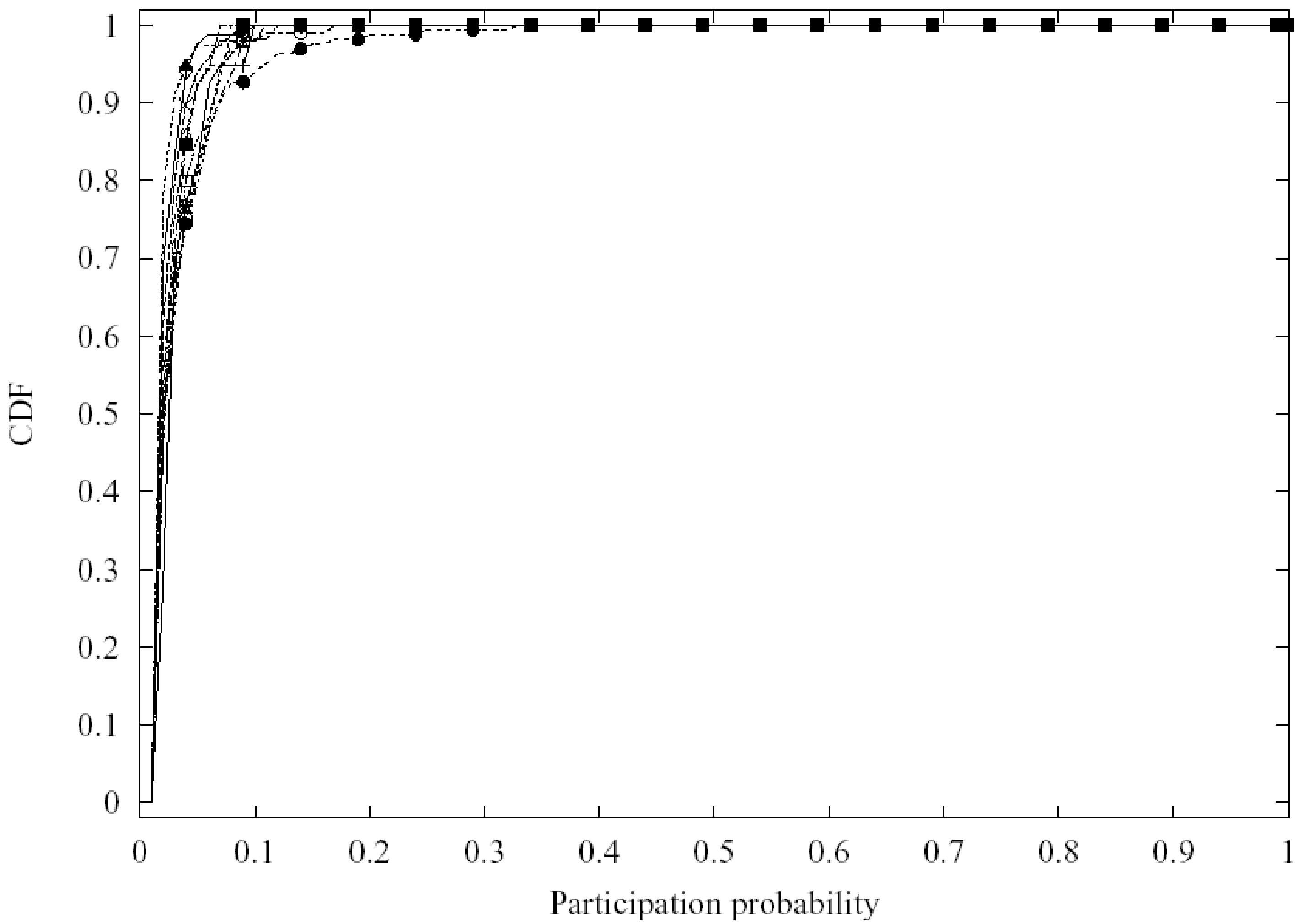
- MBONE applications (from UCSB)
  - IETF-43 (Audio and Video, Dec. 1998)
  - NASA Shuttle Launch (Feb. 1999)
  - Cumulative data sets on MBONE (1997-1999)
- Net Games (using QStat)
  - Quake I (query master server)
  - Choose 10 most popular servers (May. 2002)
- Examine three membership properties

Member Clustering



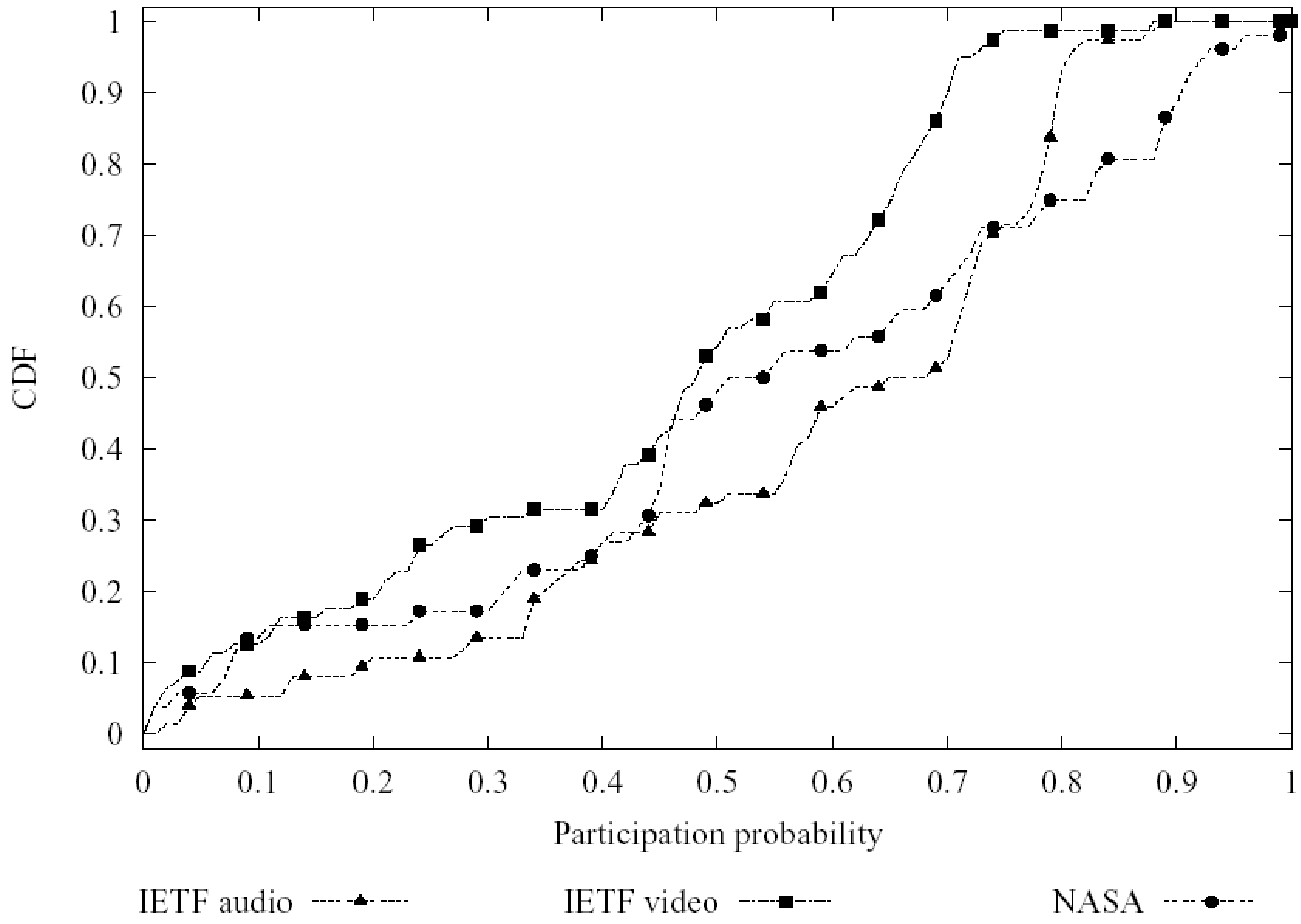
CDF of cluster size for MBONE and net games

Group Participation Probability



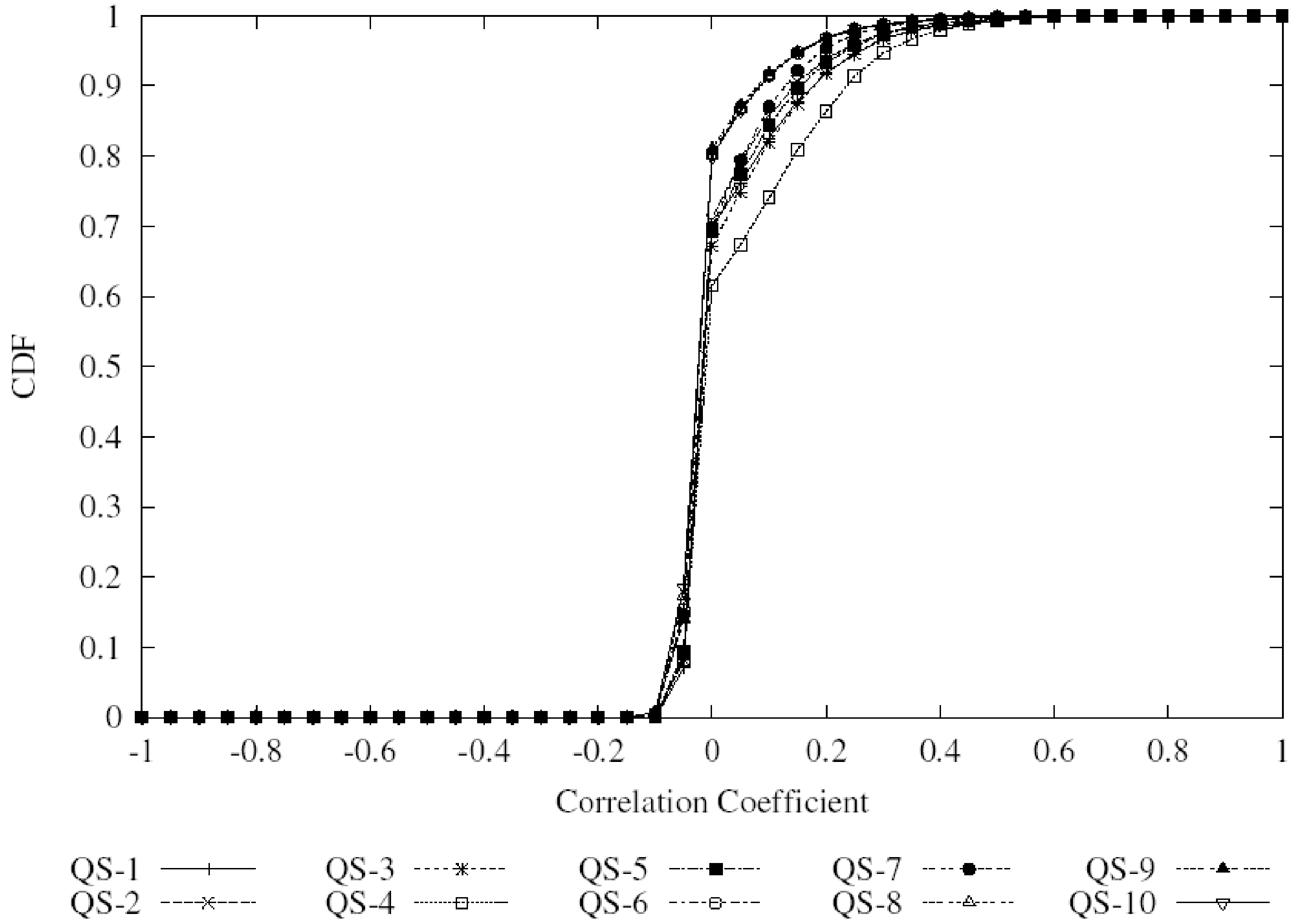
CDF of participation probability for Net Game data sets

Group Participation Probability



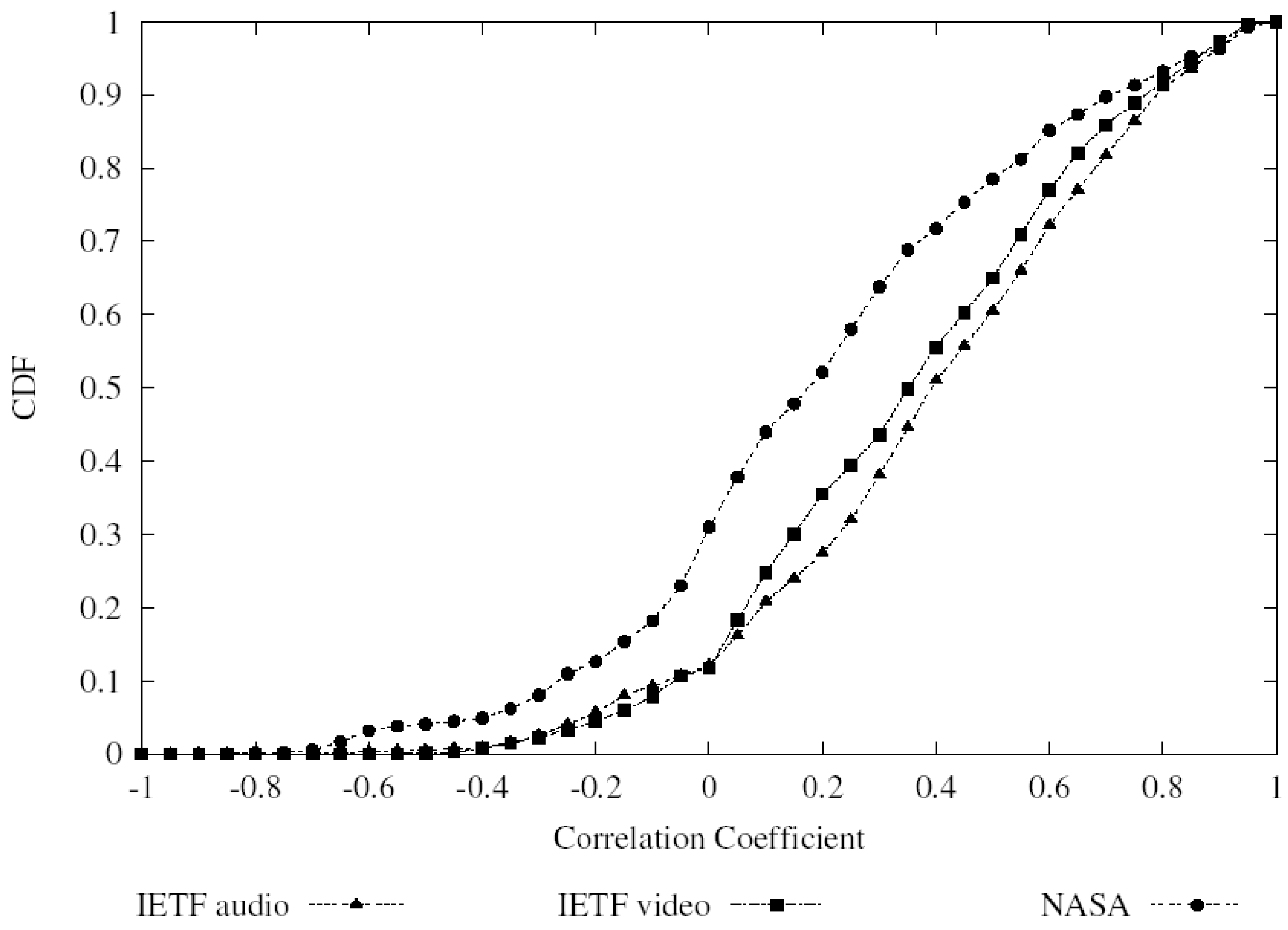
CDF of participation probability for MBONE applications

Pairwise Correlation in Group Participation



CDF of correlation coefficient for Net Game data sets

Pairwise Correlation in Group Participation



CDF of correlation coefficient for MBONE applications

Generalized Membership Model --- GEM (An Overview)

Inputs

- Network topology
- Cluster method
- Group behavior
  - Distr. of participation Prob.
  - Distr. of pairwise correlation
  - Distr. of member cluster size

↓

GEM

1. Create clusters in given topology
2. Select clusters as member clusters
   - According to input distributions
3. Choose nodes for each member clusters

↓

Outputs

Desired number of multicast groups that follow the given distributions

Member Distribution Generation

- Definition:

  K clusters: $C_1, C_2, \ldots, C_i, \ldots, C_K$

  Prob. p_i for any i in [1, K]

  Joint prob. $p_{i,j}$ for any i, j in [1, K]

  $X=(X_1, X_2, \ldots, X_i, \ldots, X_k)$: X_i is a binary indicator

  $X_i = 1$ if C_i is in the group

  $X_i = 0$ if C_i is not in the group

- Objective:

  Generate vectors $x=(x_1, x_2, \ldots, x_k)$ satisfying

  $P(X_i = 1) = p_i$ and $P(X_i = 1, X_j = 1) = p_{i,j}$

Maximum Entropy Method

- To calculate the distribution of $(X_1, X_2, \ldots, X_k)$ requires $O(2^K)$ constraints
- But we only know $O(K+K^2)$ constraints
- We use Maximum Entropy Method
  - Entropy is a measure of randomness
  - We construct a maximum entropy distr. $p^*(x)$
    - Satisfy constraints in specified dimensions
    - Keep as random as possible in unconstrained dimensions
    - i.e. maximize entropy while match given constraints

Three Cases

Considering $P(X_i=1)= p_i$ and $P(X_i=1, X_j=1)=p_{i,j}$

1. Uniform distr. without correlation (easy)

   $p_{i,j} = p_i * p_j$, and $p_i = p_j$

2. Non-uniform distr. without correlation (easy)

   $p_{i,j} = p_i * p_j$, but $p_i = p_j$ not necessary

3. Non-uniform distr. with pairwise correlation

   Neither $p_{i,j} = p_i * p_j$ nor $p_i = p_j$ necessary

   Need to calculate the maximum entropy distr. $p^*(x)$

Entropy decreases from case 1 to case 3

Calculate the Maximum Entropy Distribution

The maximum entropy distr. $p^*(x)$ is the solution for:

$$p^*(x) = arg\ \ max\{-\int p(x) log p(x) dx\}$$

Subject to

$$\int x_i x_j p(x) dx = p_{i,j}, when\ \ i \neq j$$

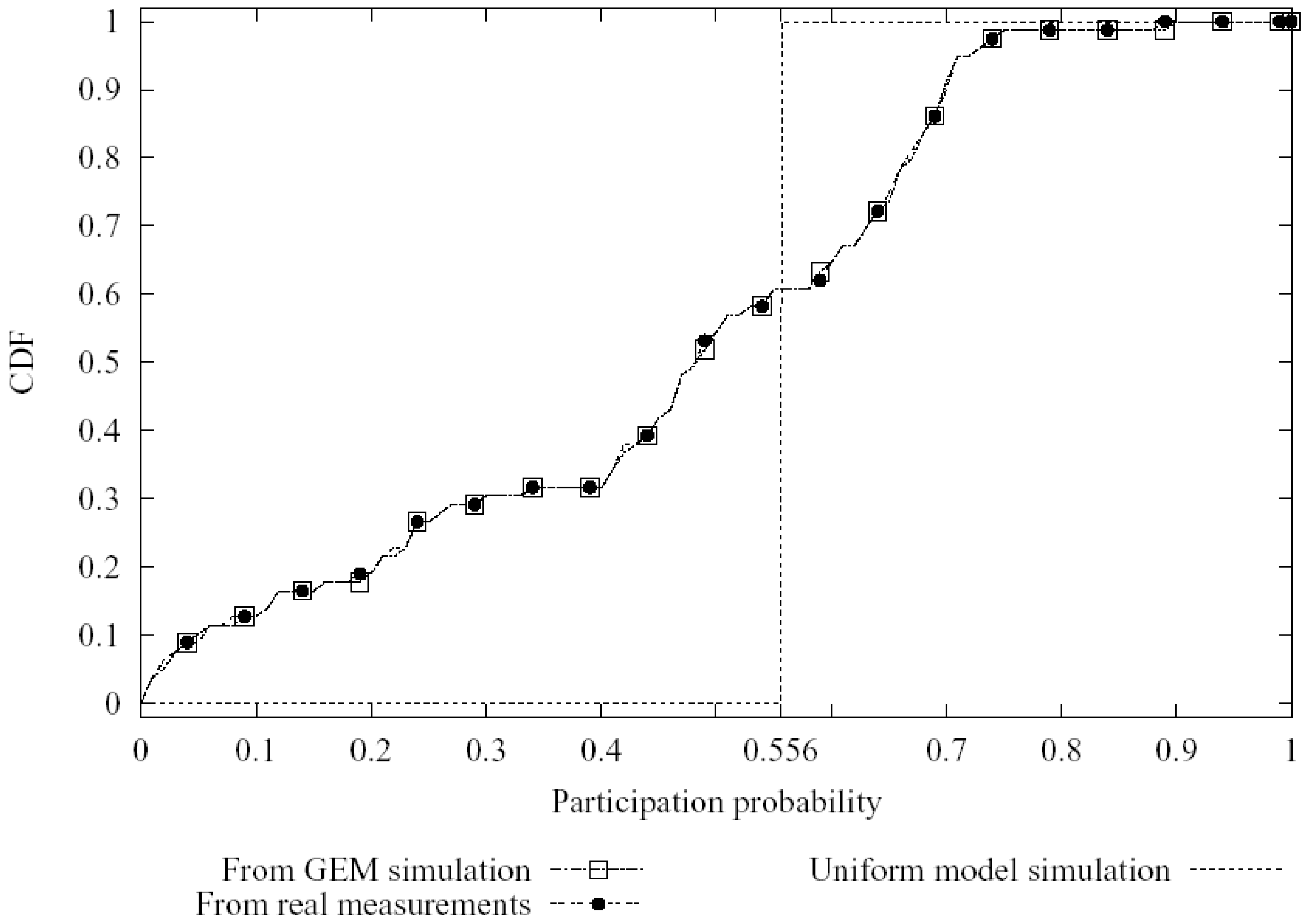
and $\int x_i p(x) dx = p_i$ and $\int p(x) dx = 1$

Use lagrange multipliers and numerical method to construct *p*(x),* Then Gibbs Sampler to sample it

Experimental Validation

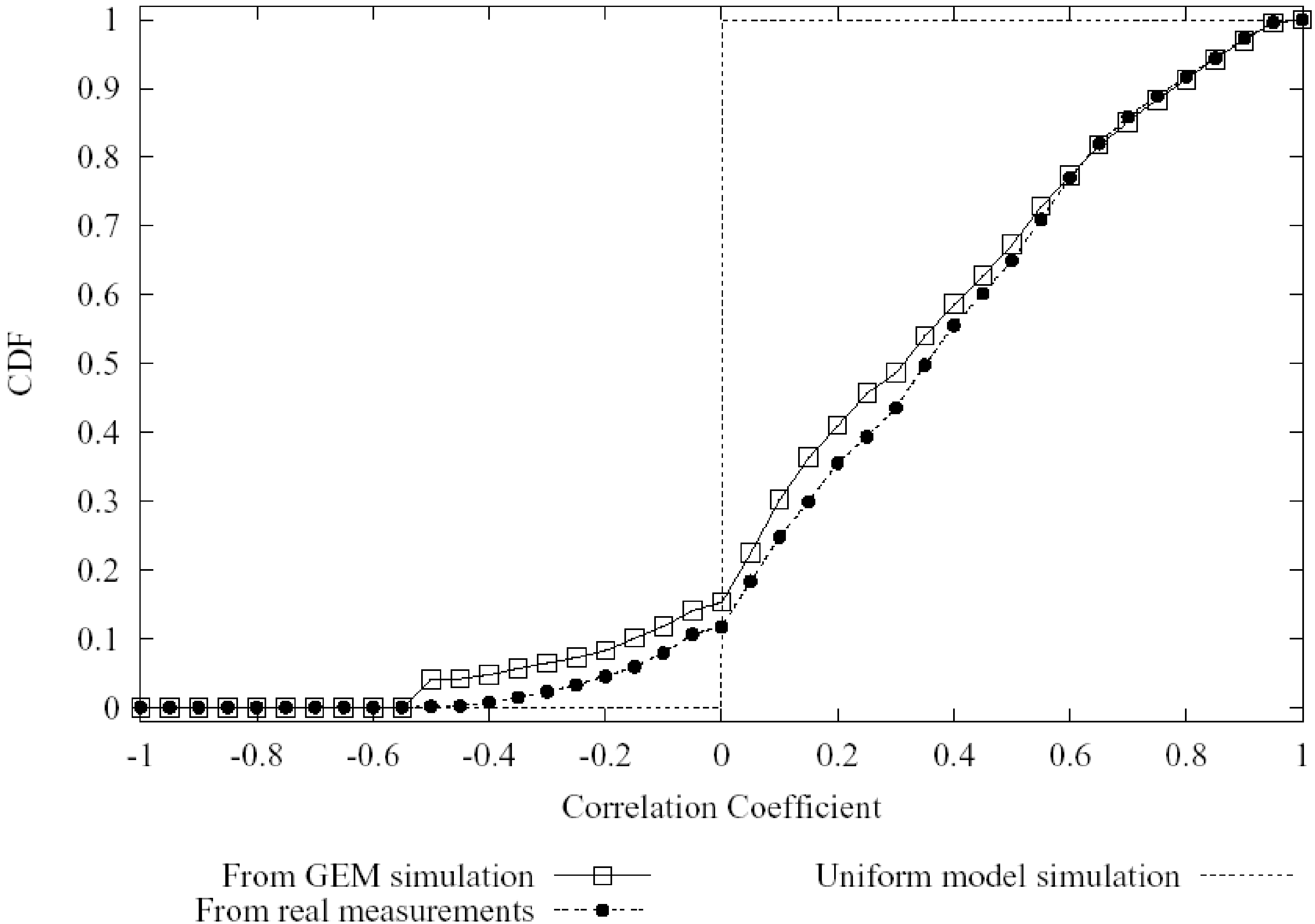
- Our Goal:
  - GEM can regenerate groups satisfying given distributions
  - Distributions are from real measurement
- Focus on the challenging case 3
- Use IETF-43 and NASA data sets
- Consider two membership properties
  - Group participation probability
  - Pairwise correlation in group participation

Group Participation Probability



Participation probability distribution for IETF43-Video

Pairwise Correlation in Group Participation



Joint probability distribution for IETF43-Video

Conclusions

- Uniform random model
  - Can capture net games approximately
  - But not realistic for MBONE applications
- GEM: a generalized membership model
  - Three cases (case 1: uniform random model)
  - Realistic membership can be regenerated

Future Work

- Study more applications
  - Different applications have different distributions

- Beyond multicast
  - Web-caching, peer-to-peer

- Beyond wired network
  - Wireless adhoc networks, sensor networks ...

Questions?

jcui@cse.uconn.edu

http://www.cse.uconn.edu/~jcui

THANKS!!!