Building Dependable Peer-to-Peer Systems

Koen Vanthournout

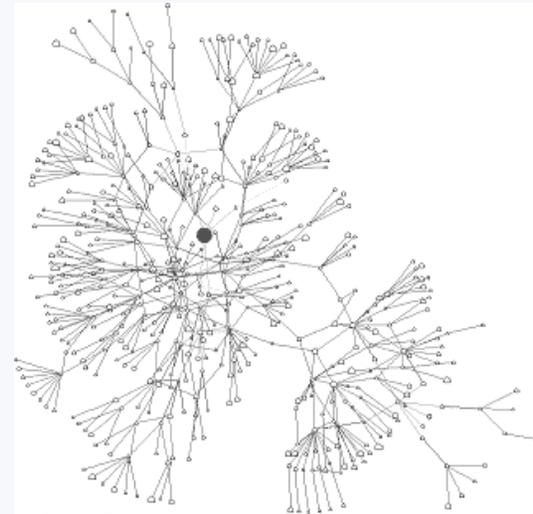Geert Deconinck

Ronnie Belmans

Introduction: P2P Evolution

- Origins: (pirate) filesharing:
  - o **Napster, Gnutella**
- Sound scientific basis:
  - o **CAN, P-Grid, Chord, Pastry, Tapestry, etc.**
- Filesharing → Resource discovery
- More widely applied:
  - o **grid applications, large networked automation, etc.**

→ **Dependability becomes an issue**

- Semantic routing network
- Properties:
  - No central coordinator
  - Self-organizing
  - All decissions based on local data only
- Components:
  - Overlay network
  - Query forwarding strategy
  - Application

- # Node failures
  - Fail-silent, crash semantics

- # Communication failures
  (TCP/IP style communication is assumed)
  - Temporarily unreachable nodes
  - Network partitions
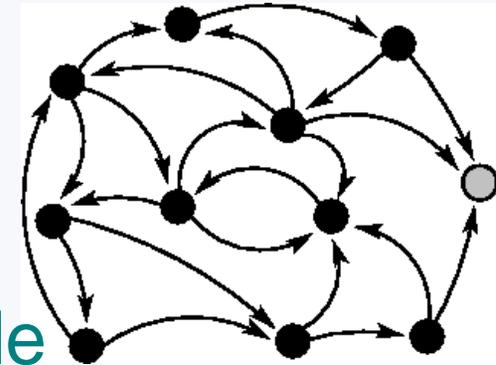
Failure Detection

- TCP/IP ⟹ Failure is detected if:
  - No communication channel can be set up
  - A communication channel is broken

- Network and node failures are undistinguishable
  - ↳ They must be treated equally

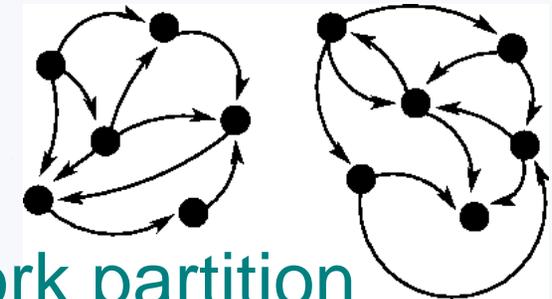- Network timeout is important

- ## Ghost nodes

  - Failed or unreachable node
  - Yet links still point to the node

- ## Network partitions

  - Communication network partition
  - When a large number of nodes fail
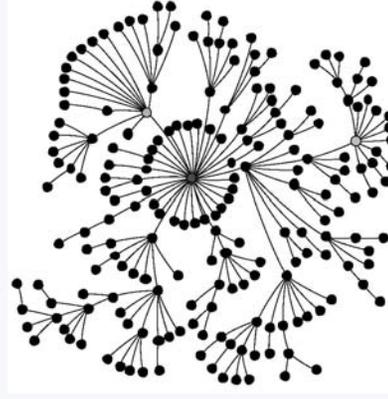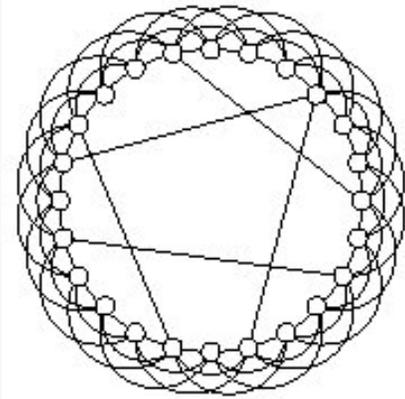
- What techniques allow:

  - Graceful degradation of the overlay

  - Spontaneous recovery after repair

- While exploiting as much as possible the existing self-organization algorithms

- Overlay desirable properties:
  - Regular
  - Small diameter

- Small-world & Scale-free models

- Highly tolerant to failures

(Albert & Barabási)

- P2P networks are highly dynamic
- Therefor, all nodes periodically reconverge
- Failures are changes
- Same mechanism allows for graceful degradation, posterior to failure detection
- But, speed vs. network load

Introduction

FailureTypes

Techniques

Test P2P

Results

Conclusion

- Every communication = failure detection

- Need for periodic communications
  (Upper limit to detection time)

- Dynamic P2P networks require periodic communications for update purposes

- Communication partitions + graceful degradation
- Simultaneous failure of large percentage of nodes

⟹ Overlay partitions
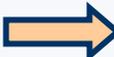
- Communication partitions + graceful degradation
- Simultaneous failure of large percentage of nodes

➡ Overlay partitions

- If no pointers => only manually repairable
- 'Deceased list':
  - Small FIFO in every node
  - Contains pointers to last detected failed nodes
  - Nodes periodically try to contact deceased nodes

➡ **Scalable**

Test P2P Network
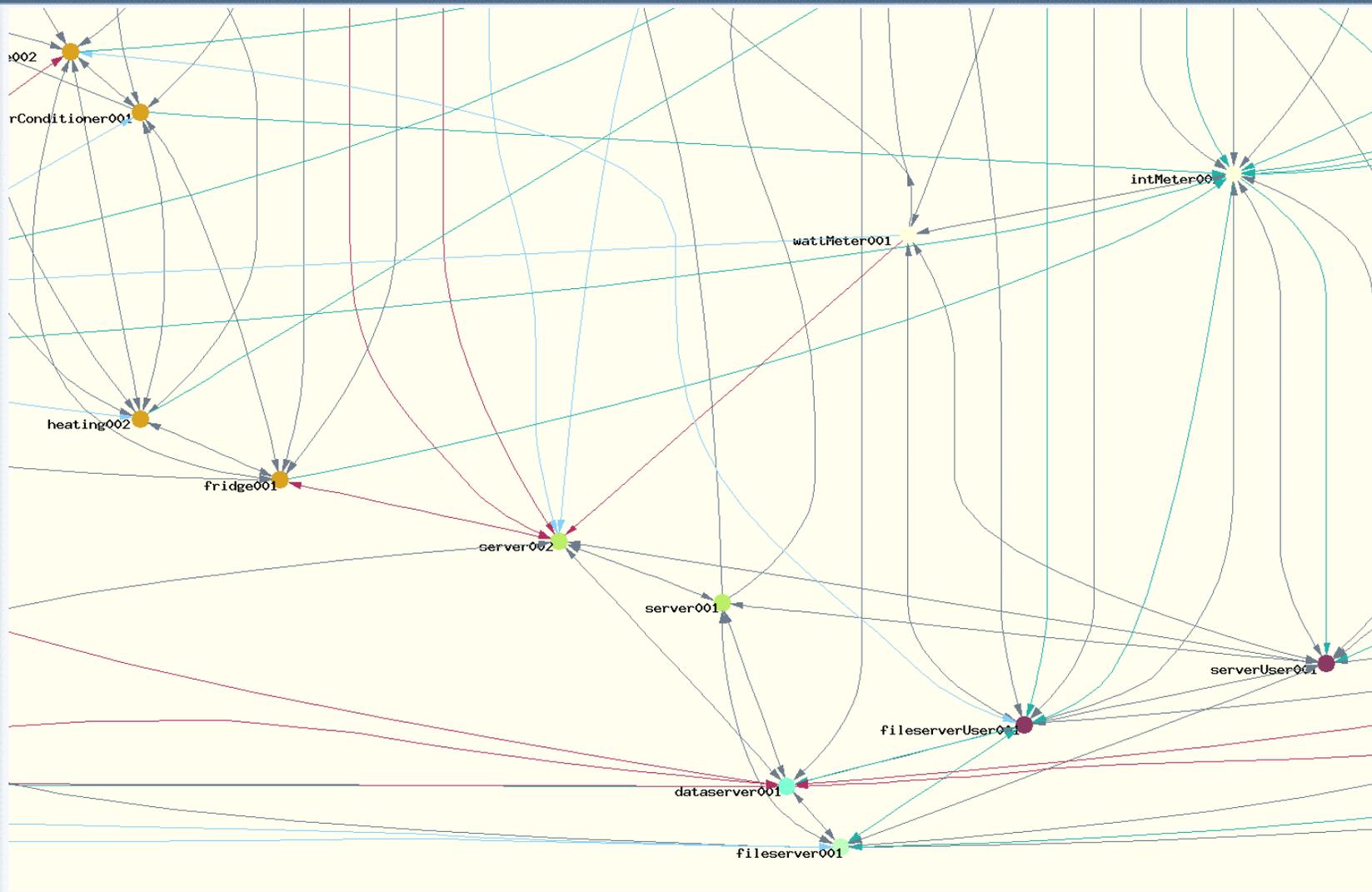
- Generic resource discovery network
- XML resource description files
- 'Distance' metric used for periodic self-organization
- Groups nodes by functionality and interests
- 50% chance for far link (small world)
- Periodic polls for XML file updates
- Deceased list of size 3

Test P2P Network

Introduction

FailureTypes

Techniques

Test P2P

Results

Conclusion

failure detection and restoration times (5%)

Introduction

FailureTypes
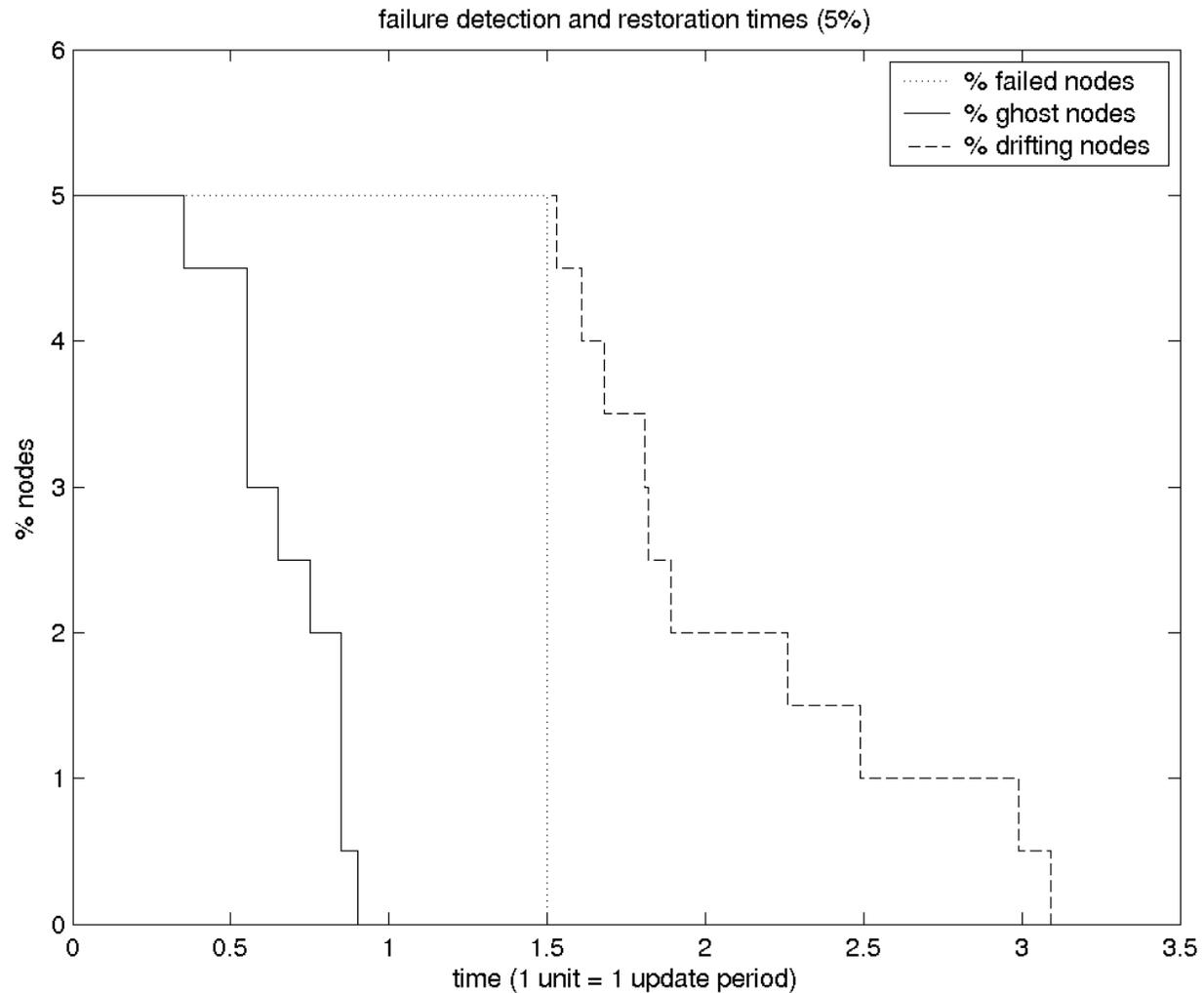
Techniques

Test P2P

Results

Conclusion

failure detection and restoration times (30%)
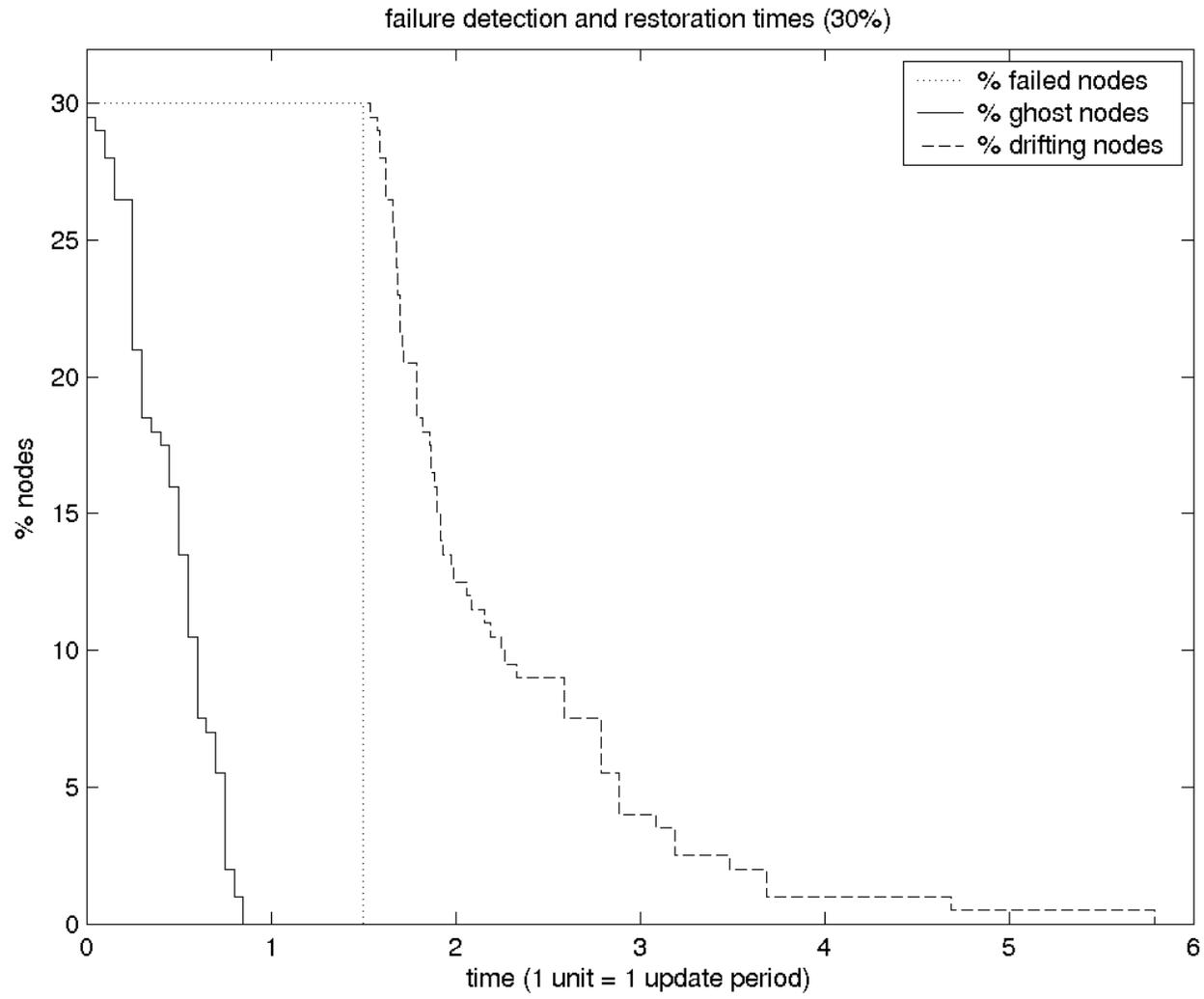
Introduction

FailureTypes

Techniques

Test P2P

Results

Conclusion

9 partitions
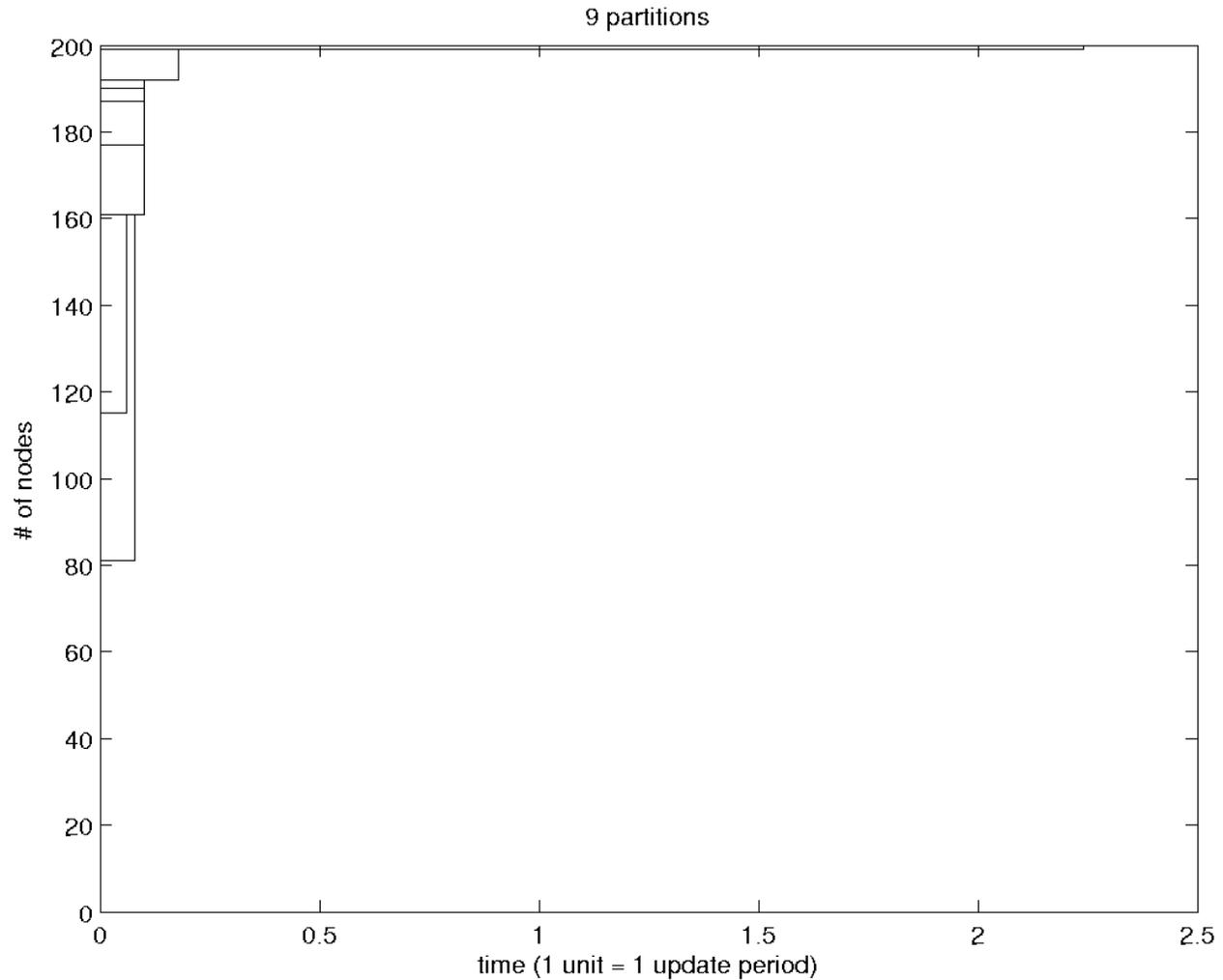
Conclusions

- ## P2P is more widely applicable
- ## Dependability becomes an issue

- ## Several techniques to improve the dependability have been proposed
- ## Simulations prove there effectiveness

- ## Future work:
  - ### Influence of failures on query forwarding
  - ### More simulations with larger numbers of nodes

Introduction

FailureTypes

Techniques

Test P2P

Results

Conclusion