Video Popularity Distributions and Potential of Peer-Assisted

政策・メディア研究科 後期博士課程

Mohamad Dikshie Fauzie

dikshie@sfc.wide.ad.jp

MAUI 2014/04/28

Status Updates

- Formal: 7/7/2010
- Kouchoukai: 8/20/2013
- Sechi: 8/28/2013
- Saishusinsha deadline: 8/28/2014?
- Conference: 1
- Journals: 2

Setting

- Peer-to-Peer (P2P) usage is changing. (e.g peer-assisted Content Delivery Network(CDN))
- Analytical tools & measurements help us make good decisions. (e.g. bandwidth trade off, incentive for users, energy consumption).

Problems

- Current model is not good enough:
  - limited measurement in closed environment [Dale et al, (2008): in PlanetLab]
  - simulators [Al-Hamra et al, (2007): discrete event simulator]
- We need to measure in real world
- Nobody knows swarm topology in real bittorrent network

Contributions

- First real world measurements of swarm topology
- Proposing new model for global energy consumption of peer-assisted CDN
- In both, found new results that extend or contradict prior work

ToC Dissertation

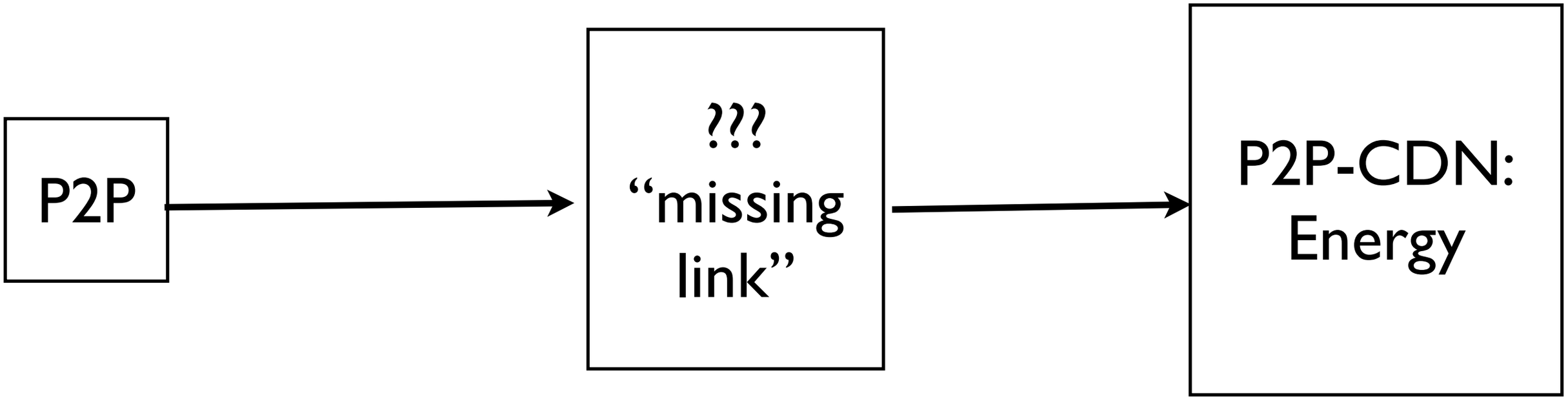
Kochoukai Notes

1. P2P topology dynamics
2. Video popularity in P2P
3. How to check energy consumption

ToC Dissertation

1. Introduction
2. P2P Content Delivery
3. Characteristic of Bittorrent Swarms
4. Peer-Assisted CDN delivery
5. Energy Saving
6. Conclusion and Future Work

Story

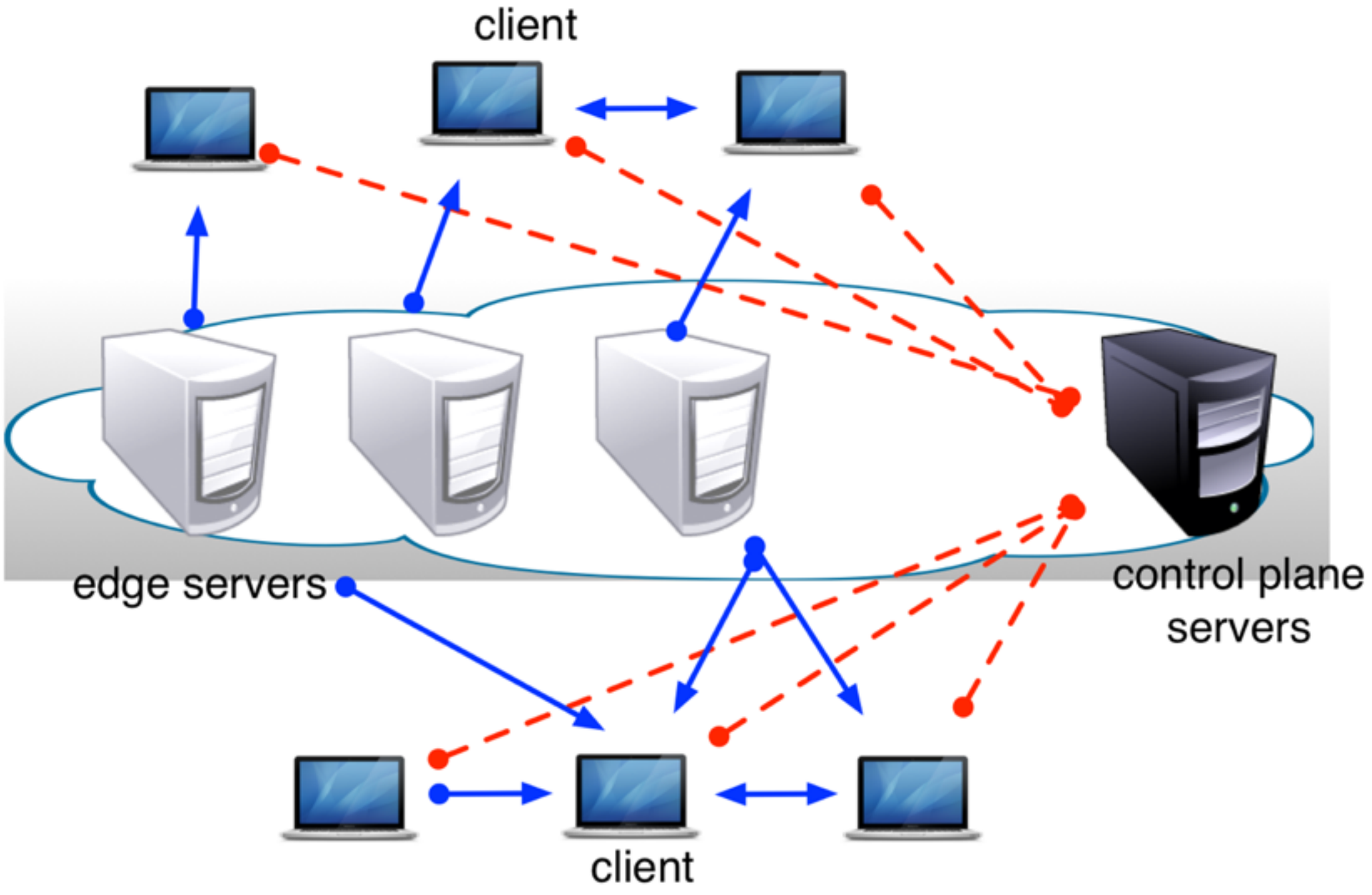


One journal is required to fill the gap between P2P topic and energy topic Peer-Assisted CDN.

Setting

- Peer-to-Peer (P2P) usage is changing. (e.g peer-assisted Content Delivery Network(CDN))
- Analytical tools & measurements help us make good decisions. (e.g. bandwidth trade off, incentive for users, energy consumption).

What's Peer-Assisted CDN



Research Question

- Current P2P-CDN model only considering popularity based on local-system
  - counting of frequency access from peers
- Can I add VoD popularity model to assist/ help P2P-CDN system?
  - while maintaining same good metrics or even make it better

General Methodology

1. VoD popularity evolution characterization
2. Modified current cache replacement strategy with my strategy that utilised VoD popularity evolution model.
   - I use PROP [1] as basis.
   - [1] Guo et al., “Design and Evaluation of a Scalable and Reliable P2P assisted proxy for on demand streaming media delivery. IEEE TKDE 2006.
3. Feed to Simulator

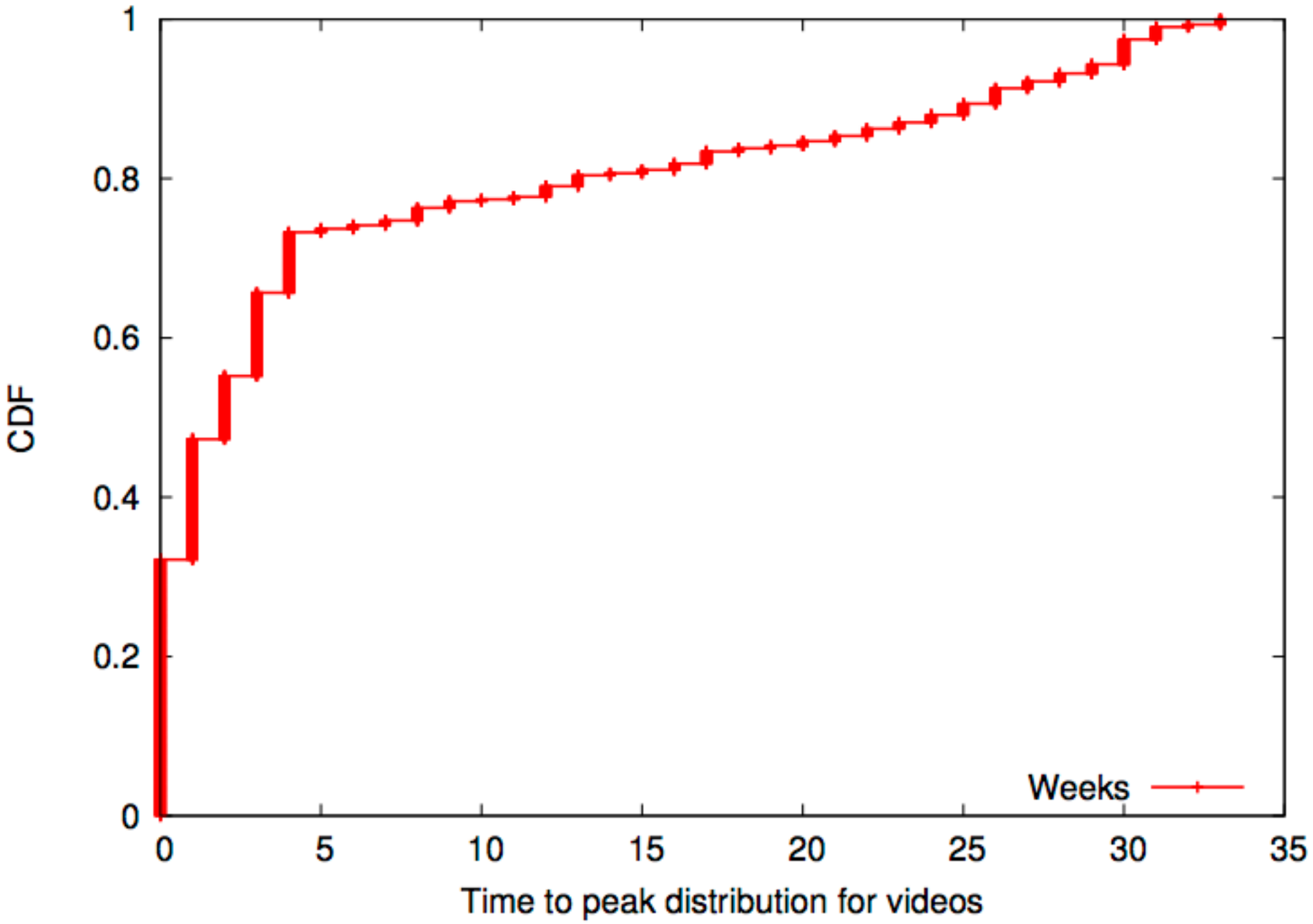
VoD Popularity (1)

- Take Youtube VoD as example.
- Data-sets (36 weeks) from Borghol et al.,
  - “Characterizing and modeling popularity of user generated videos”, performance evaluation, 2011.
- Added with my data-sets (8 weeks).
- Estimate when a video is in peak popularity phase.

VoD Popularity (2)

- There are 3 phases of characterization of popularity evolution:
  - before-peak, at-peak, and after-peak.
  - time-to-peak video is the age at which they attain their peak popularity since upload.
- definition “at peak”:
  - at which its weekly viewing rate is the highest within our measurement period.

VoD Popularity (3)



- the time to peak distribution shows that a large fraction of the videos, approximately three quarters of them, peak within the first 6 week since their upload.

- Distribution: mixture between exponential and uniform. for exponential part using MLE we can determine alpha=0.6. greater than 6 weeks: uniform U(6,d).

VoD Popularity (4)

- Q: How to estimate video's at-peak?
- A: sample N values from time-to-peak distribution and determine the number of videos that peak at week i.

Current Cache Strategy (1)

- PROP's CDN cache strategy
- PROP's Peer cache strategy
- My modification to PROP's peer cache strategy

Current Cache Strategy (2) PROP's CDN cache strategy

- We use PROP's method to estimate the popularity video: $$P_j = min\left\{\frac{n_j^r}{t_j^r - a_j}, \frac{1}{t - t_j^r}\right\}$$

- Long-term requests: $\frac{n_j^r}{t_j^r - a_j}$
- Recent request rate: $\frac{1}{t - t_j^r}$
- The smallest popularity is chosen as candidate to be replaced when the CDN cache is full.

Current Cache Strategy (3) PROP's Peer cache strategy

- We use PROP's method to calculate utility function for peer cache strategy:

$$u = \frac{(f(p)-f(p_{min}))\times(f(p_{max})-f(p))}{r^{\alpha+\beta}}$$

- p represent popularity of the video.
- p_min represents estimation of min popularity
- p_max represents estimation of max popularity
- r represent number of replica
- f(p) represent monotonic non-decreasing function, we choose log(p).

Current Cache Strategy (4)
Modified PROP's Peer cache strategy

- Because I can estimate at-peak week video, I modified PROP's utility function as follows:
- Before-peak and After-peak: $u = \frac{f(p)-f(p_{min})}{r^{\alpha+\beta}}$
- At-peak: $u = \frac{f(p_{max})-f(p)}{r^{\alpha+\beta}}$

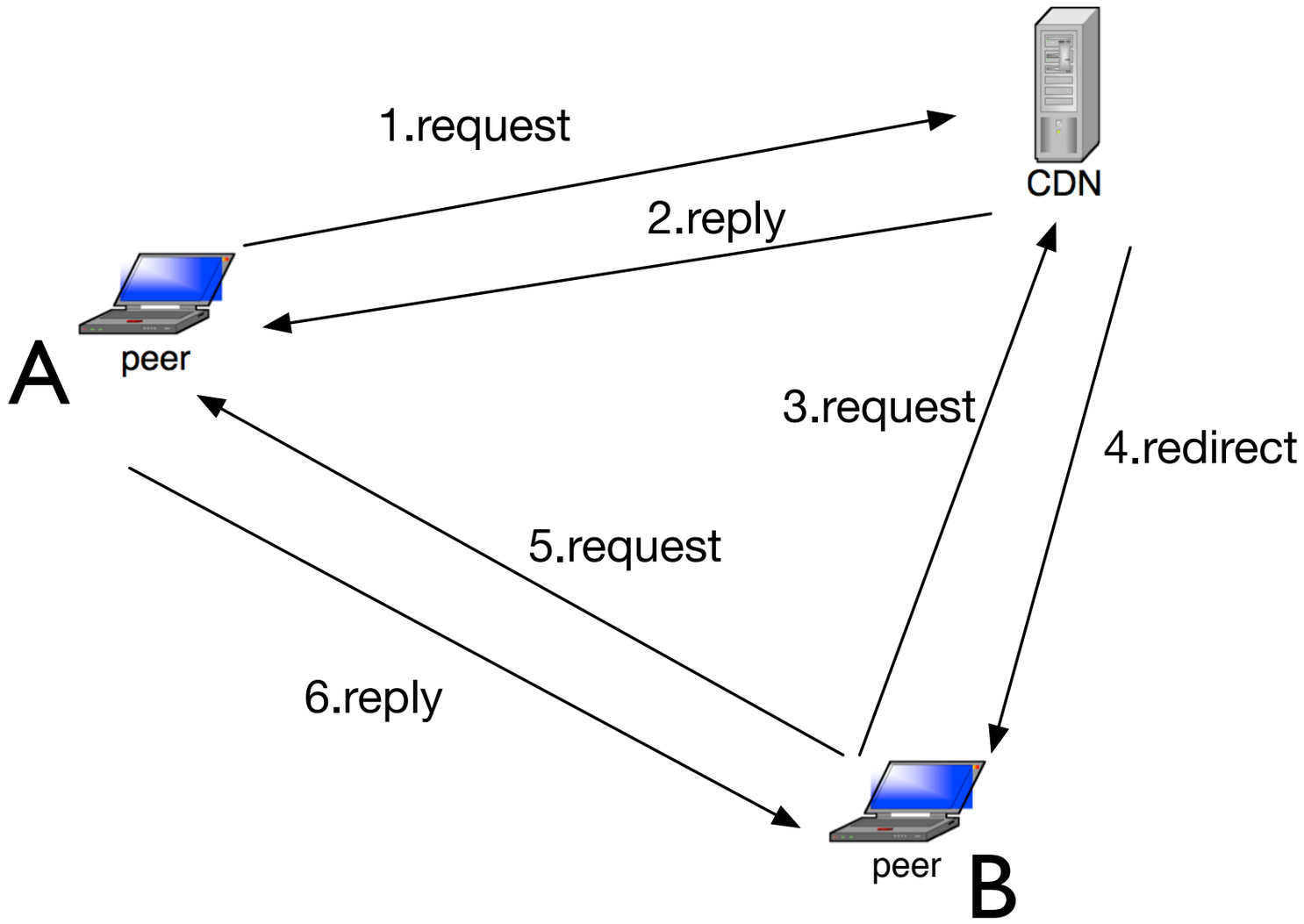
Simulator (1)

1. Simulator Design
2. Simulator Parameter

Simulator (2): Design

- Event driven simulator is developed using python
- Peer request to CDN assume to be Poisson process with mean rate (lambda) = 1.1, with 3600 times request per hour
- Time is divided into rounds, during a round a peer request a video

Simulator (3): Design



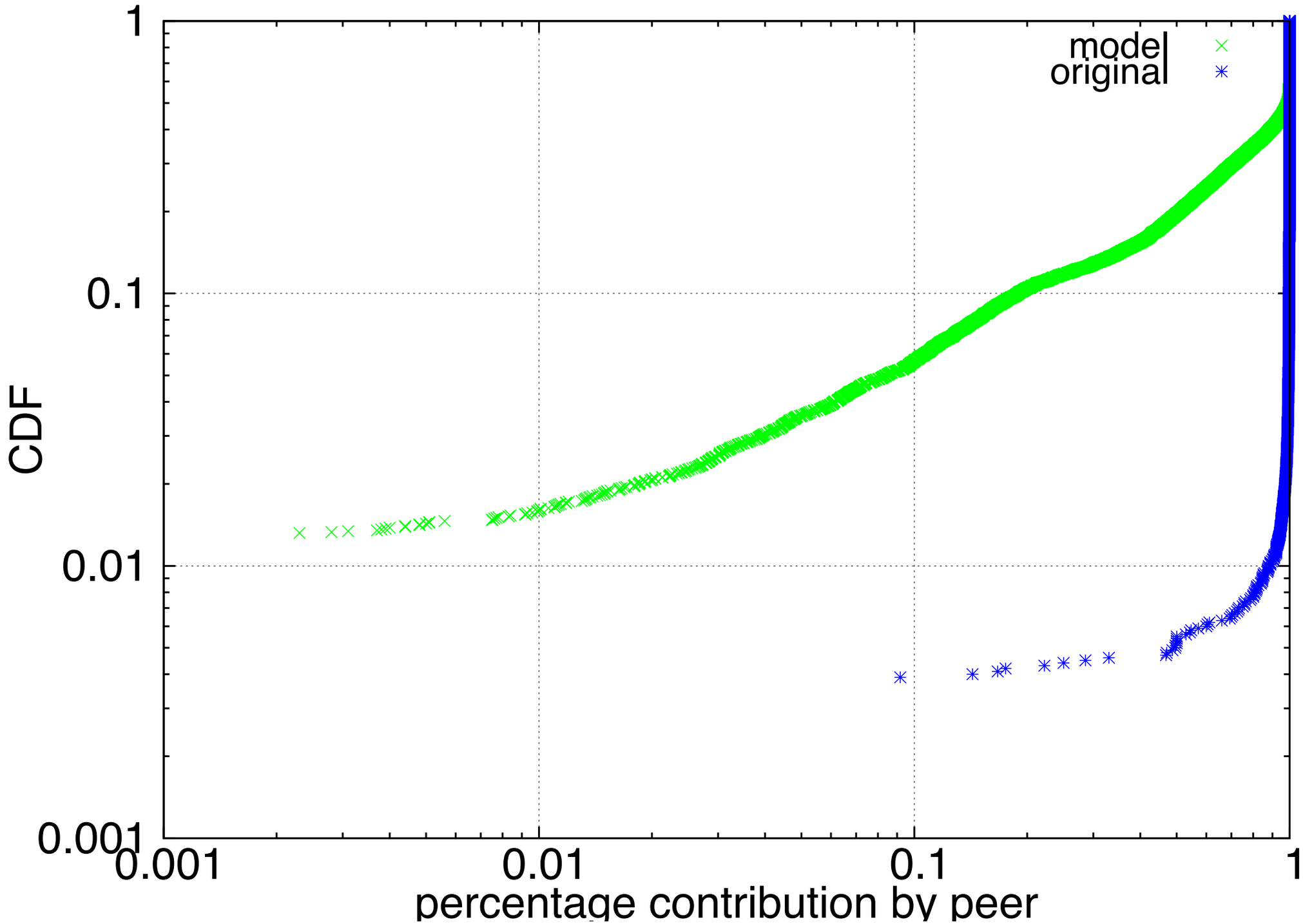
Simulator (4): Parameters

- Length: 360 days
- Video size random: between 1MB and 200MB
- Peer capacity: 500MB and 1000MB
- CDN capacity: 10000MB
- Number of peers: 100000 peers
- Number of videos: 10000 videos
- Symmetric peer load/download rate: 1Mbps

Evaluation

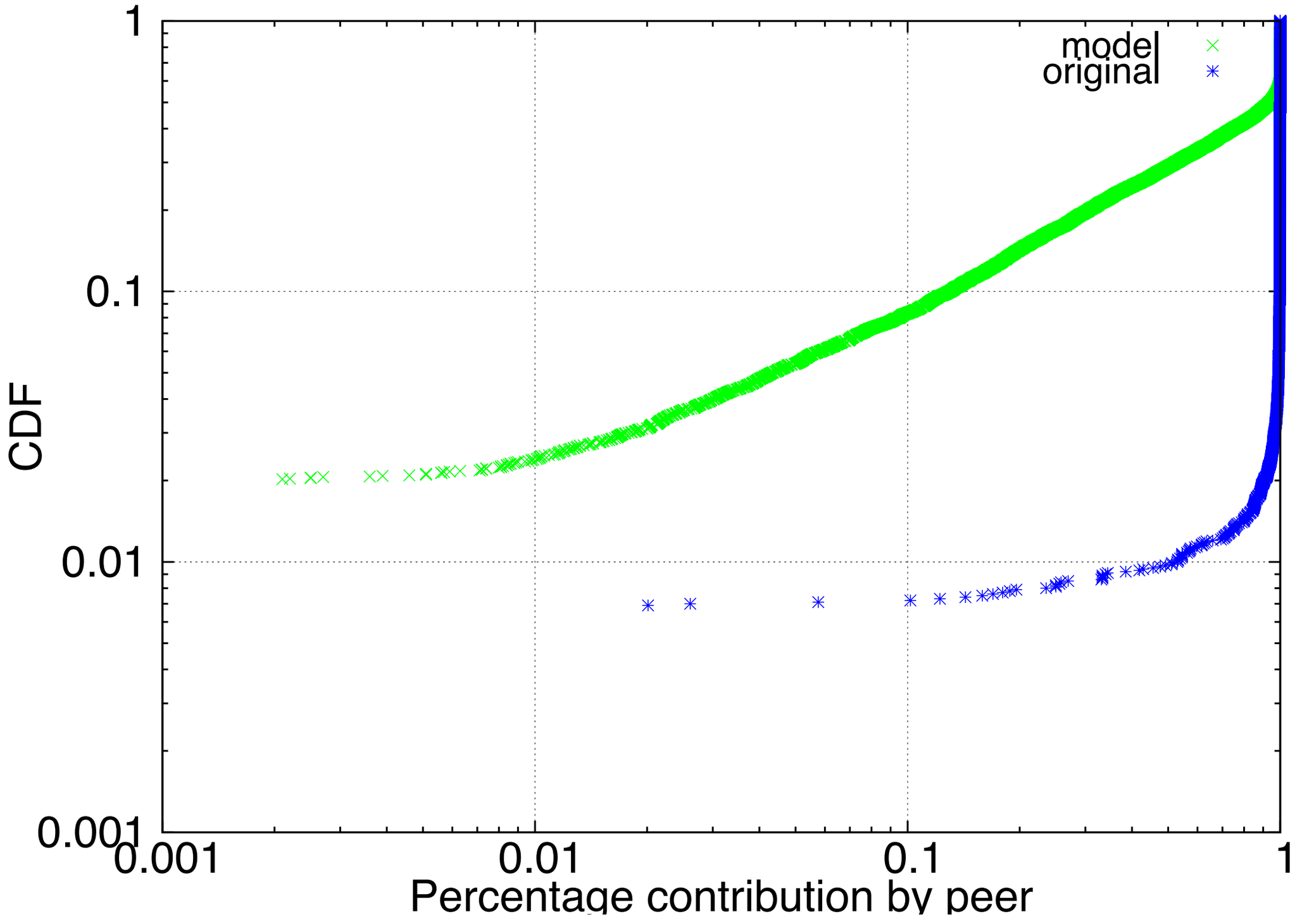
- Compare PROP and my work.
- Metrics:
  - Percentage of peer contributions
  - #Replicas

Evaluation (1): Peer Contribution



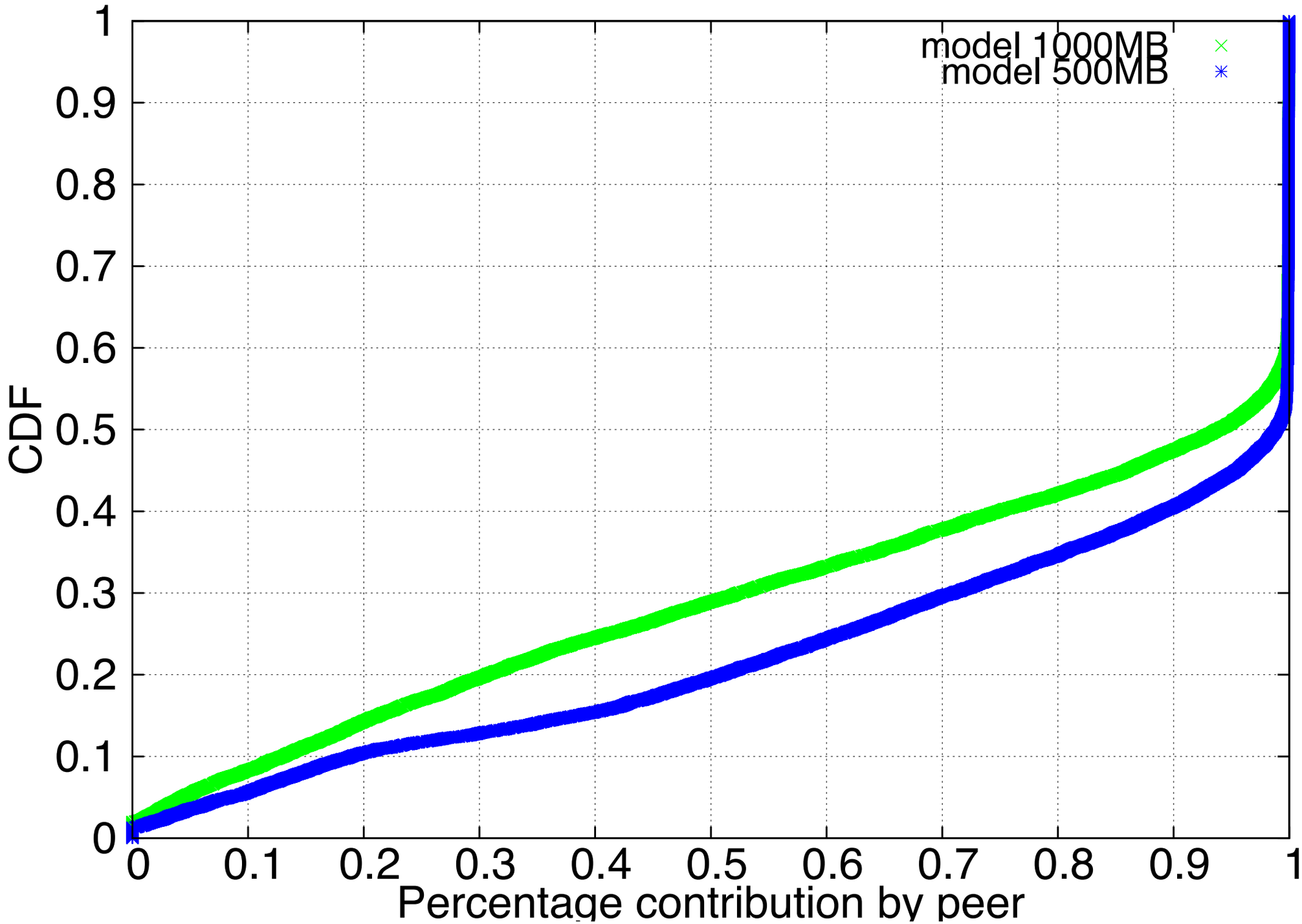
- Peer capacity = 500MB

Evaluation (2): Peer Contribution



- Peer capacity = 1000MB

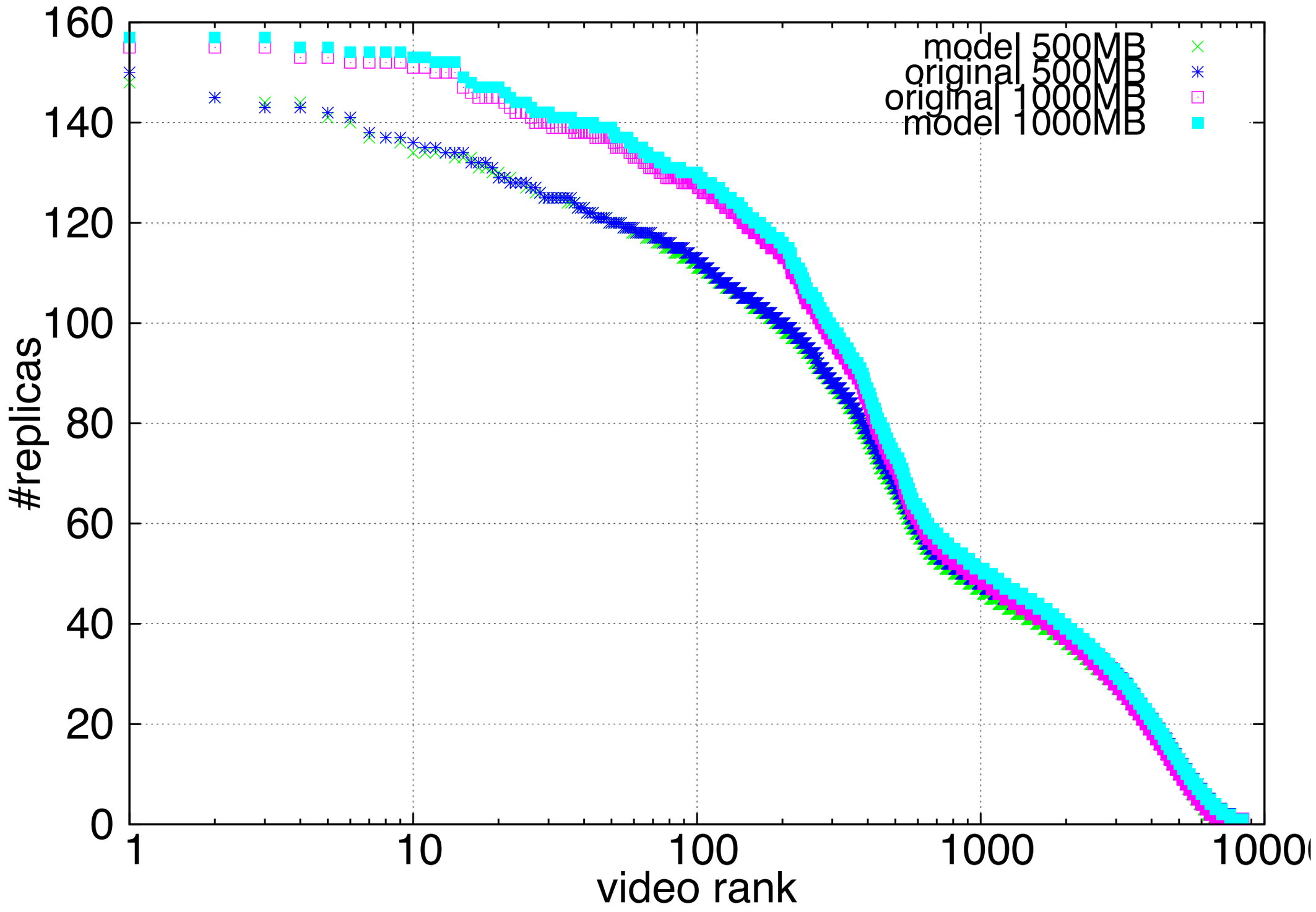
Evaluation (3): Peer Contribution



Evaluation (4): Peer Contribution

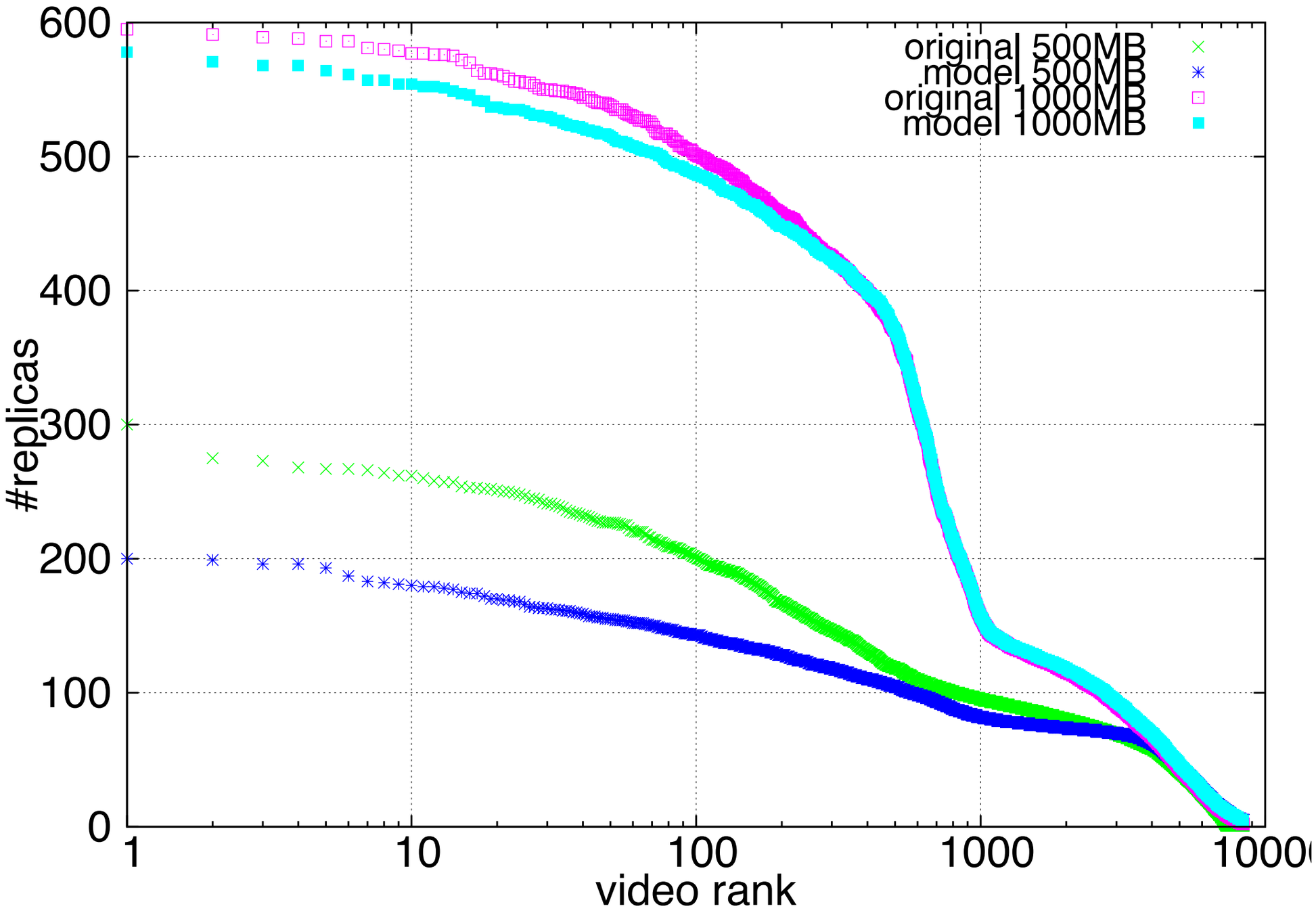
- Q: Are the both results (comparing PROP and model) significant?
- A: Use K-S statistics for significance testing.
  - for both model (500MB and 1000MB capacity) we get p-value: 0.5e-005 and 0.46e-005
  - since p-value less than 1%, the results are significant

Evaluation (5): #Replicas



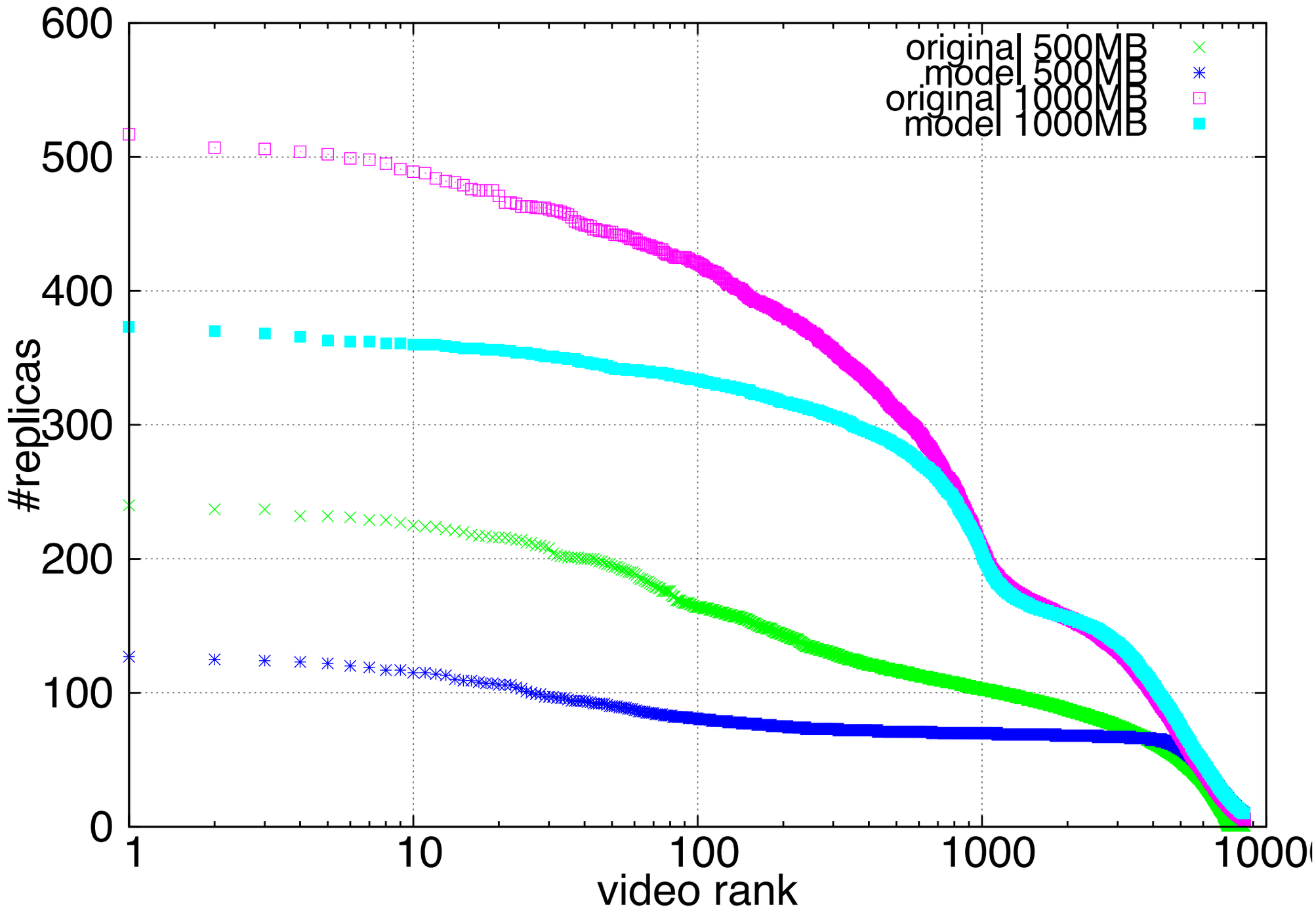
- snapshot t=10day. (t= 864000)

Evaluation (6): #Replicas



- snapshot t=6week (t=3628800 sec)

Evaluation (7): #Replicas



- snapshot t=10week (t= 6048000 sec)

Closing Remarks

- Propose modification to PROP gives good results:
  - More peer contribution
  - Lower replicas
- Contribution to dissertation:
  - since we have higher peer contribution we can move some of energy from CDN to peer!

Schedule

- May 1st: send to rdv@ for english grammar checking & fix the mistakes
- May 10: submit to Hindawi ISRN Communication and Networking
  - in case of Jin Nakazawa and H.Tokuda先生 (2012): 34 days from submission to acceptance.
- May 10th ~: Add this work to dissertation and cleanup dissertation.

Finish