

Christina Heider

19. October 2012 – Master Thesis Defense

# Measuring Freenet

# Outline

1. Background
2. Goal
3. Measurement Setup
4. Results
  1. Network Size & Participants
  2. Churn
  3. File Requests
5. Conclusion & Outlook

# Outline

1. **Background**
2. Goal
3. Measurement Setup
4. Results
  1. Network Size & Participants
  2. Churn
  3. File Requests
5. Conclusion & Outlook

# Freenet

## Goal

- Distributed Storage
- Anonymous information management

## Implementation

- Each member contributes disk space
- Each file owns a key
- Indirect routing
- Two modes: Darknet & Opennet

# Why measuring Freenet

- Freenet provides anonymity and censorship resilience
- Darknet examined
- Opennet analyzed only after first release
- Characteristics of current Opennet not known
  - Ongoing development
  - New release every 3 - 4 months

# Outline

1. Background
2. **Goal**
3. Measurement Setup
4. Results
  1. Network Size & Participants
  2. Churn
  3. File Requests
5. Conclusion & Outlook

# Open Questions



Who is using Freenet?

How dynamic is set of members in the network?

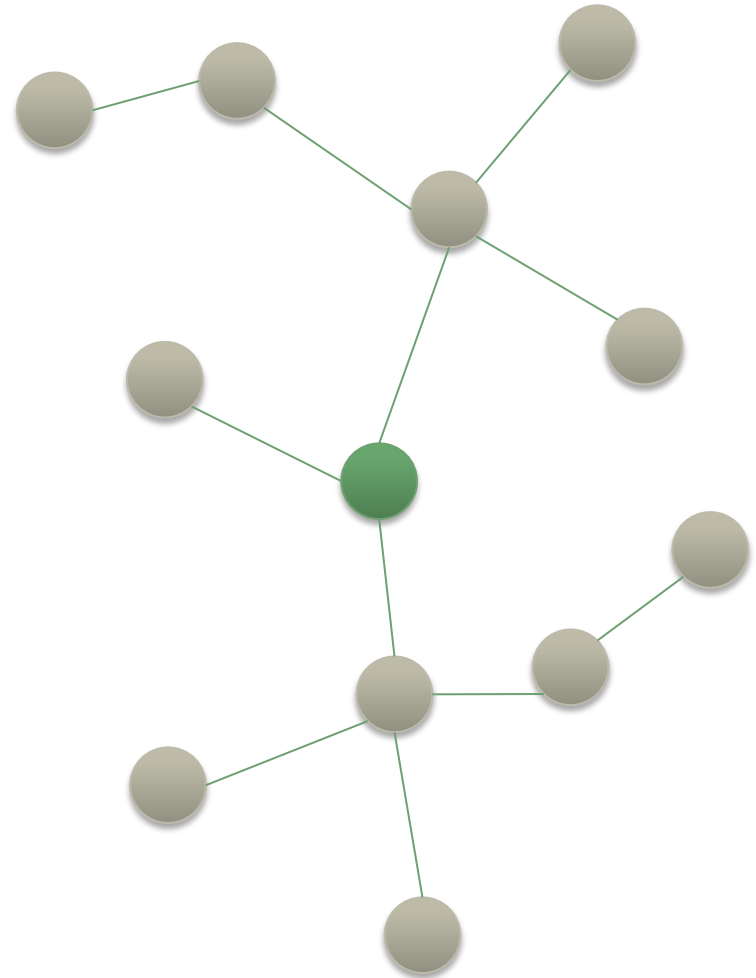
What structure does the network have?

What kind of content is shared?

How are Opennet and Darknet connected?

# Considerations

- Node represents user
  - IP address
  - Location
  - Etc...
- Node has limited view
- Detailed information about direct neighbors

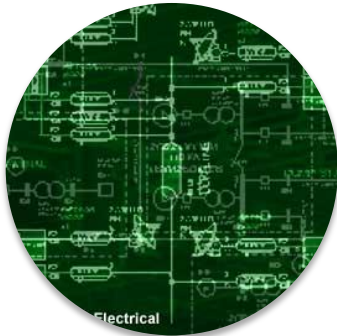




# Outline

1. Background
2. Goal
3. **Measurement Setup**
4. Results
  1. Network Size & Participants
  2. Churn
  3. File Requests
5. Conclusion & Outlook

# Measurement Methodology



## Instrument nodes

- Connect to Opennet
- Different Locations



## Passive Logging

- All incoming and outgoing messages
- Location and IP updates
- Discovery of nodes



## Active probing

- Ping messages to target locations
- Probe messages

Measurement Duration: 2 Months

# Setup

- 5 machines available
  - 4 machines: 5GB RAM, Dual core 2.0 GHz
  - 1 server: shared with others
- Node needs resources for
  - Basic functioning
  - Active probing
  - Maintaining many connections
  - Copying data
- 11 nodes on each machine → 55 nodes

# Measurement Areas

## Network Size & Participants

- Unique locations & IP addresses
- IP characteristics, e.g. Provider & Geolocation

## Churn/Online time

- Global metrics
- Per-peer metrics
- Correlation between metrics

## Topology

- Discovered fraction
- Link distribution, Clustering

## File Requests

- File transfer overhead
- File popularity

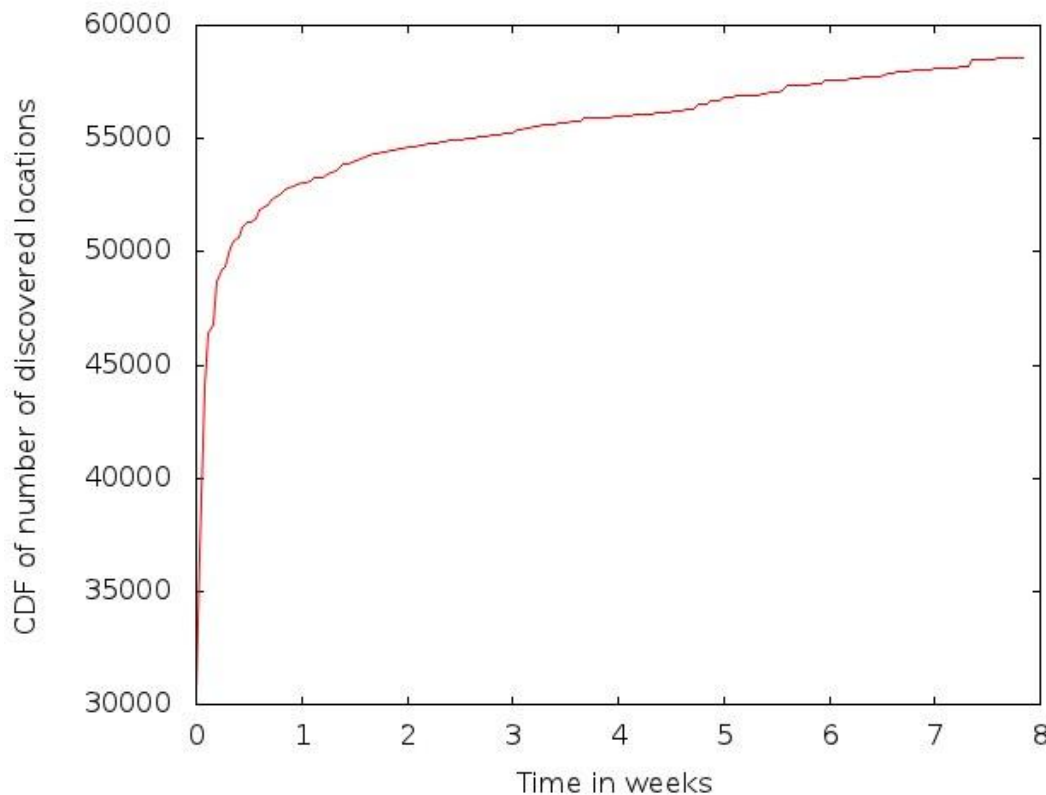
## Darknet

- Bridge node behavior

# Outline

1. Background
2. Goal
3. Measurement Setup
4. Results
  1. **Network Size & Participants**
  2. Churn
  3. File Requests
5. Conclusion & Outlook

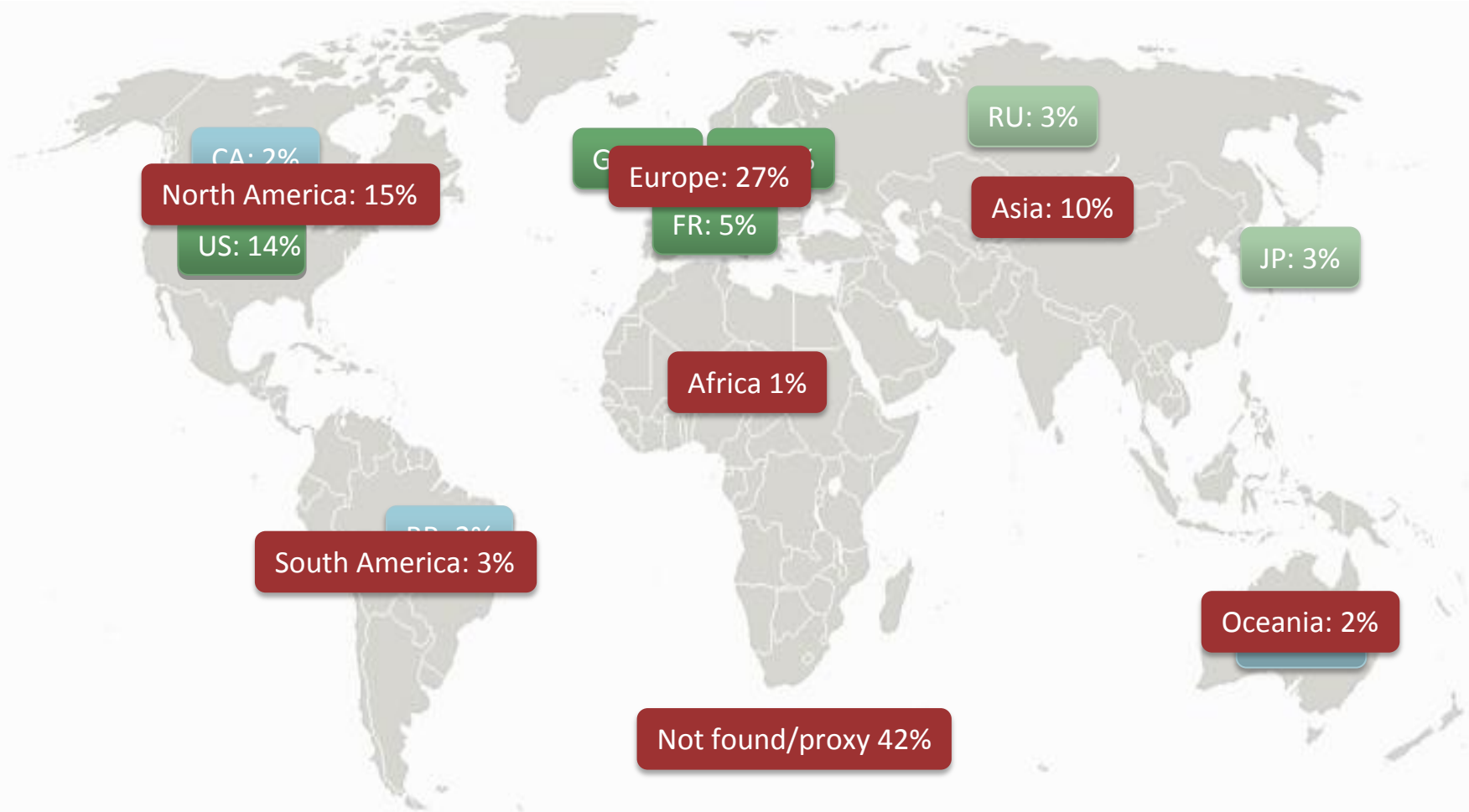
# Number of nodes



- In total: 58, 571 unique locations
- Number of unique locations corresponds to number of nodes
- Not all of them could be found again after discovery

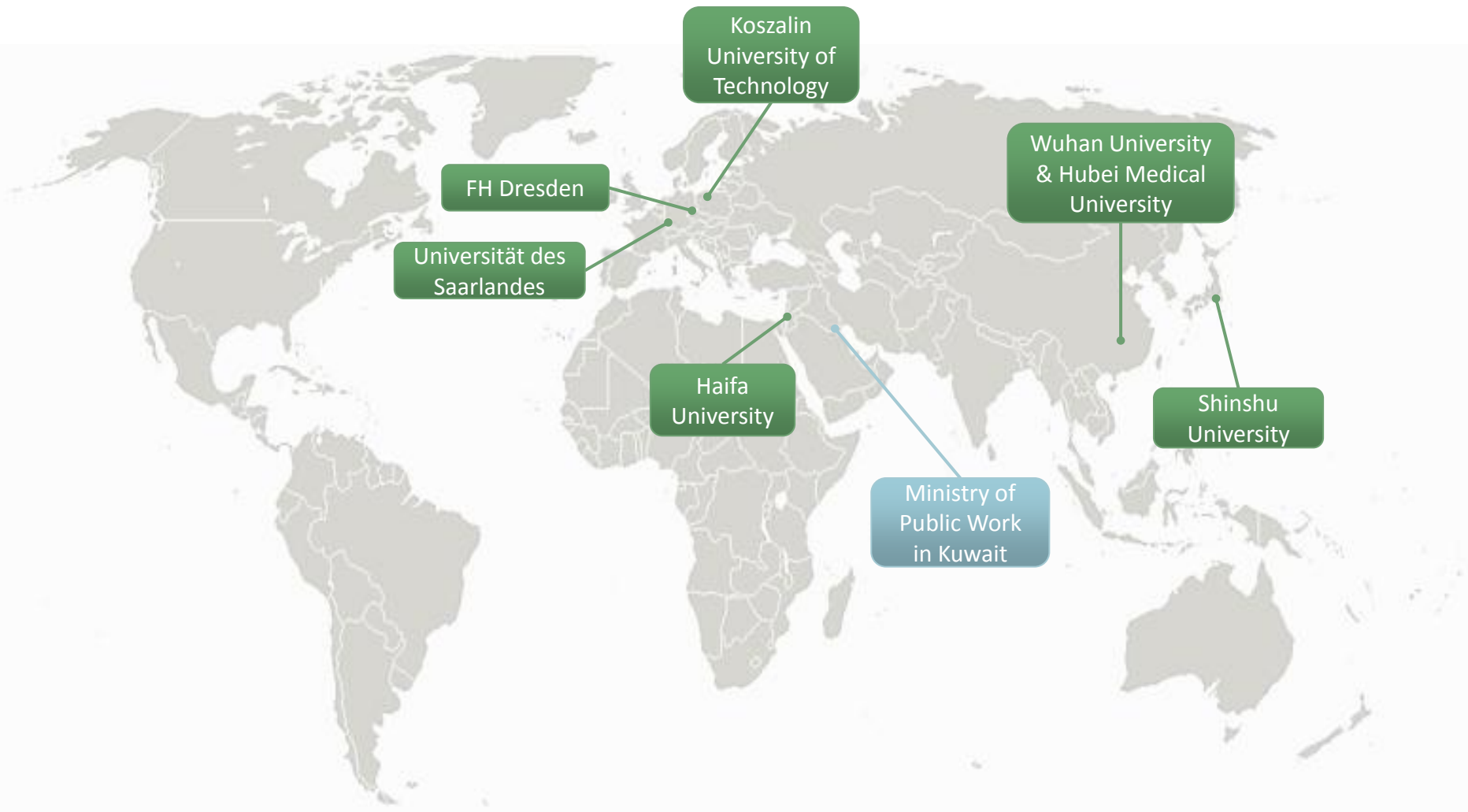
# Network members

## Geolocation



# Network members

## Universities and Institutions

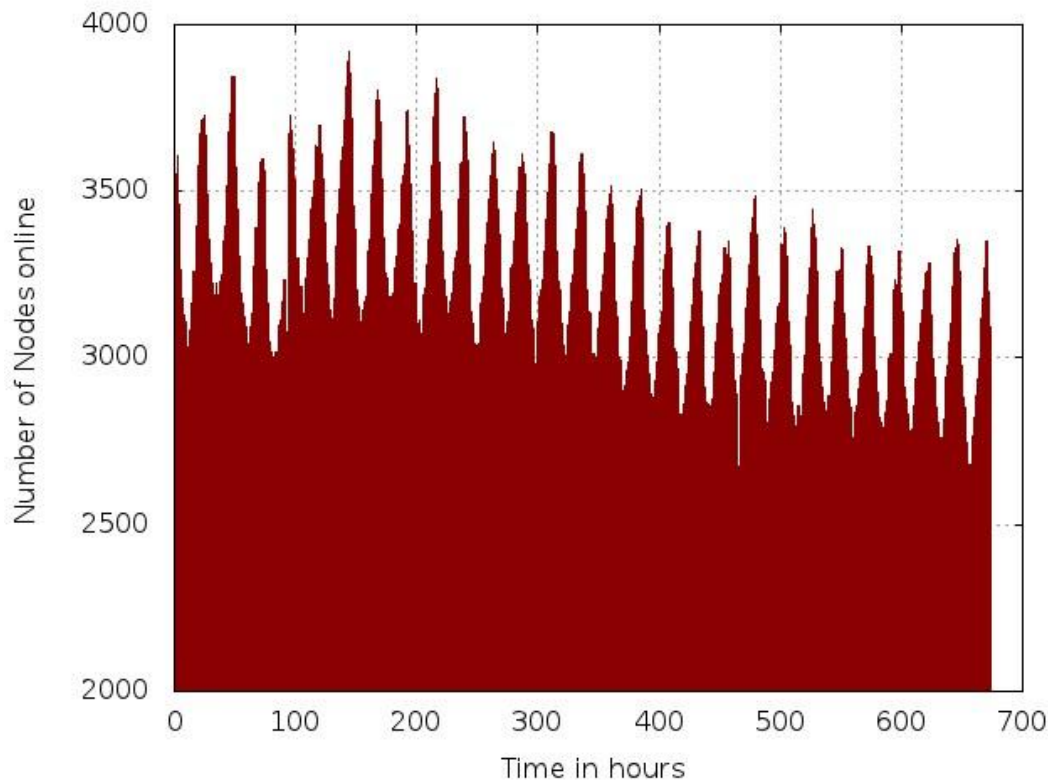




# Outline

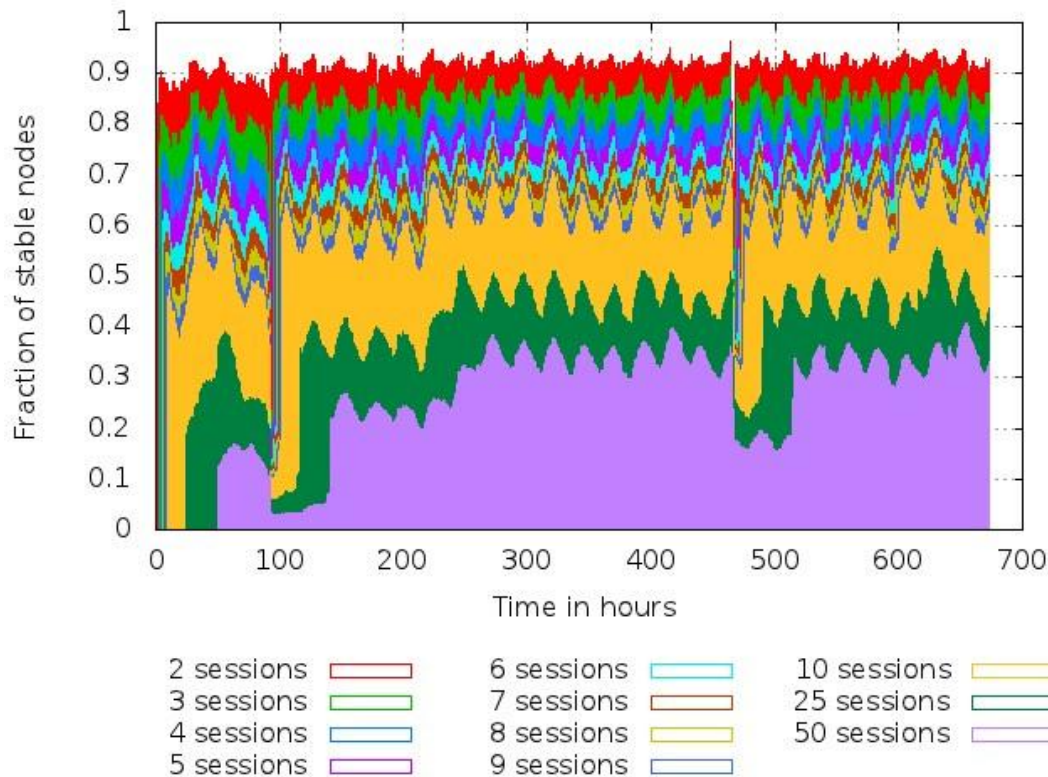
1. Background
2. Goal
3. Measurement Setup
4. Results
  1. Network Size & Participants
  2. **Churn**
  3. File Requests
5. Conclusion & Outlook

# Online nodes



- 15,000 target locations
- Ping message once per hour
- Daily pattern
- On average 3,207 nodes found -> 20.7%
  - 58,571 locations discovered
  - Estimation of ~ 12,120 nodes online each hour

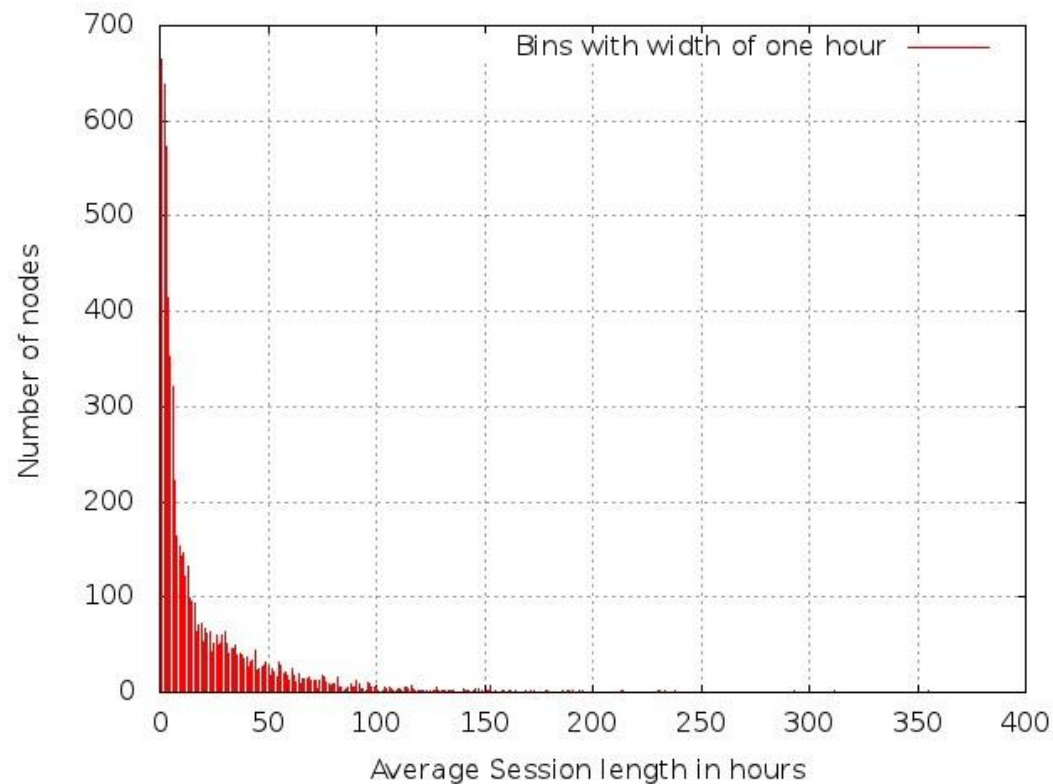
# Churn rate



- Stable set of nodes for various time frames
- Churn rates lower than in other P2P networks
  - 0.1 per hour
  - 0.71 in 25 hours
  - Gnutella: 0.016 per minute
  - PPLive: 10-180 peers per minute

# Average session length

- Average session length is 30.3 hours
  - 8.8% higher than 50 hours
  - 0.44% higher than 150 hours
- Highest measured value is 357 hours



# Outline

1. Background
2. Goal
3. Measurement Setup
4. Results
  1. Network Size & Participants
  2. Churn
  3. **File Requests**
5. Conclusion & Outlook

# Administration Overhead

- Compare amount of messages sent for maintaining the overlay with file transfer messages
- About 37% of all messages belong to a file transfer

Message Types	Fraction
Data transfer	0.1897
Data requests, data replies	0.1776
File inserts	0.0071

# Most popular content

- 6.8 Mio keys observed
- 1000 most popular keys are requested in 30.3% of all requests
- Content attached to keys is inferred from inserter



# Outline

1. Background
2. Goal
3. Measurement Setup
4. Results
  1. Network Size & Participants
  2. Churn
  3. File Requests
5. **Conclusion & Outlook**



# Conclusion

## Almost 60,000 nodes discovered

- Approx. 12,000 users online on average
- Most users from Europe and America

## Node behavior different from other P2P networks

- Lower churn rate
- Higher session length

## File requests

- High overhead for overlay maintenance
- Mostly legal content popular

# Future Work

- Run measurements for longer period
  - More data
  - Seasonal changes
- Exploit new feature for probing nodes in latest version
- Use measurement results to develop improvements

Thank you!

**Questions?**