Policies for Efficient Data Replication in P2P Systems

IEEE ICPADS'13

João Paiva, Luís Rodrigues

Instituto Superior Técnico / Inesc-ID, Lisboa, Portugal

December 18, 2013

Policies for **Efficient** Data Replication in P2P Systems

- Taking in mind the tradeoffs involved in policy design
- Previous approaches too single-minded for specific metrics

Policies for **Efficient** Data Replication in P2P Systems

- Taking in mind the tradeoffs involved in policy design
- Previous approaches too single-minded for specific metrics

Policies for **Efficient** Data Replication in P2P Systems

- Taking in mind the tradeoffs involved in policy design
- Previous approaches too single-minded for specific metrics

Metrics

We consider the following metrics:

1. Monitoring costs: "pinging" the owners of replicas
2. Data Transfer costs: copying data as nodes join or leave
3. Load Unbalance costs: per-node distribution of load

Neighbour Replication vs Most-Available Replication

Neighbour Replication (NR)

description: Each node replicates its data on its R neighbours

motivation: Simplicity and scalability

Most-Available Replication (MAR)

description: Data is placed on nodes predicted to be most reliabe

motivation: Data Transfer costs

Neighbour Replication vs Most-Available Replication

Neighbour Replication (NR)

description: Each node replicates its data on its R neighbours

motivation: Simplicity and scalability

Most-Available Replication (MAR)

description: Data is placed on nodes predicted to be most reliabe

motivation: Data Transfer costs

Neighbour Replication vs Most-Available Replication

Neighbour Replication (NR)

description: Each node replicates its data on its R neighbours

motivation: Simplicity and scalability

Most-Available Replication (MAR)

description: Data is placed on nodes predicted to be most reliabe

motivation: Data Transfer costs

Neighbour Replication vs Most-Available Replication

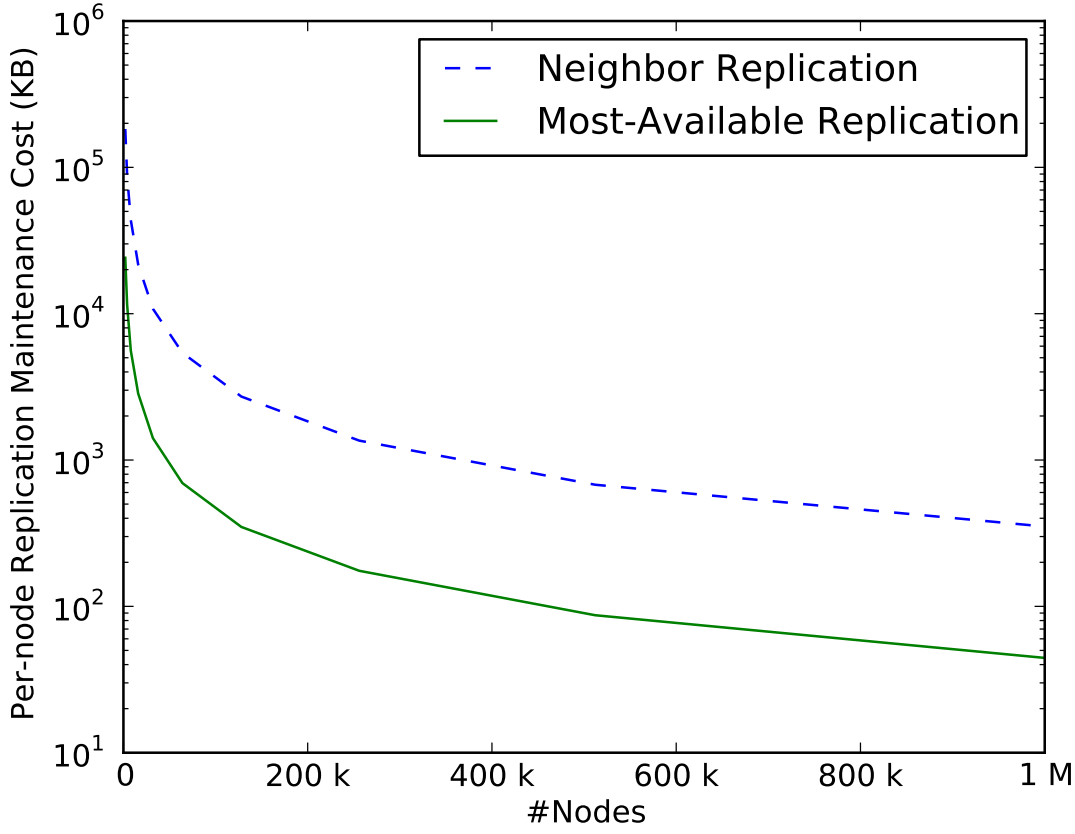


Figure : Data Transfer Costs

TÉCNICO LISBOA

inesc id lisboa

Neighbour Replication vs Most-Available Replication

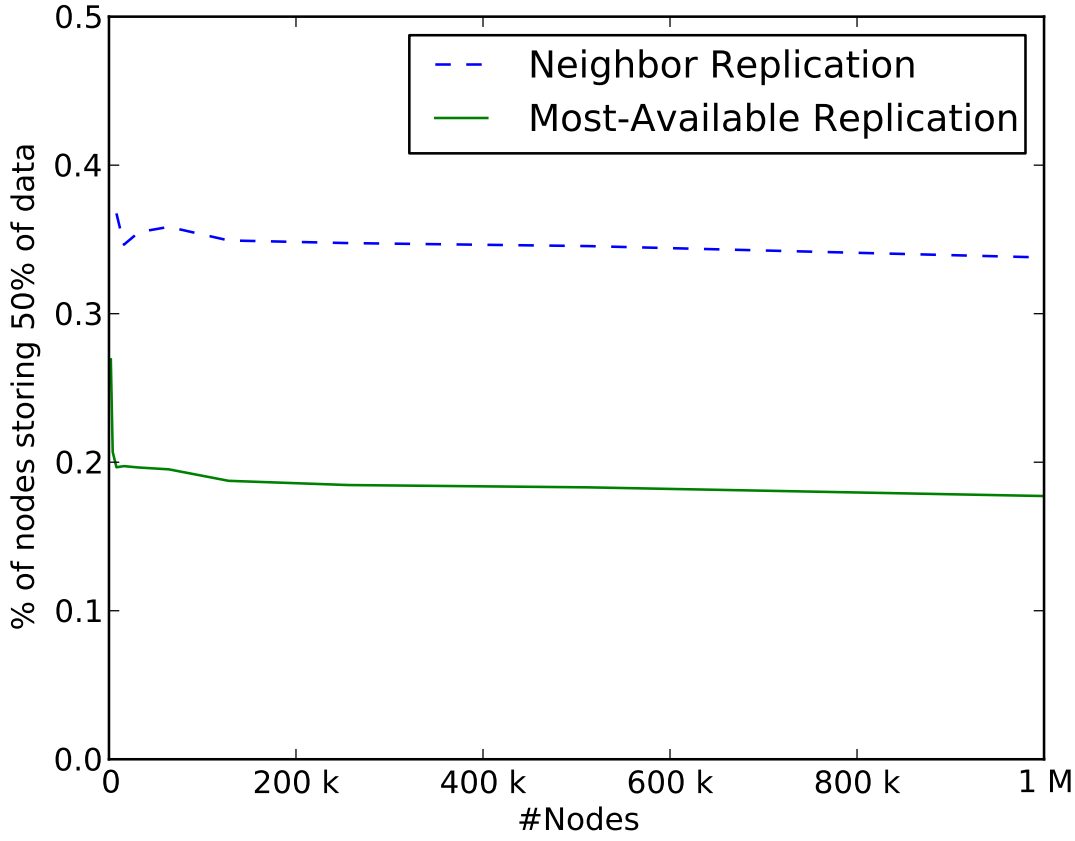


Figure : Load Unbalance Costs

TÉCNICO LISBOA

inesc id lisboa

Contributions

1. Catalog of existing solutions according to new taxonomy
2. New replication policies
3. Performance model for evaluating metrics

Outline

Policy Classification

1. **State Information**
   - Oblivious: No information on state of each peer (e.g. N.R.)
   - Informed: Depends on peer state information (e.g. M.A.R)

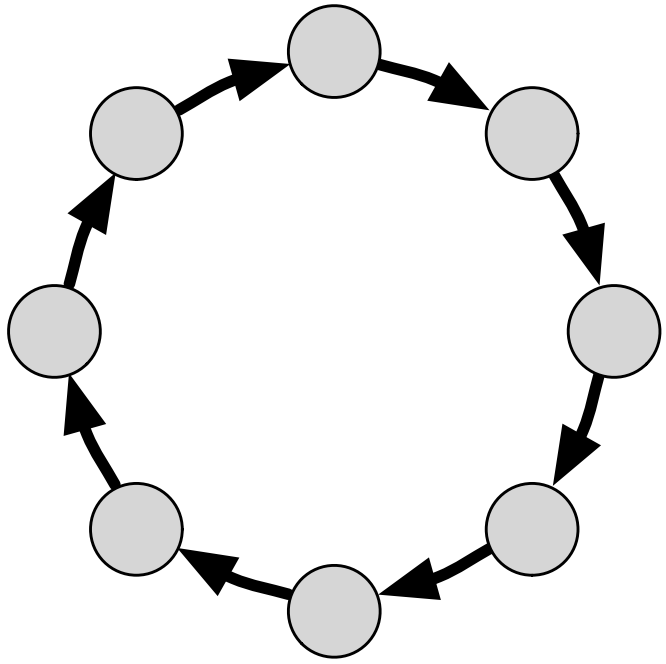
2. **Replica Placement**
   - Consistent Hashing: No control on data placement (e.g. N.R.)
   - Directory-Based: Precise control over data placement (e.g. M.A.R.)

3. **Topology**
   - Plain: Single identity per node (e.g. N.R, M.A.R)
   - Virtual Servers: Several identities per node
   - Logical Groups: (introduced next)

Logical Groups

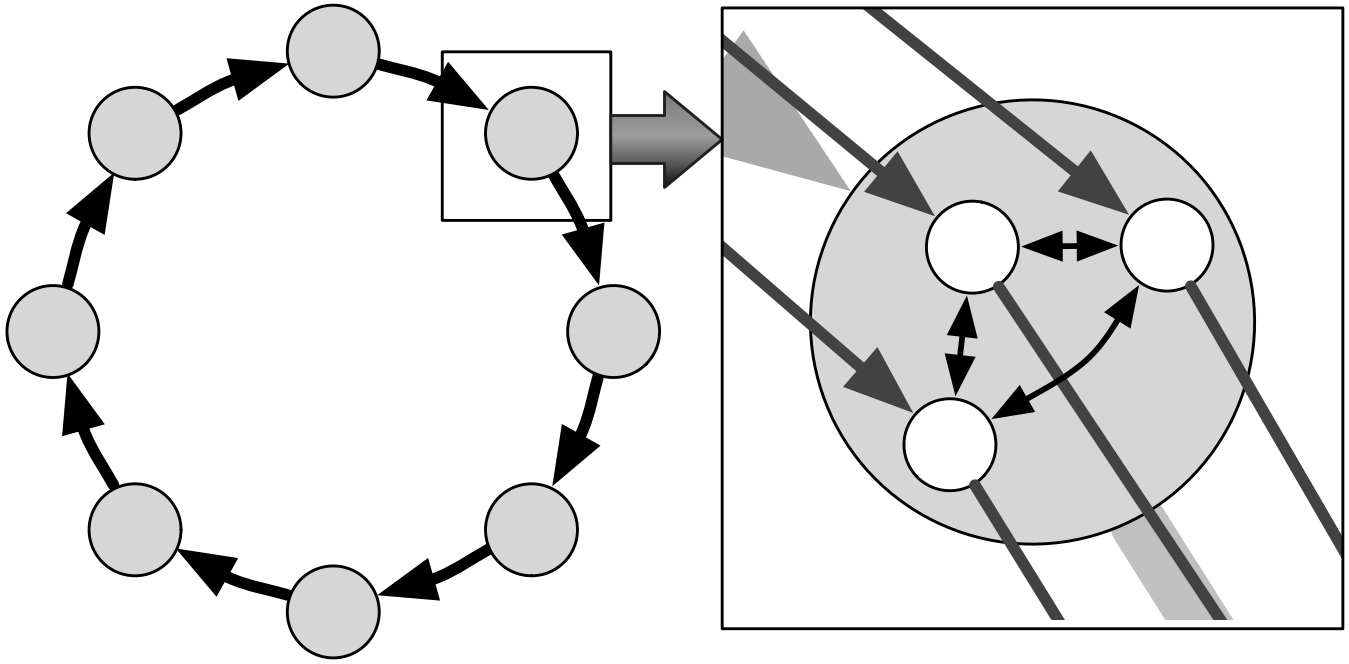
- Self-contained replication groups
- Act as single nodes in DHT
- No pre-defined node positions in network

Logical Groups



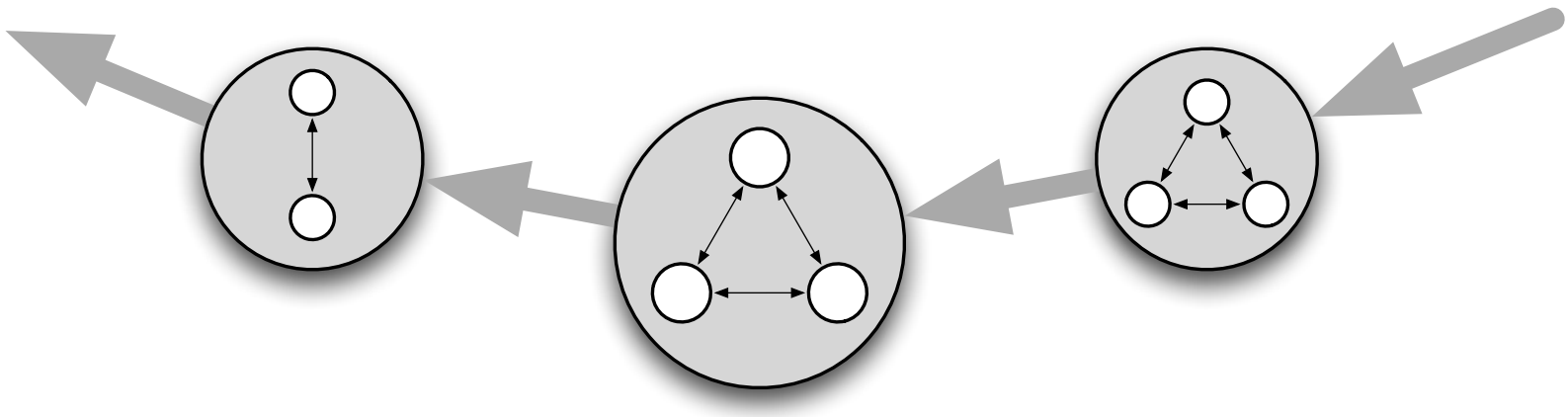
- Ring-based overlay: Composed of logical group of nodes

Logical Groups

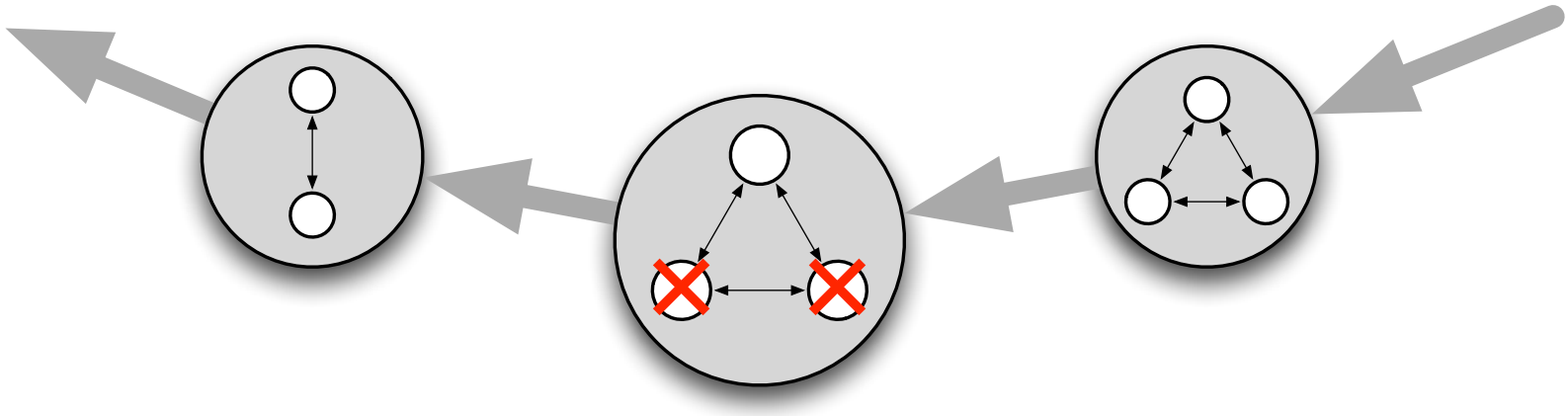


- Ring-based overlay: Composed of logical group of nodes

Logical Groups: Dynamic topology



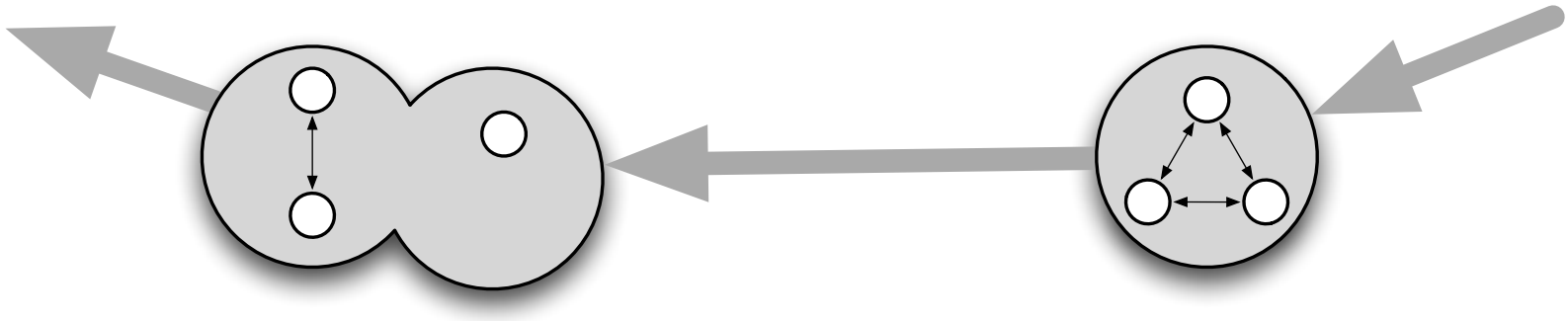
Logical Groups: Dynamic topology



TÉCNICO LISBOA

inesc id lisboa

Logical Groups: Dynamic topology



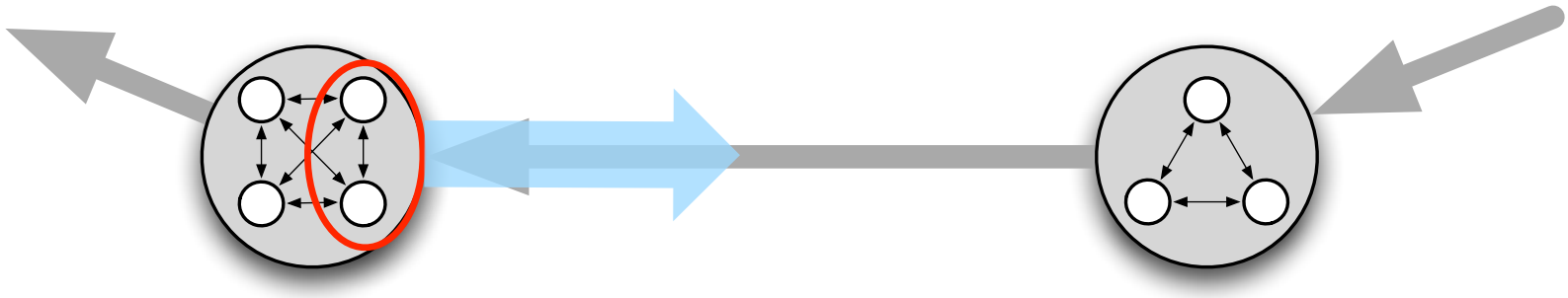
TÉCNICO LISBOA

inesc id lisboa

Logical Groups: Dynamic topology

Logical Groups: Dynamic topology

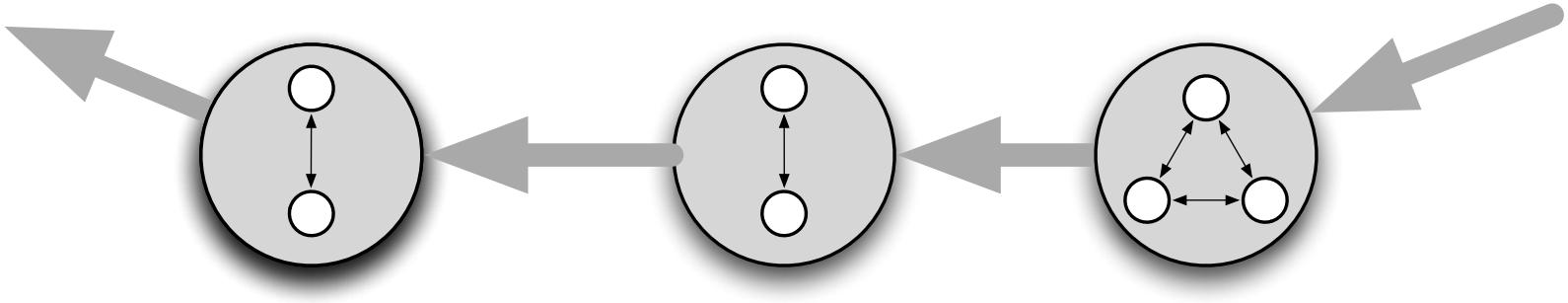
Logical Groups: Dynamic topology



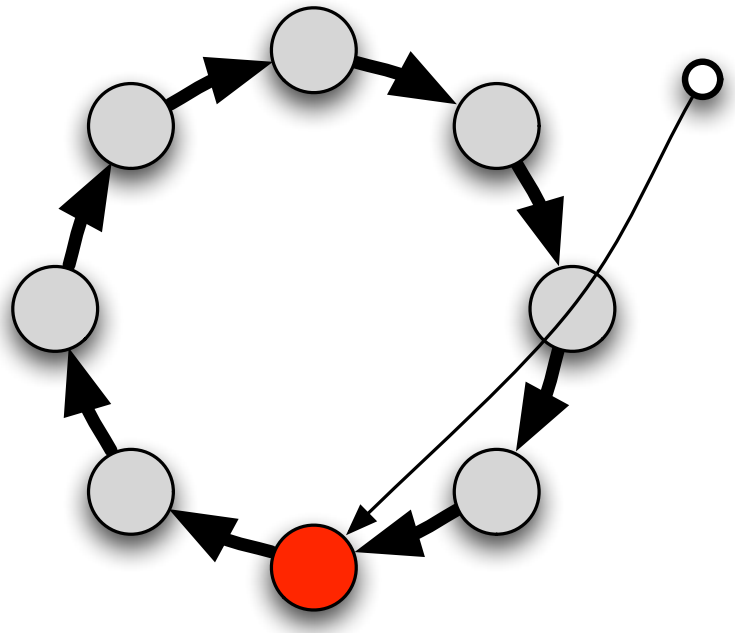
TÉCNICO LISBOA

inesc id lisboa

Logical Groups: Dynamic topology

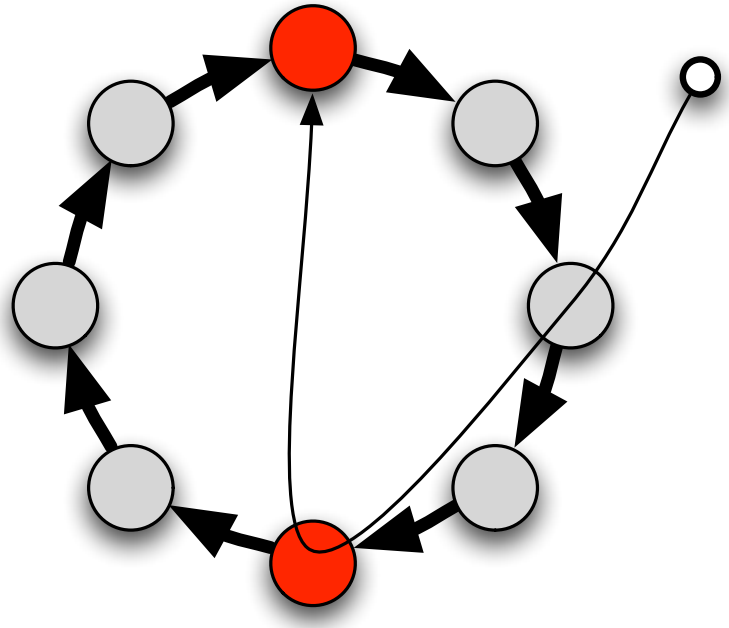


Logical Groups: node join



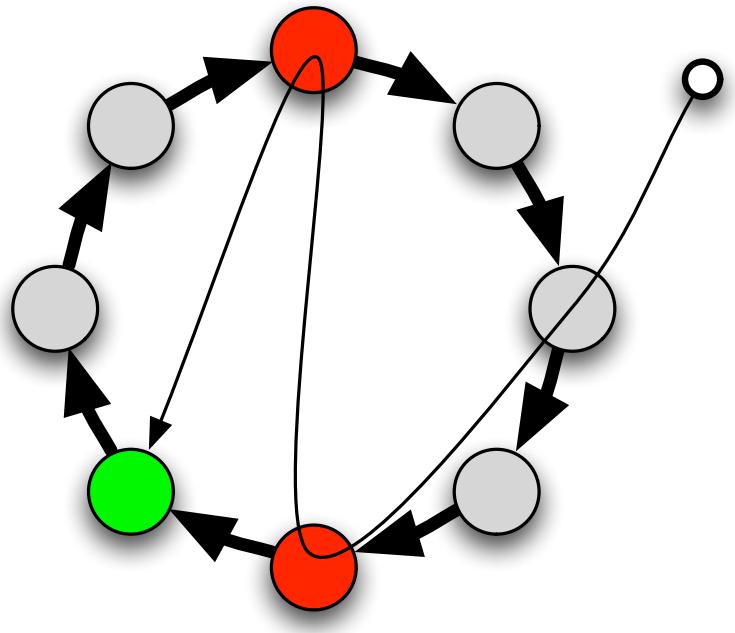
- Nodes can select where to join

Logical Groups: node join



- Nodes can select where to join

Logical Groups: node join



- Nodes can select where to join

Logical Groups: R-LB Policy

Resilient Load Balancing (R-LB) policy

description: Small groups are merged with neighbour groups; new nodes joing groups with highest load; when groups are split, per-group load is maintained

motivation: Resilience and Load-Balancing

Policy Map

		primary performance target			
		none	**monitoring**	**load balancing**	**bandwidth**
Oblivious	Plain		Neighbour Replication	Multi-Publication	RelaxDHT
	VServers			Virtual Servers	
	Groups				
Informed	Plain				Most-available, Regularity-based
	Groups			R-LB	

Policy Map

		primary performance target			
		none	**monitoring**	**load balancing**	**bandwidth**
Oblivious	Plain		Neighbour Replication	Multi-Publication	RelaxDHT
	VServers			Virtual Servers	
	Groups				
Informed	Plain				Most-available, Regularity-based
	Groups			R-LB	

Outline

Policy Map

		primary performance target			
		none	**monitoring**	**load balancing**	**bandwidth**
Oblivious	Plain		Neighbour Replication	Multi-Publication	RelaxDHT
	VServers			Virtual Servers	
	Groups				Supersize-me
Informed	Plain				Most-available, Regularity-based
	Groups			R-LB	

Supersize-Me

Supersize-Me policy

description: Avoid merging groups: Larger Groups; Nodes join on smaller groups

motivation: Data Transfer costs

Policy Map

		primary performance target			
		none	**monitoring**	**load balancing**	**bandwidth**
Oblivious	Plain		Neighbour Replication	Multi-Publication	RelaxDHT
	VServers			Virtual Servers	
	Groups				Supersize-me
Informed	Plain				Most-available, Regularity-based
	Groups		Avoid-Surplus	R-LB	

TÉCNICO LISBOA

inesc id lisboa

Avoid-Surplus

Avoid-Surplus policy

description: Reduce maintenance costs: Smaller groups; Nodes join on larger groups

motivation: Monitoring costs

Policy Map

		primary performance target			
		none	**monitoring**	**load balancing**	**bandwidth**
Oblivious	Plain		Neighbour Replication	Multi-Publication	RelaxDHT
	VServers			Virtual Servers	
	Groups				Supersize-me
Informed	Plain				Most-available, Regularity-based
	Groups		Avoid-Surplus	R-LB	
			Hotter-On-Ephemeral		

Hotter-on-Ephemeral (HonE)

HonE policy

description: Load Balancing of R-LB + Bandwidth Usage of M.A.R.: Most unreliable nodes (ephemeral) in groups with fewer objects (hotter).

motivation: Monitoring costs, Load-Balancing, Data Transfer costs

Policy Map

		primary performance target			
		none	**monitoring**	**load balancing**	**bandwidth**
Oblivious	Plain		Neighbour Replication	Multi-Publication	RelaxDHT
	VServers			Virtual Servers	
	Groups				Supersize-me
Informed	Plain				Most-available, Regularity-based
	Groups	Random	Avoid-Surplus	R-LB	Preemptive replacement
			Hotter-On-Ephemeral		

TÉCNICO LISBOA

inesc id lisboa

Outline

Performance Model

Idealized system as baseline for comparison

- Perfect Load Balancing
- All nodes with same probability of failing
- All load preserved after failures and joins

Experimental Parameters

- Real trace of Peer-To-Peer
- 1 million unique peers
- 100.000 objects
- Zipf distribution of load
- Replication degree 6

Experimental Results

	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

TÉCNICO LISBOA

inesc id lisboa

Experimental Results

	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

TÉCNICO LISBOA

inesc id lisboa

Experimental Results

	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

TÉCNICO LISBOA

inesc id lisboa

Experimental Results

	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

TÉCNICO LISBOA

inesc id lisboa

Experimental Results

	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

TÉCNICO LISBOA

inesc id lisboa

Experimental Results

	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

TÉCNICO LISBOA

inesc id lisboa

Experimental Results

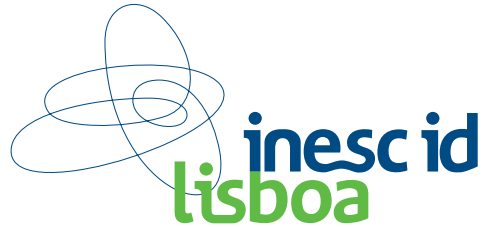
	Monitoring	Data Transfer	Load Unbalance
Idealized	1.00	1.00	1.00
Neighbor Rep.	1.00	1.00	1774.1
Most-available	0.07	0.13	2365.5
R-LB	0.71	0.52	1.1
Avoid Surplus	0.76	0.41	308.5
Supersize-me	1.07	0.79	1.1
HonE	0.61	0.28	1.1

Outline

Conclusions

1. Catalog of existing solutions according to new taxonomy
2. New policies based on group topologies
3. New policy with particularly interesting trafeoffs
4. Performance model for evaluating policies

Thank you



Behaviour of HonE over time

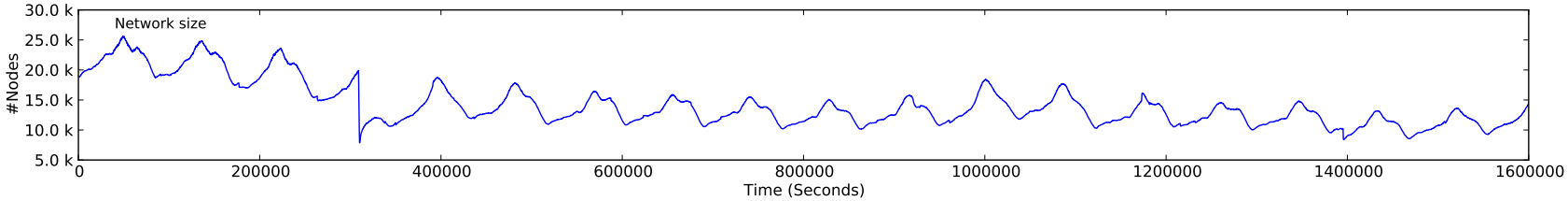


Figure : Network size over time.

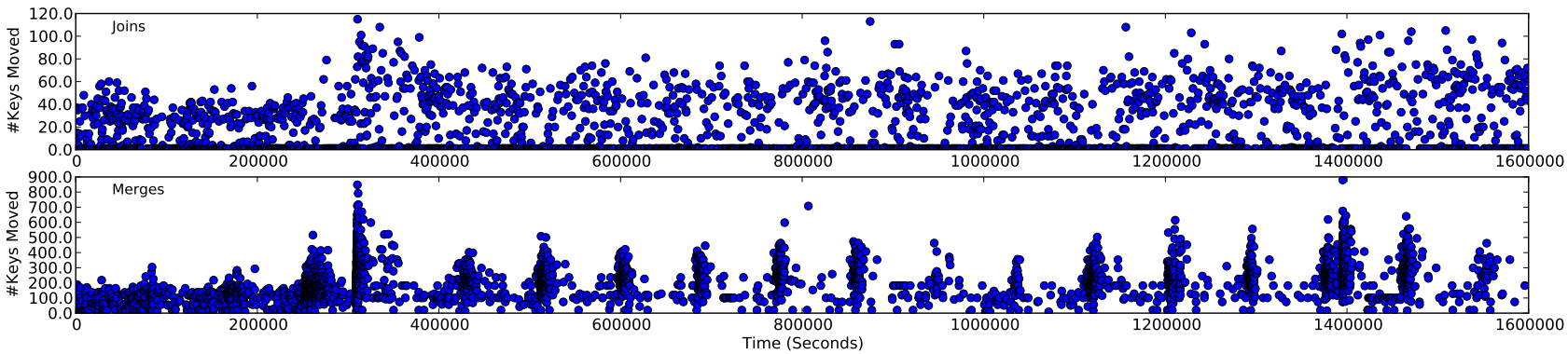


Figure : Behaviour of HonE over time.

TÉCNICO LISBOA

inesc id lisboa

Behaviour of R-LB over time

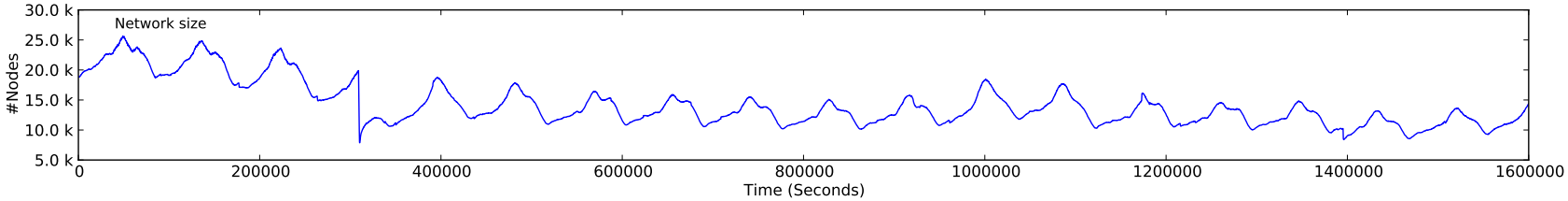


Figure : Network size over time.

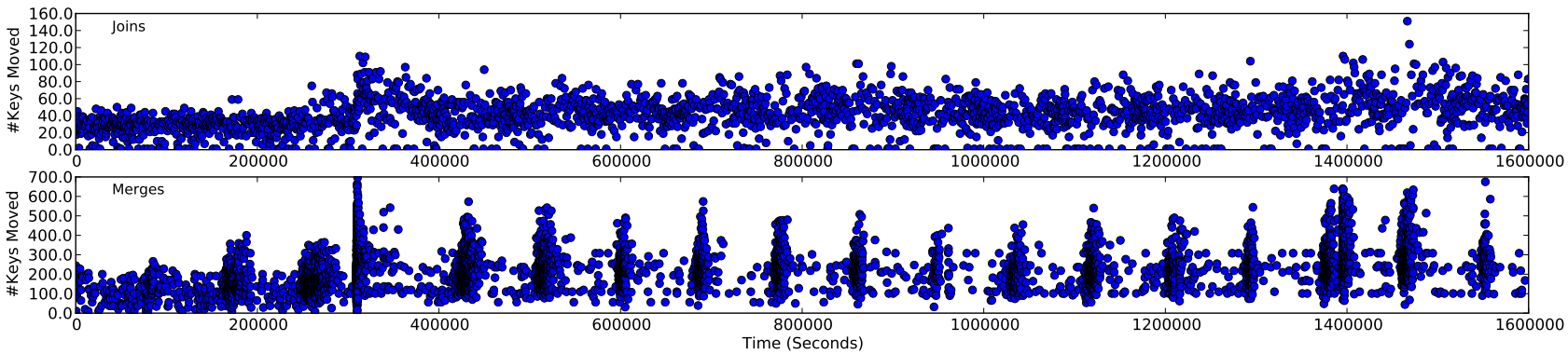


Figure : Behaviour of R-LB over time.

TÉCNICO LISBOA

inesc id lisboa

Behaviour of Supersize-me over time

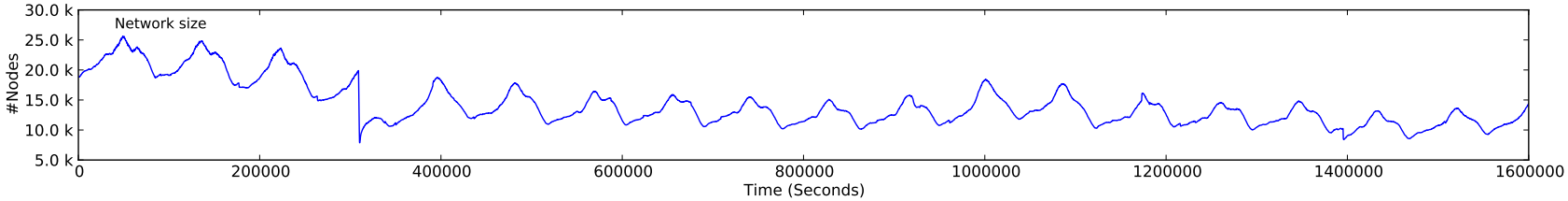


Figure : Network size over time.

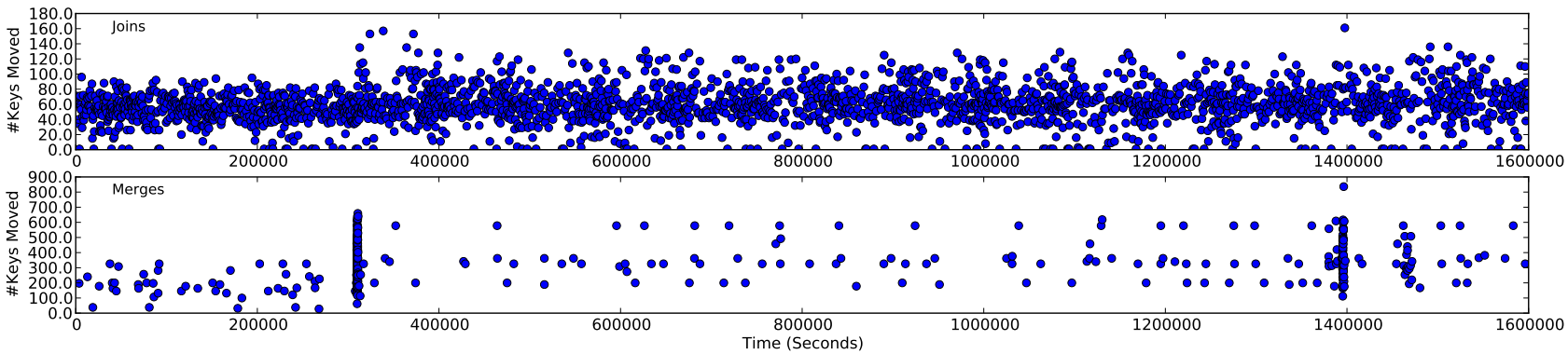


Figure : Behaviour of Supersize-me over time.

TÉCNICO LISBOA

inesc id lisboa