EPFL

A Control Plane in Time and Space for Locality-Preserving Blockchains

Arnaud Pannatier
Master Thesis

***Pr. Bryan Ford**, Advisor*
***Pr. Pawel Szalachowski**, Expert*
***Cristina Basescu**, Supervisor*

Lausanne - 11.02.20

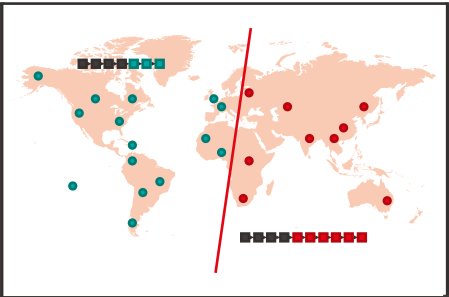
École
polytechnique
fédérale de
Lausanne

Outline

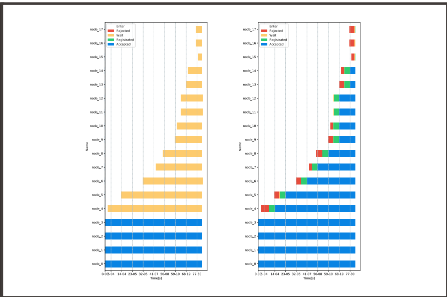
1. Some problems of traditional blockchains

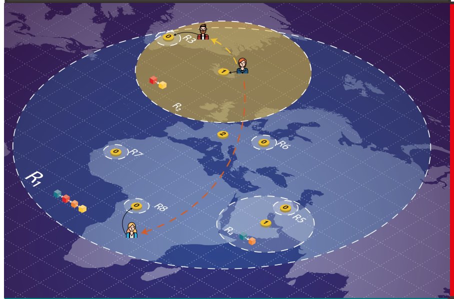
WWIII Scenarios

Time for validation



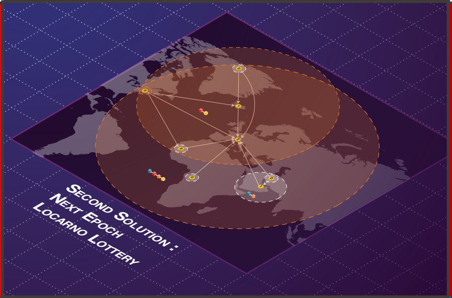
4. Results





2. Context: Nyle

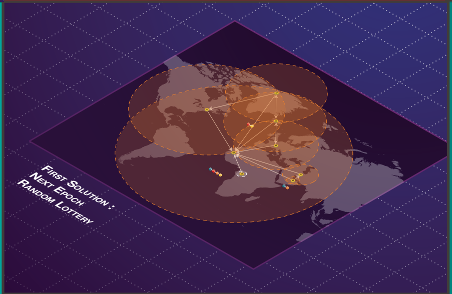
Using region replication to defeat the problems



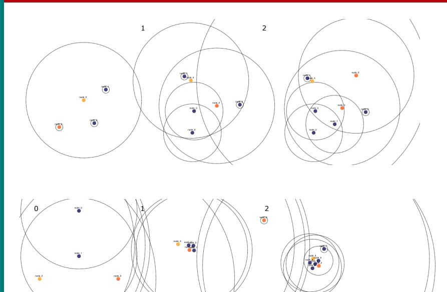
5. Improvements

3. My work

Adapt the regions to node modifications

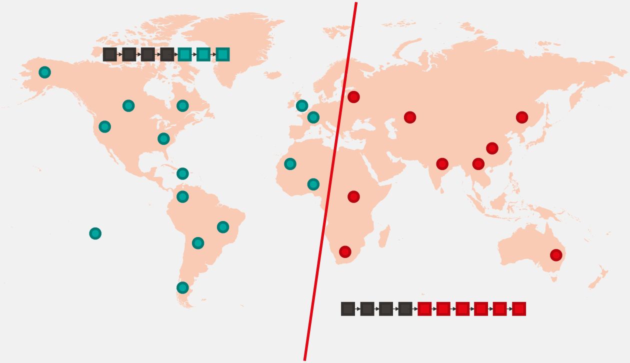


6. Conclusion

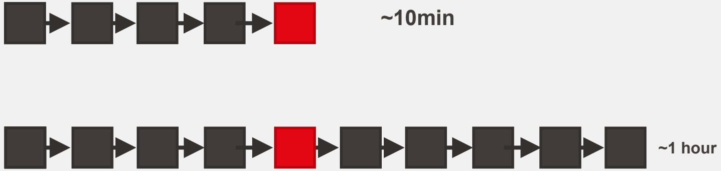


Problems of traditional blockchains

World War III Scenarios

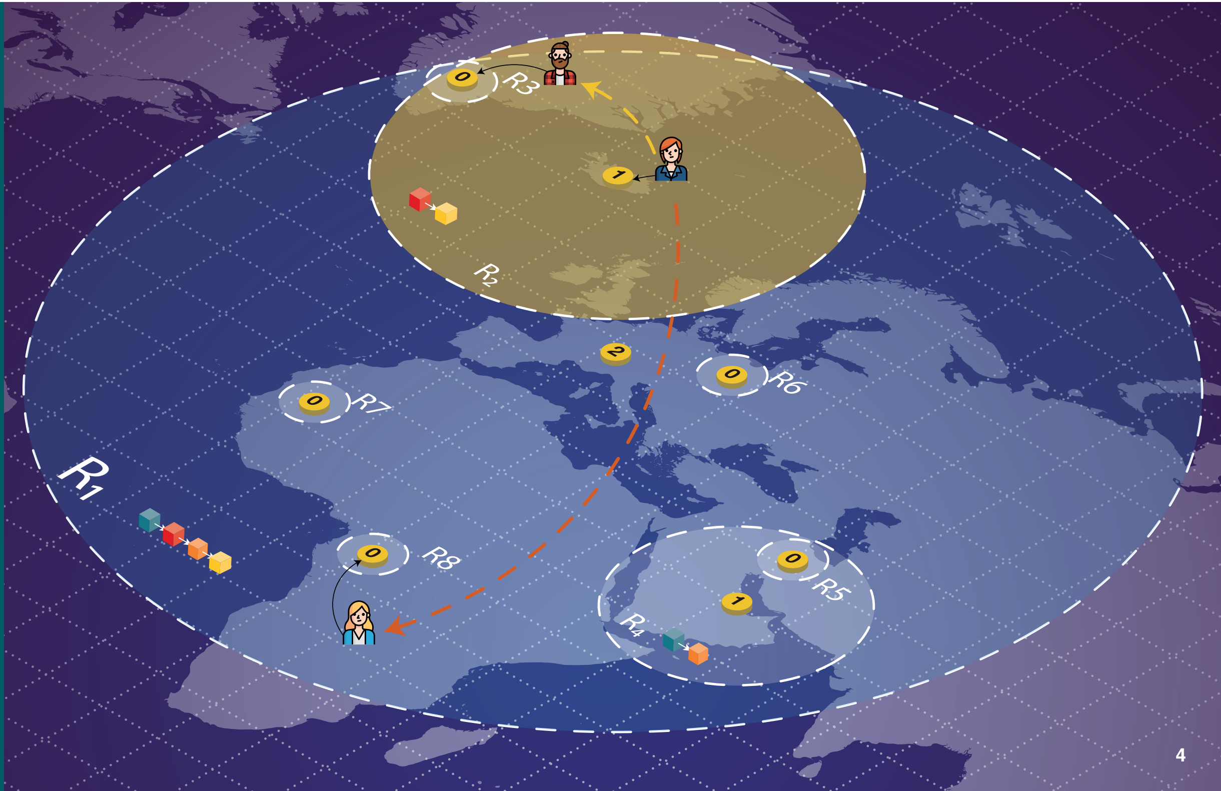


Time for validation



Context : Nyle

- Enhances blockchains with locality
- The system replicated in *regions*
- The worst case latency for any pair of nodes is a small multiple of their network latency (RTT)



Context : Nyle

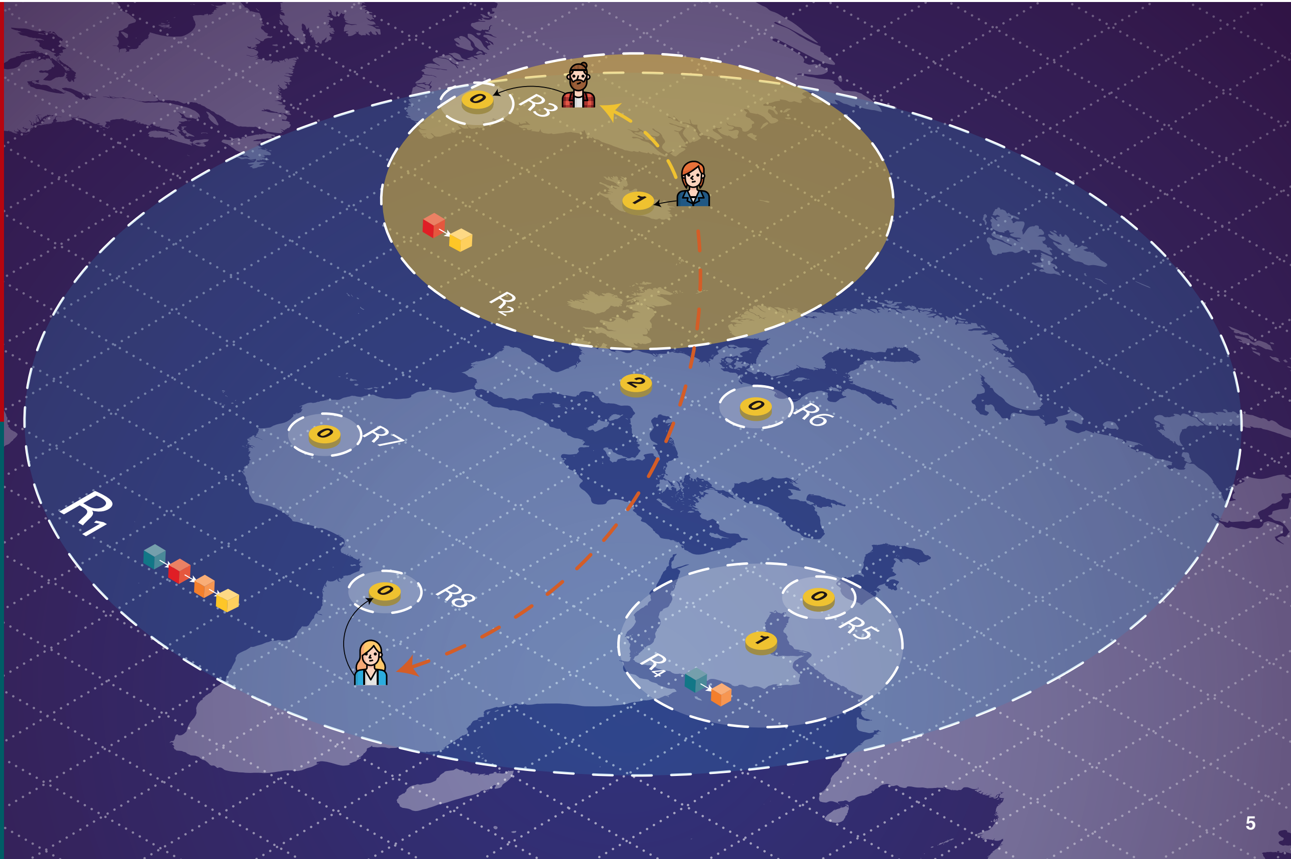
Replicates the system in regions, from local to global

World War III Scenarios

If a global partition occurs, the system still works in regions that are not split by a partition

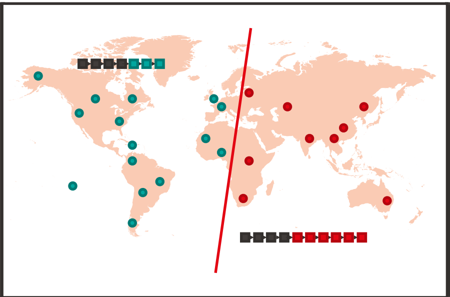
Time for validation

Transactions can be validated in regions

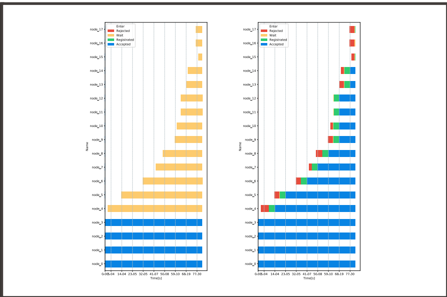


Outline

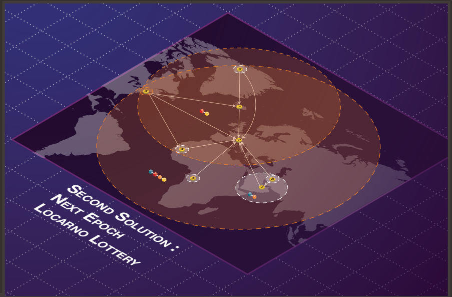
1. Some problems of traditional blockchains
WWIII Scenarios
Time for validation



4. Results

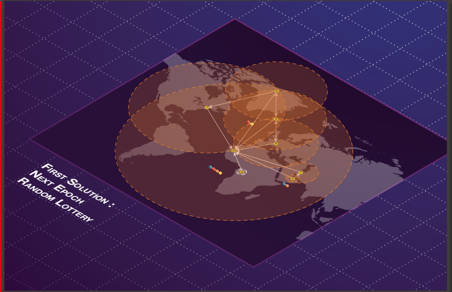


2. Context: Nyle
Using region replication to defeat the problems

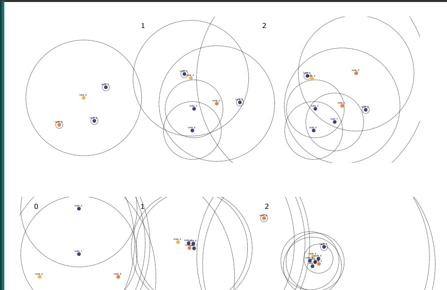


5. Improvements

3. My work
Adapt the regions to node modifications



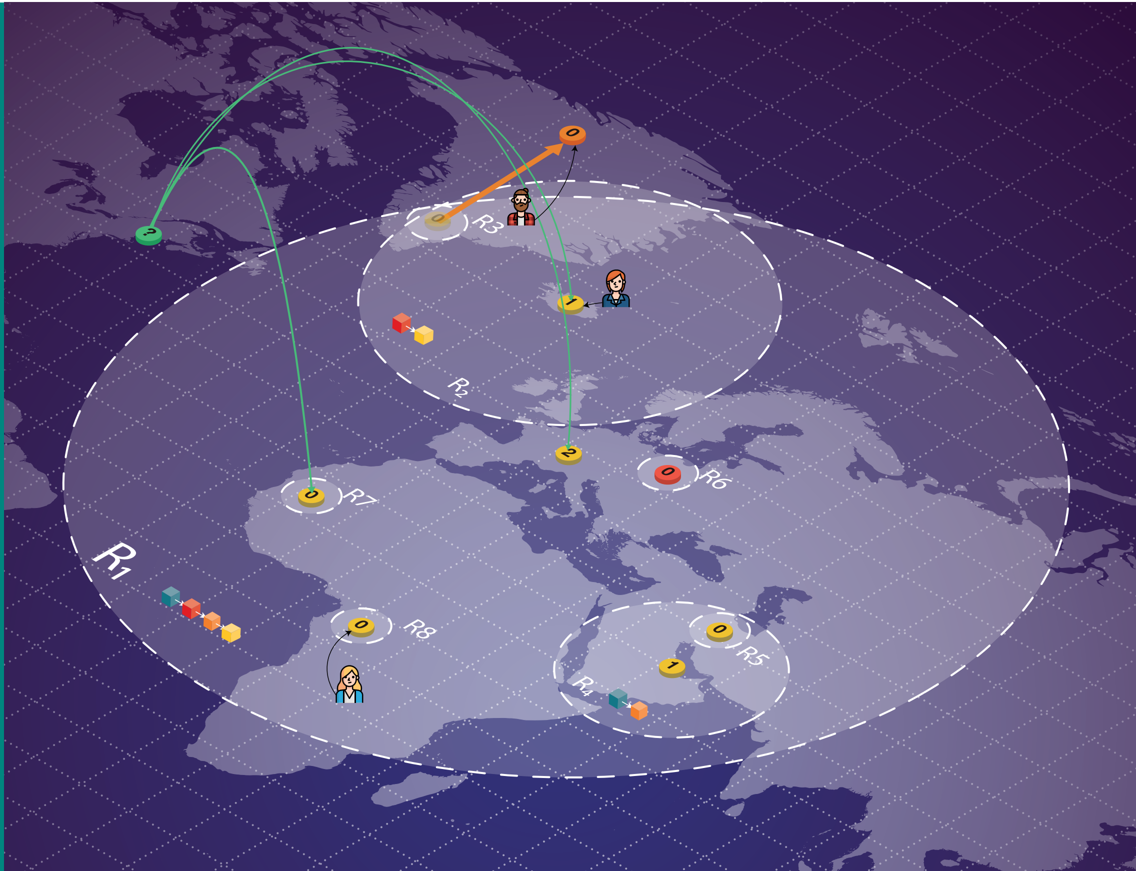
6. Conclusion



Master Thesis : Movement & Churn

What if nodes move, join or leave ?

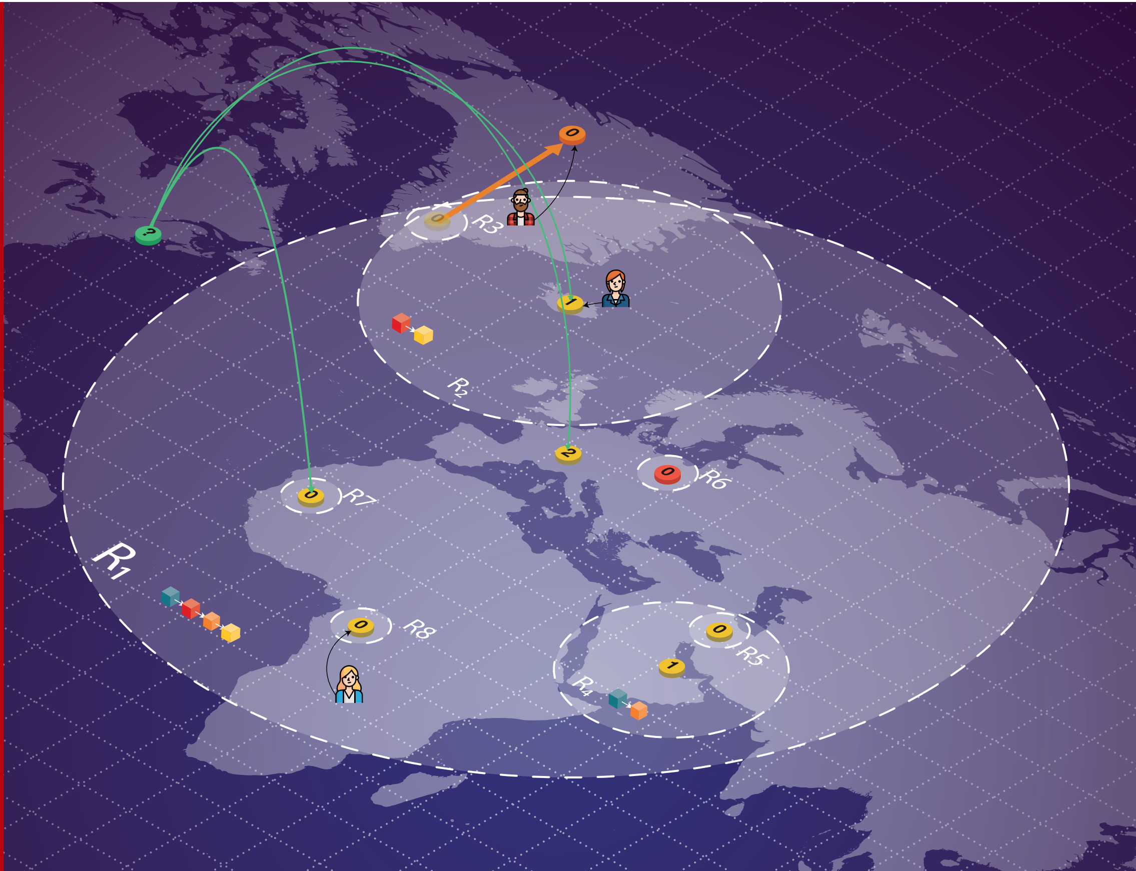
We know how to create regions for a **static system**, but we need to find a way to **adapt** the region as the system **evolves**



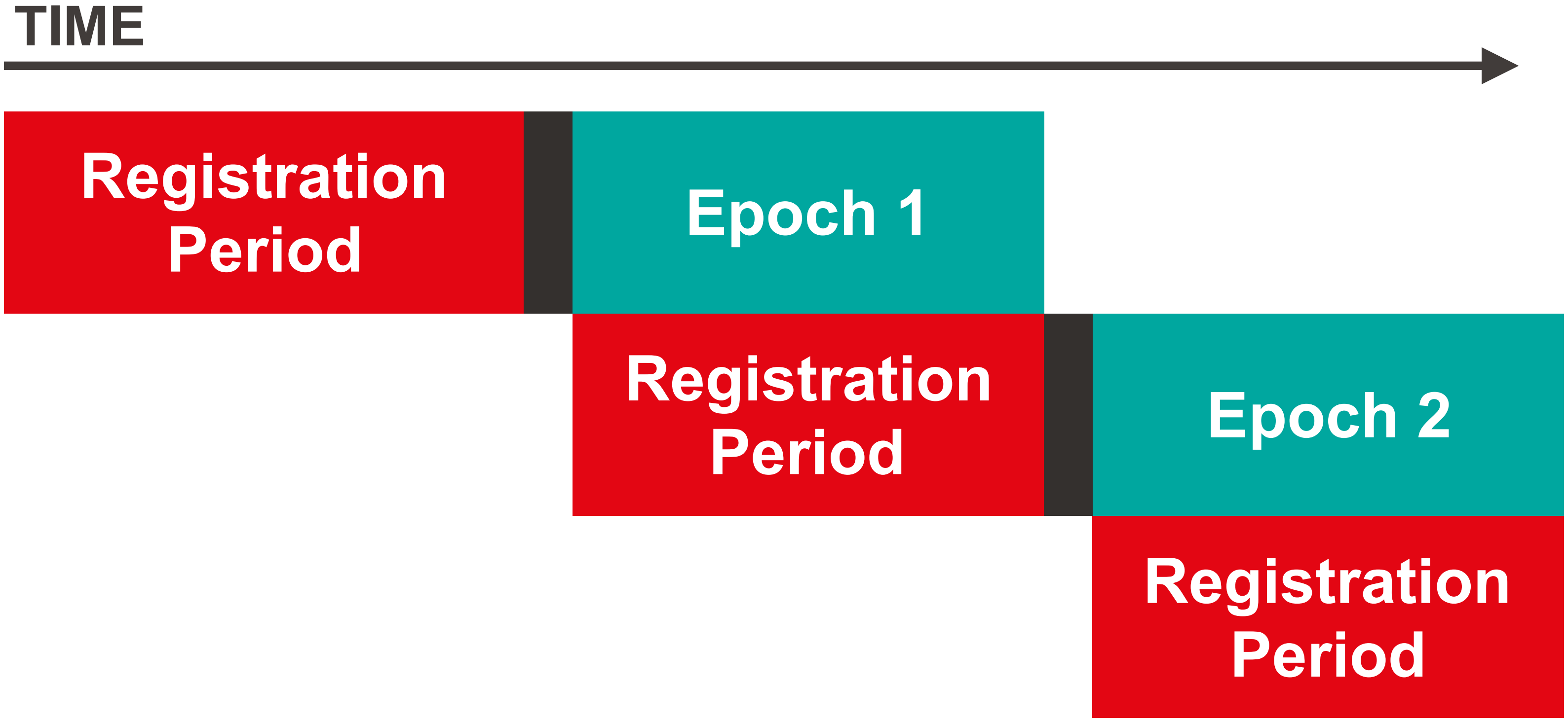
Master Thesis : Movement & Churn

Need a Control Plane !

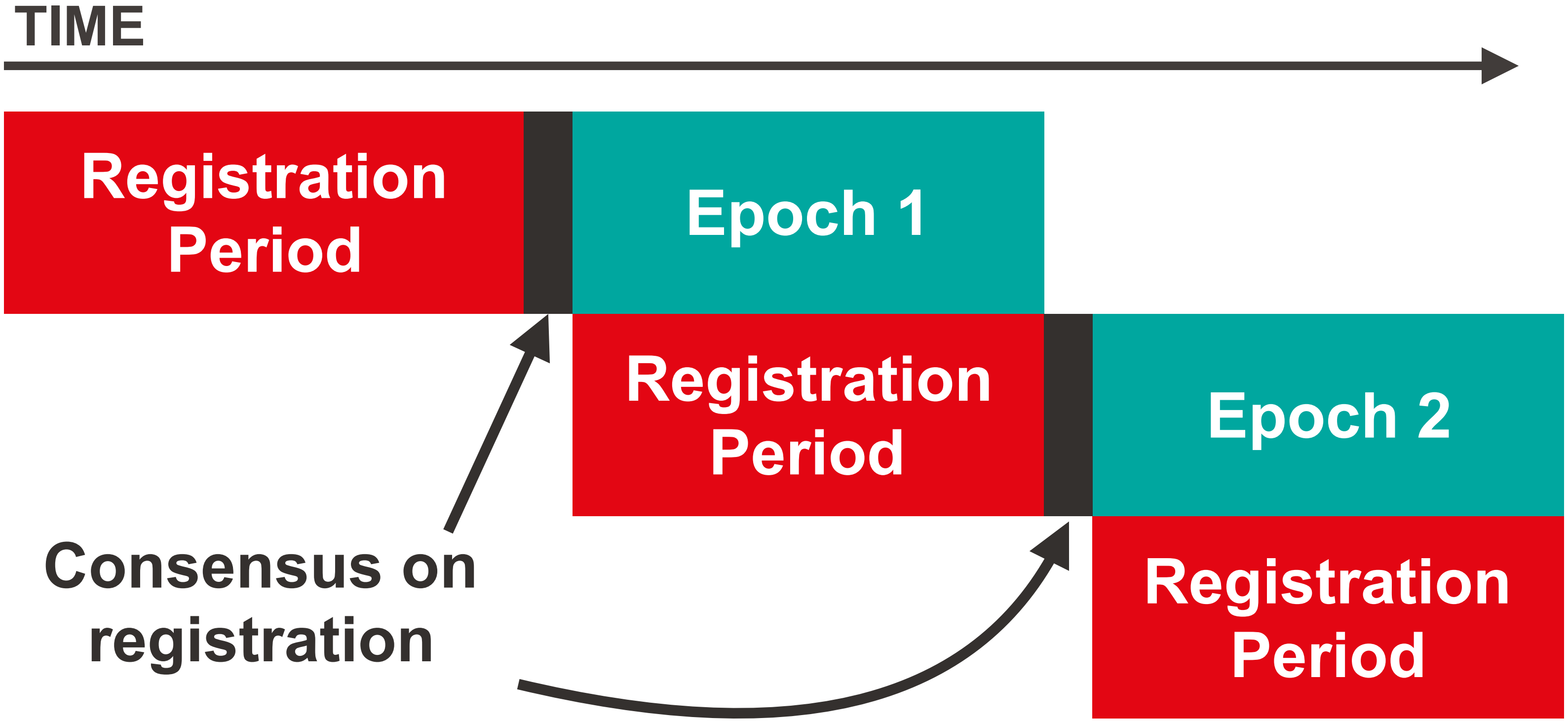
We need a **Control Plane for Locality Preserving Blockchains** : a protocol that can adapt the regions through time



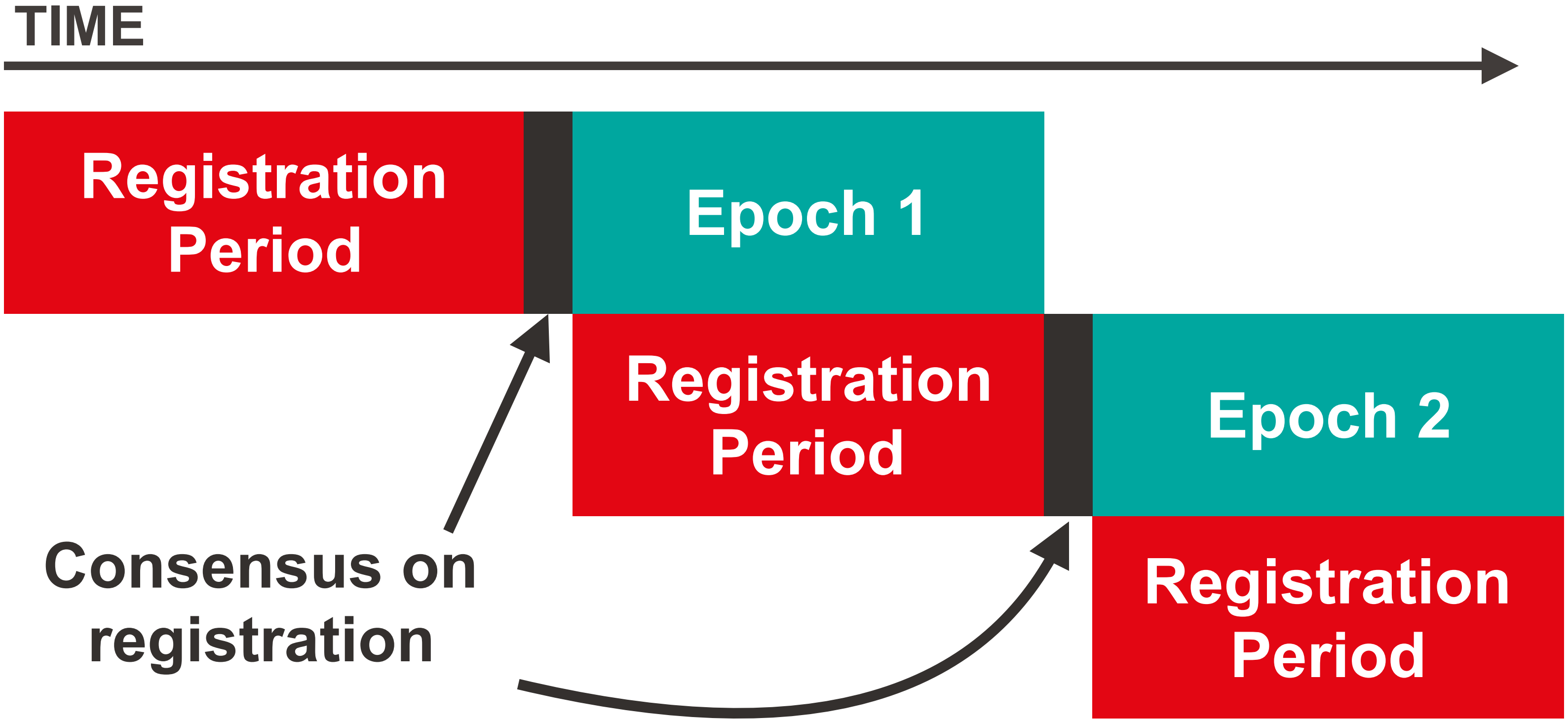
Control Plane: Protocol



Control Plane: Protocol

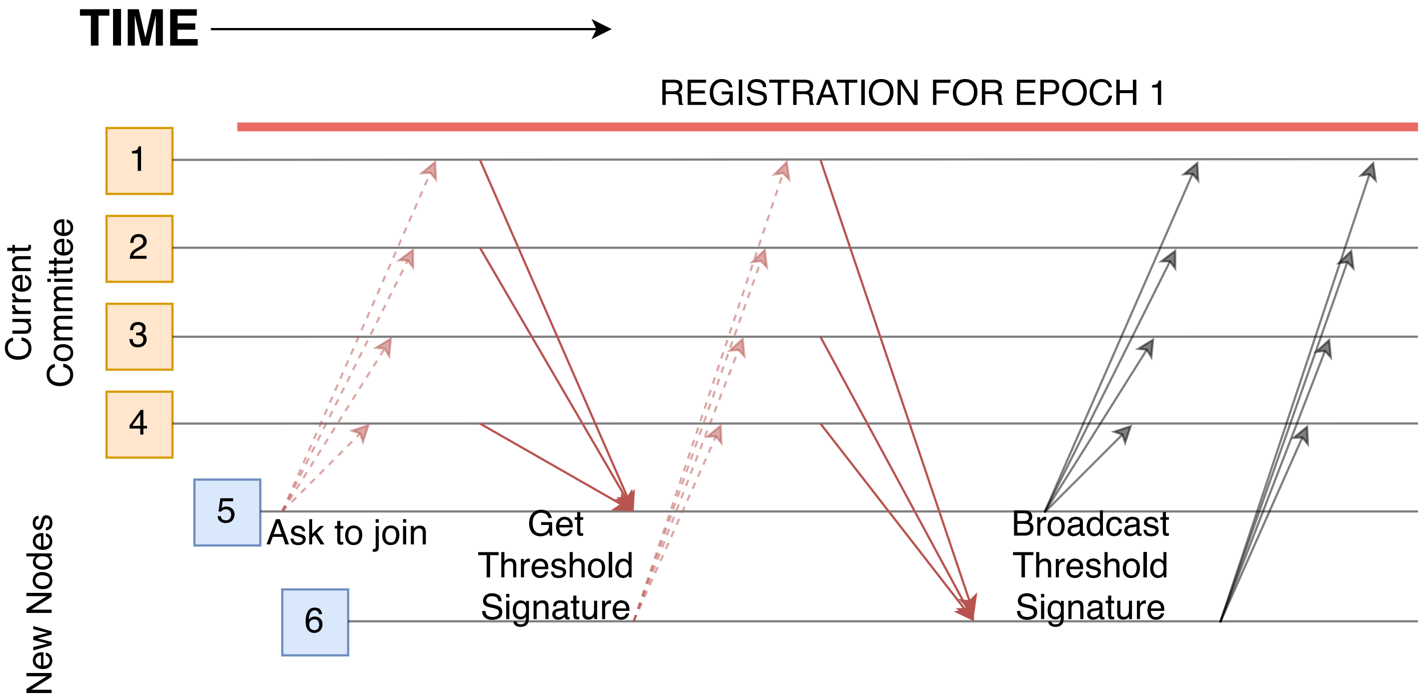


Control Plane: Protocol

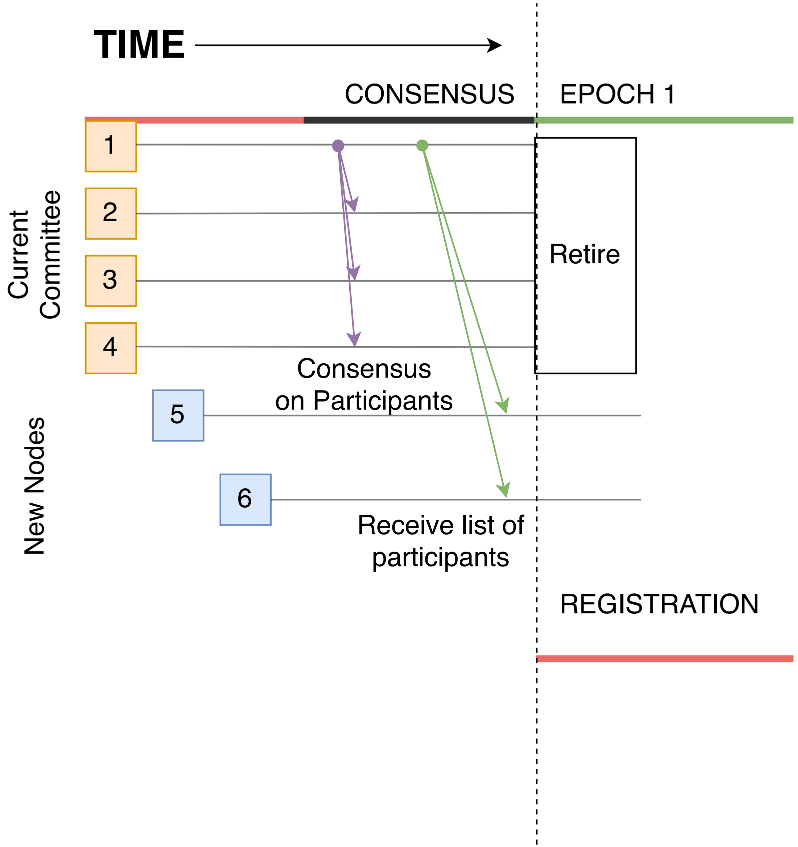


As the time is split in defined period, nodes needs *synchronized clocks.*

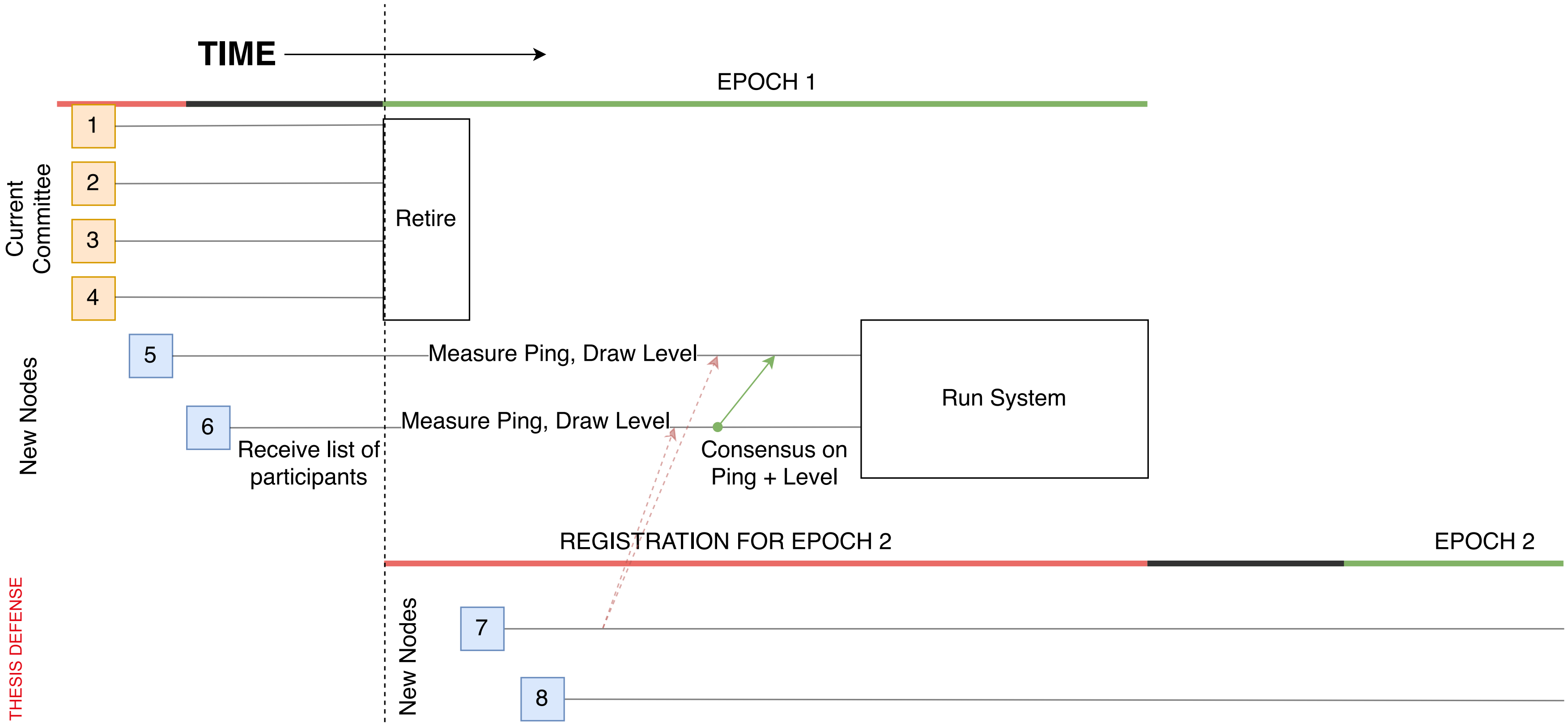
Registration Period



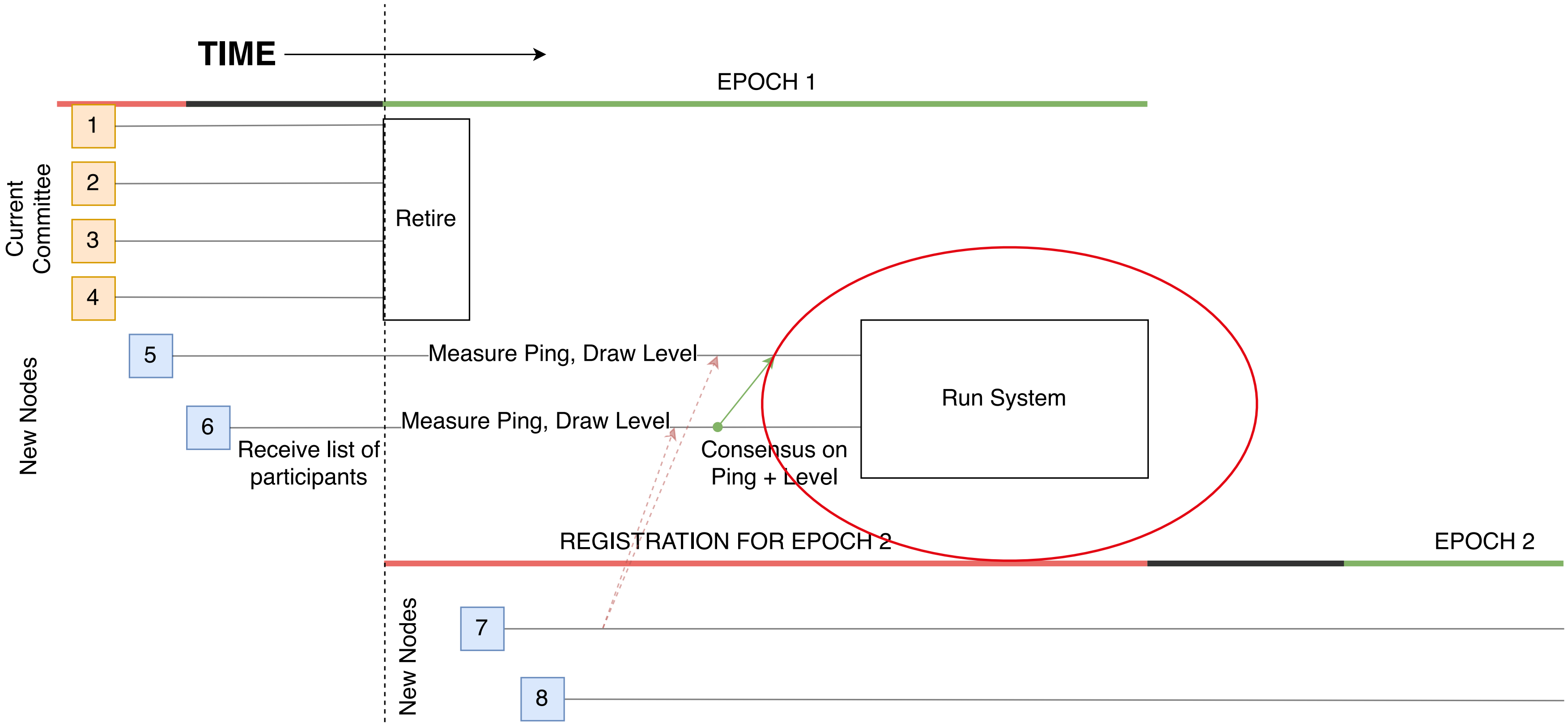
Consensus on Registration



Epoch

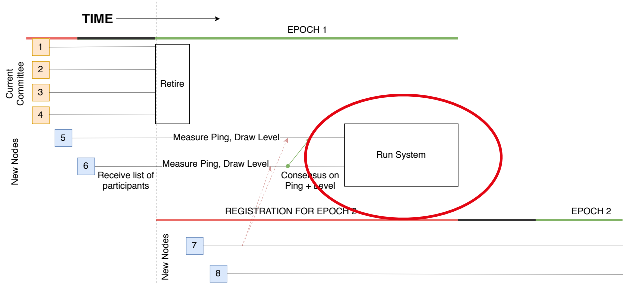

TIME
EPOCH 1
Current Committee
1
2
3
4
Retire
New Nodes
5
Measure Ping, Draw Level
6
Measure Ping, Draw Level
Receive list of participants
Consensus on Ping + Level
Run System
REGISTRATION FOR EPOCH 2
EPOCH 2
New Nodes
7
8

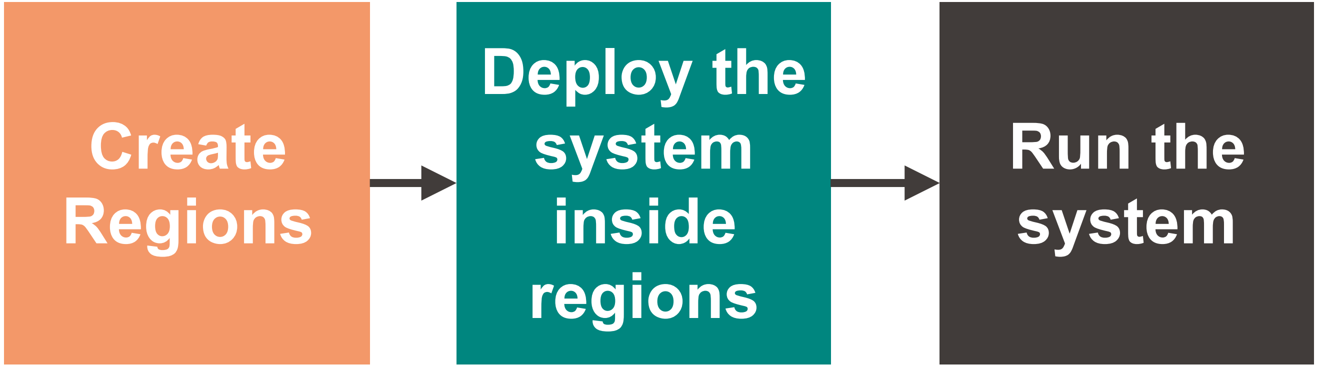
Epoch


TIME
EPOCH 1
Current Committee
1
2
3
4
Retire
New Nodes
5
Measure Ping, Draw Level
6
Measure Ping, Draw Level
Receive list of participants
Consensus on Ping + Level
Run System
REGISTRATION FOR EPOCH 2
EPOCH 2
New Nodes
7
8

MASTE

Running System



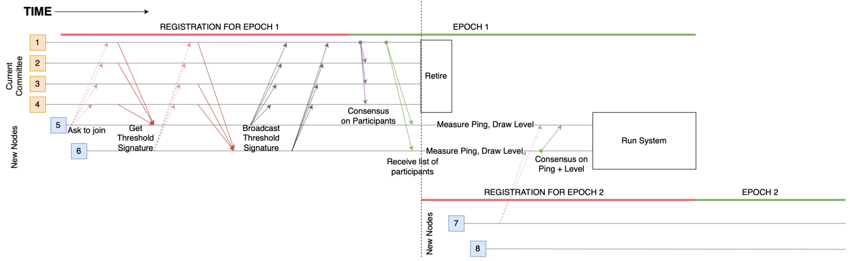


Create Regions

Deploy the system inside regions

Run the system

Security Analysis



- Delay Attacks
- Man-in-the-middle
- Malicious nodes
- Adversaries have limited computational power

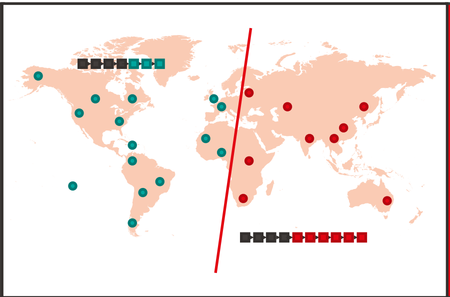
Message	Sufficient Delay	Signature
Registration Request	Admission Refused	All Signed
Threshold Signature on request		
Broadcasting Threshold Signature		
Consensus on Participants	View Change	
Consensus on pings and levels		

Outline

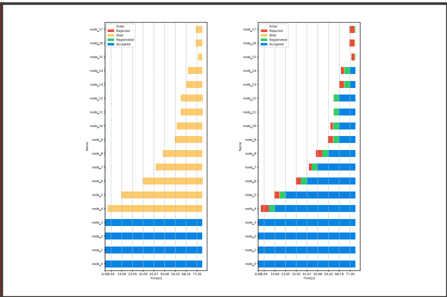
1. Some problems of traditional blockchains

WWIII Scenarios

Time for validation

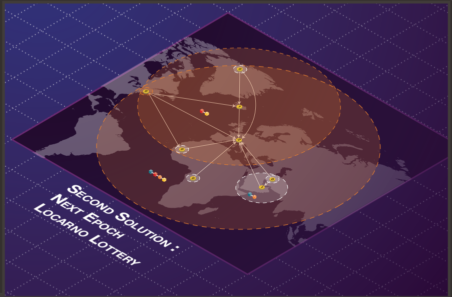


4. Results



2. Context: Nyle

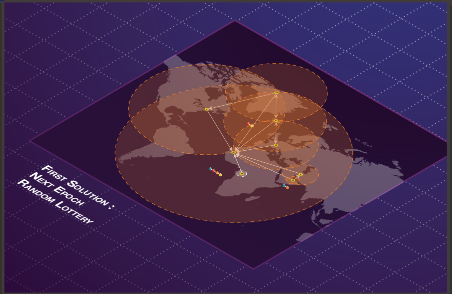
Using region replication to defeat the problems



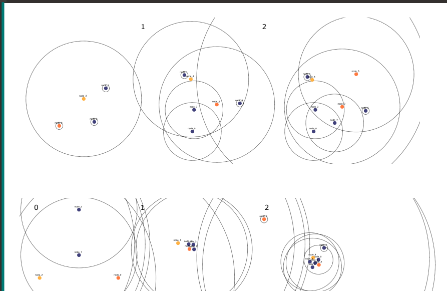
5. Improvements

3. My work

Adapt the regions to node modifications

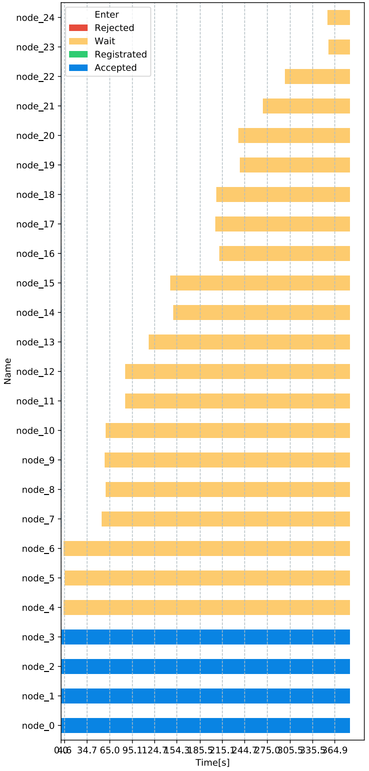


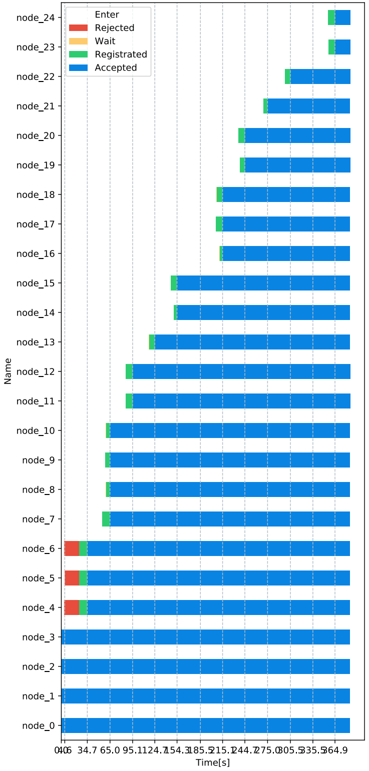
6. Conclusion



Control Plane: Results

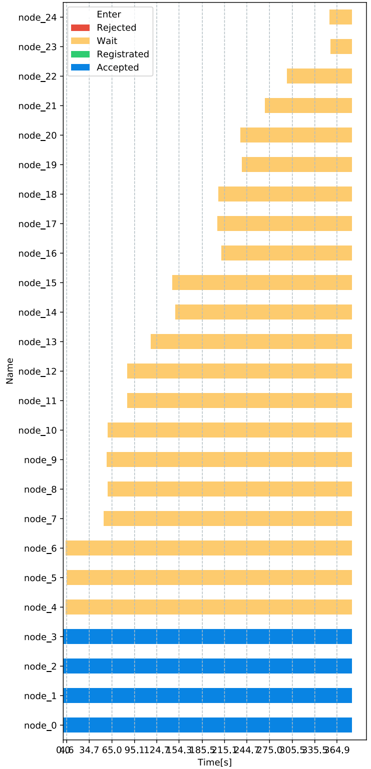
Without Control Plane



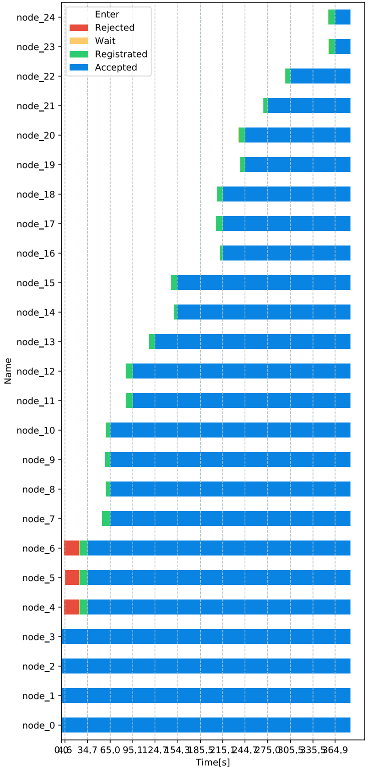


With Control Plane

Control Plane: Results



Without Control Plane

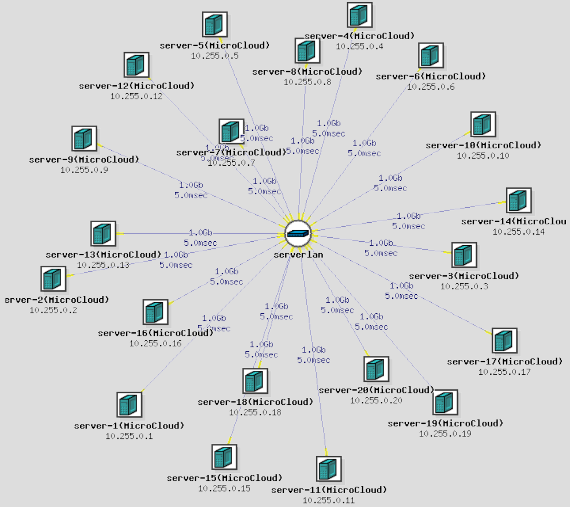


With Control Plane

Parameters of the experiment

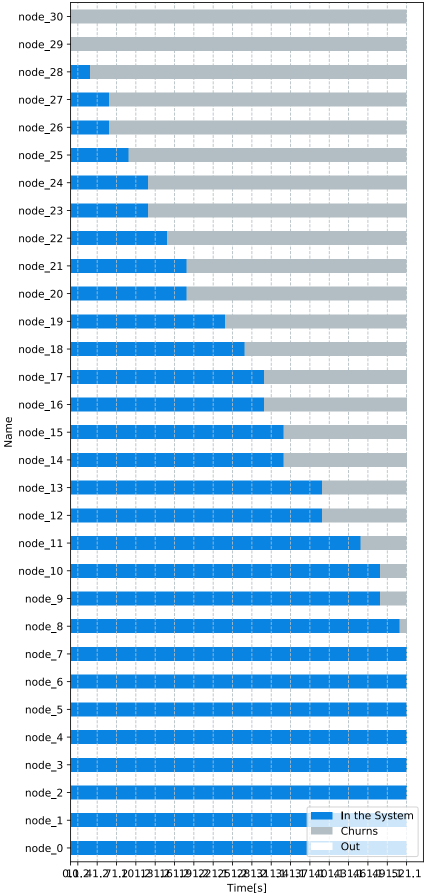
Hardware
20 MicroCloud nodes
Linked to a central LAN
Delay of links : 5 ms
Throughput of link 1.0Go
Total of 30 processes

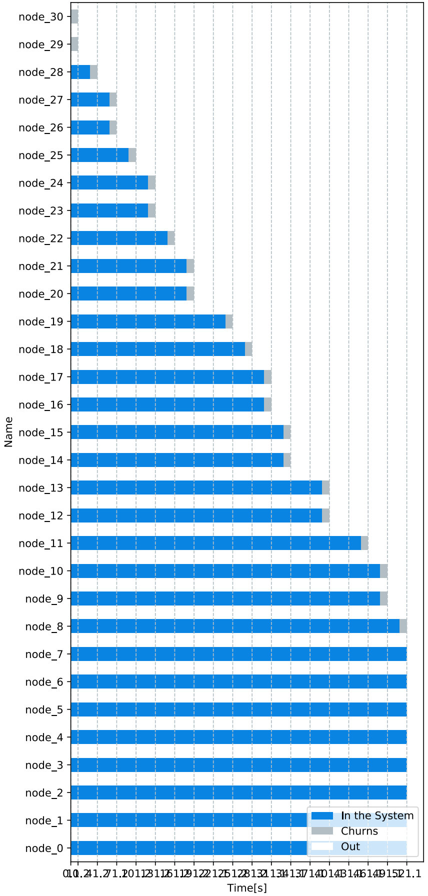
Experiment
Registration period : 10 sec
Epoch duration : 20 sec
A committee of 4 nodes is set at genesis
A random number (0-7) of nodes joins at each epoch.
Each node waits a random amount of time (between 0 and 7.5 sec) before asking for admission.
If a node failed to join at the first attempt, it will ask again for the next epoch.



Control Plane: Results

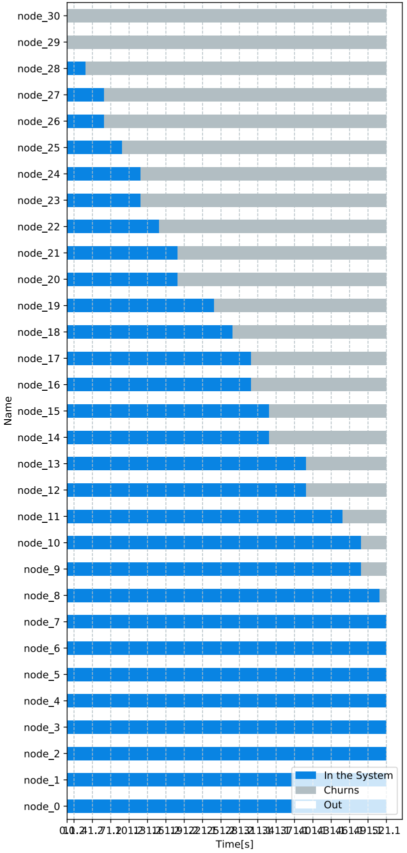
Without Control Plane



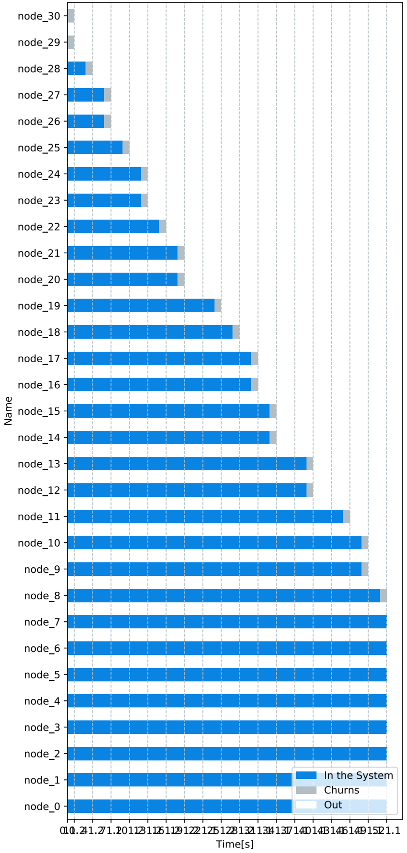


With Control Plane

Control Plane: Results



Without Control Plane

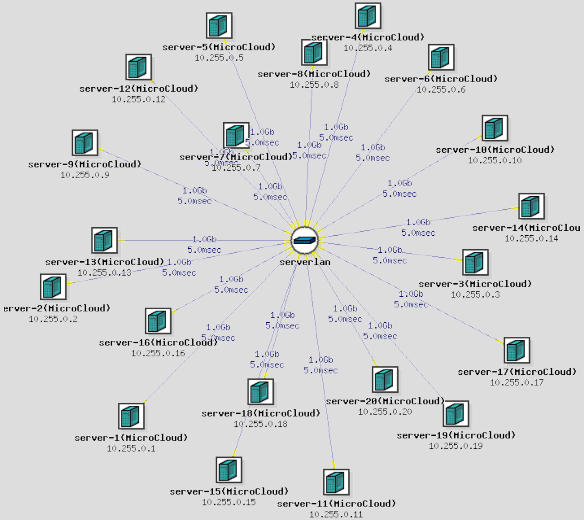


With Control Plane

Parameters of the experiment

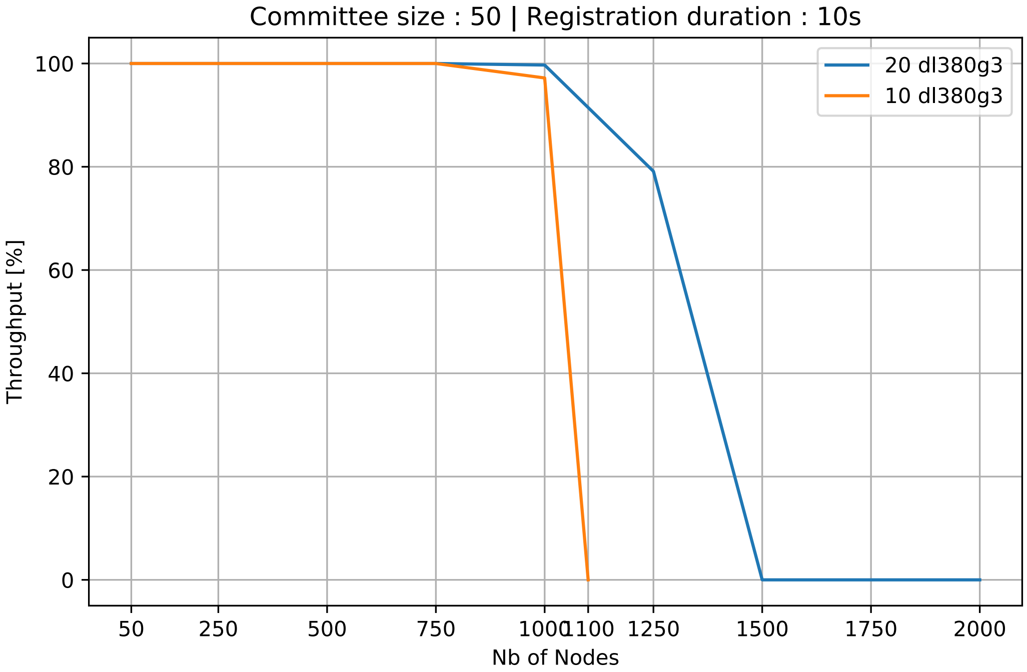
Hardware
20 MicroCloud nodes
Linked to a central LAN
Delay of links : 5 ms
Throughput of link 1.0Go
Total of 30 processes

Experiment
Registration period : 10 sec
Epoch duration : 20 sec
A committee of 30 nodes is set at genesis
A random number (0-3) of Nodes fail at each epoch.
Each node waits a random amount of time (between 0 and 7.5 sec) before failing.


server-5(MicroCloud) 10.255.0.5
server-4(MicroCloud) 10.255.0.4
server-8(MicroCloud) 10.255.0.8
server-6(MicroCloud) 10.255.0.6
server-12(MicroCloud) 10.255.0.12
server-7(MicroCloud) 10.255.0.7
server-10(MicroCloud) 10.255.0.10
server-9(MicroCloud) 10.255.0.9
server-14(MicroClou 10.255.0.14
server-13(MicroCloud) 10.255.0.13
serverlan
server-3(MicroCloud) 10.255.0.3
erver-2(MicroCloud) 10.255.0.2
server-16(MicroCloud) 10.255.0.16
server-17(MicroCloud) 10.255.0.17
server-20(MicroCloud) 10.255.0.20
server-18(MicroCloud) 10.255.0.18
server-19(MicroCloud) 10.255.0.19
server-1(MicroCloud) 10.255.0.1
server-15(MicroCloud) 10.255.0.15
server-11(MicroCloud) 10.255.0.11
1.0Gb 5.0msec

Control Plane: Experiment - Throughput

If the load on one machine becomes too large, the registration rate drops as nodes cannot complete the protocol in time



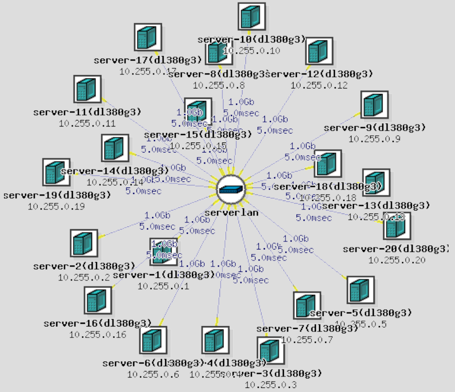
MASTER THESIS DEFENSE

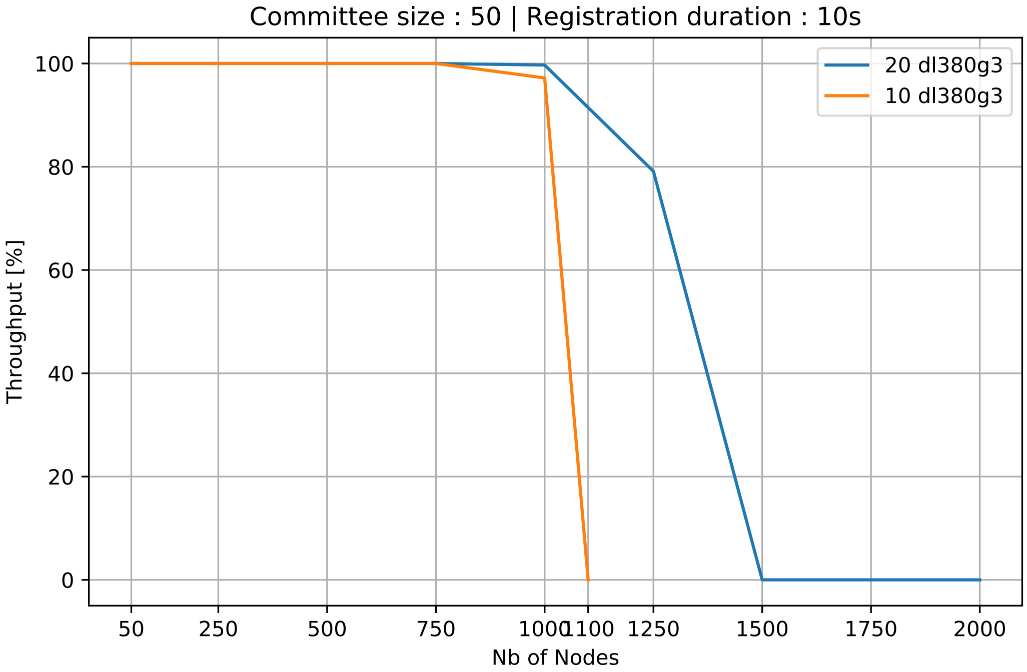
Control Plane: Experiment - Throughput

Parameters of the experiment

Hardware
20 - 10 dl380g3 nodes
Linked to a central LAN
Delay of links : 5 ms
Throughput of link 1.0Go
Total of 2000 - 1100 processes

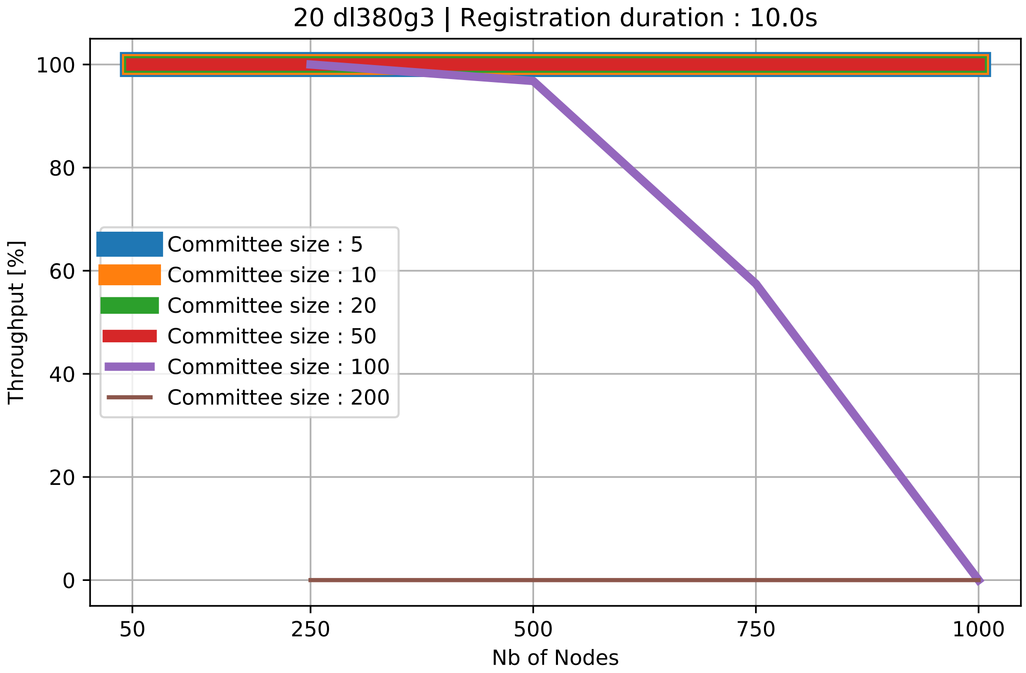
Experiment
Registration period : 10 sec
Epoch duration : 20 sec
A committee of 50 nodes is set at genesis
A certain number of nodes (depicted on the x axis) try to get accepted for the next epoch





Experiment - Committee Size

As the committee size increases, the throughput drop as the load on nodes increases

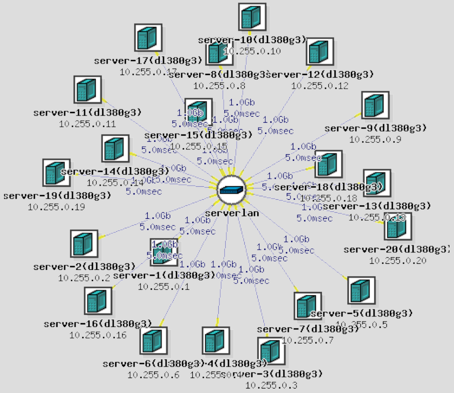


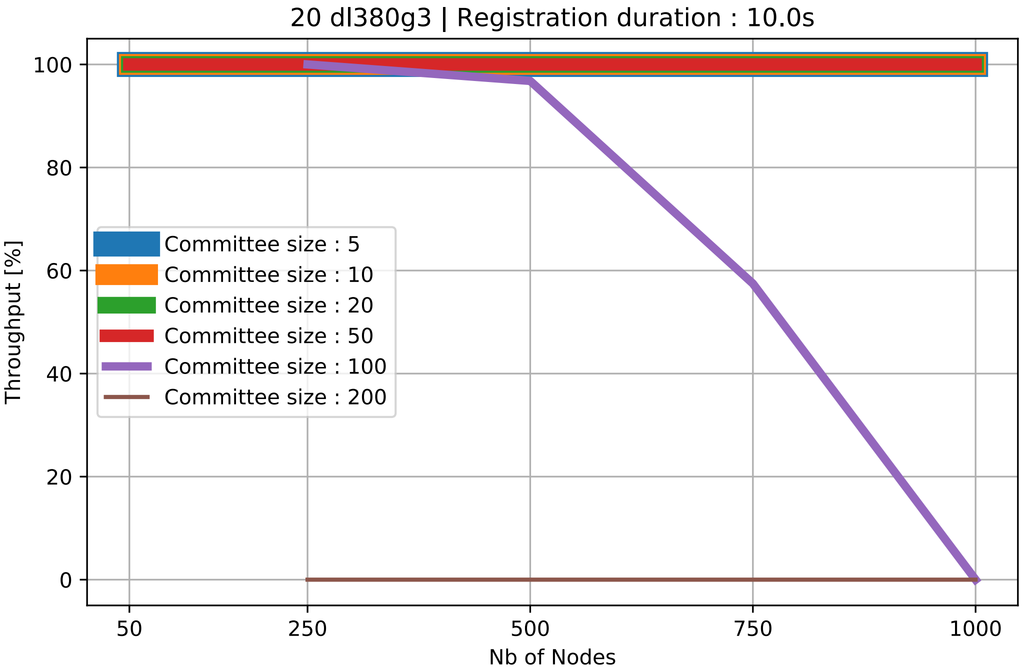
Experiment - Committee Size

Parameters of the experiment

Hardware
20 dl380g3 nodes
Linked to a central LAN
Delay of links : 5 ms
Throughput of link 1.0Go
Total of 1000 processes

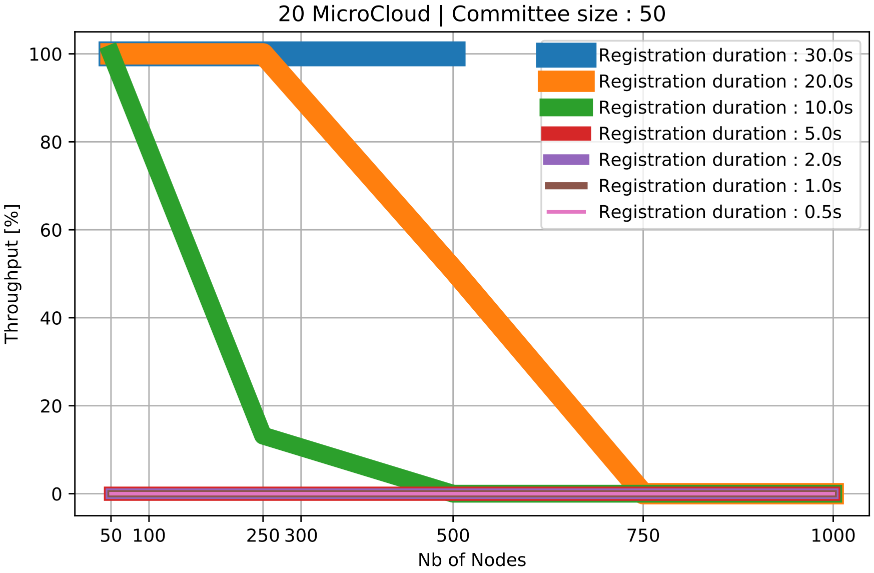
Experiment
Registration period : 10sec
Epoch duration : 20 sec
Committee Size : variable (legend)
A certain number of nodes (depicted on the x axis) try to get accepted for the next epoch


serverlan
server-1(dl380g3) 10.255.0.1
server-2(dl380g3) 10.255.0.2
server-3(dl380g3) 10.255.0.3
server-5(dl380g3) 10.255.0.5
server-6(dl380g3) 10.255.0.6
server-7(dl380g3) 10.255.0.7
server-8(dl380g3) 10.255.0.8
server-9(dl380g3) 10.255.0.9
server-10(dl380g3) 10.255.0.10
server-11(dl380g3) 10.255.0.11
server-12(dl380g3) 10.255.0.12
server-13(dl380g3)
server-14(dl380g3)
server-15(dl380g3) 10.255.0.15
server-16(dl380g3) 10.255.0.16
server-17(dl380g3)
server-18(dl380g3) 10.255.0.18
server-19(dl380g3) 10.255.0.19
server-20(dl380g3) 10.255.0.20
1.0Gb
5.0msec



Experiment - Change Duration

As the duration increases, the protocol starts to work again !

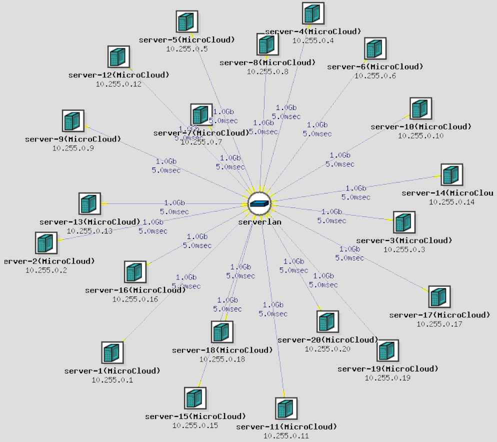


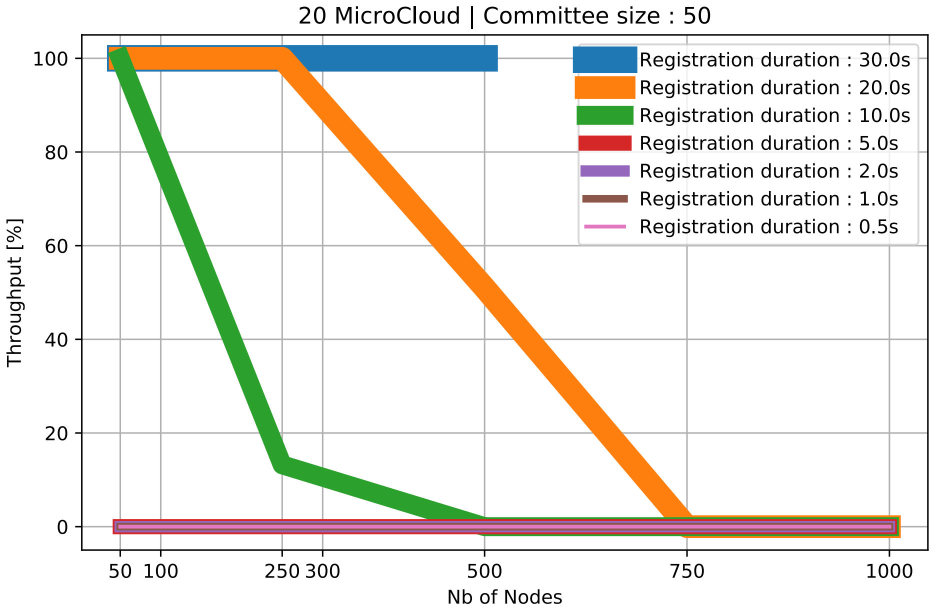
Experiment - Change Duration

Parameters of the experiment

Hardware
20 MicroCloud nodes
Linked to a central LAN
Delay of links : 5 ms
Throughput of link 1.0Go
Total 500 - 1000 processes

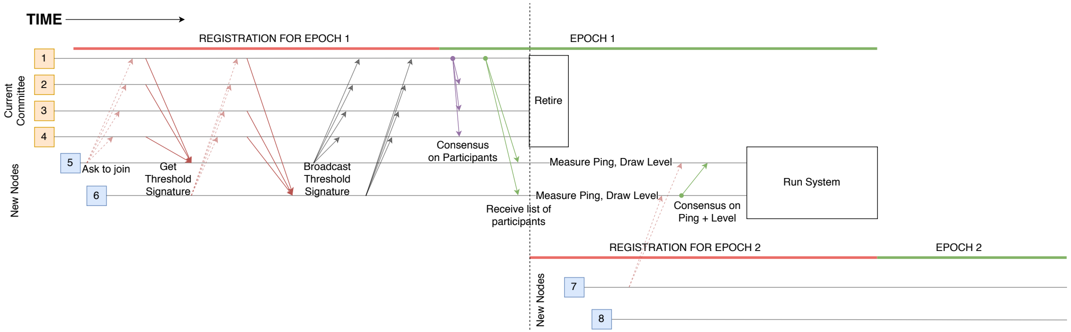
Experiment
Registration period : variable (legend)
Epoch duration : 20 sec
Committee size : 50
A certain number of nodes (depicted on the x axis) try to get accepted for the next epoch


serverlan
server-1(MicroCloud) 10.255.0.1
server-2(MicroCloud) 10.255.0.2
server-3(MicroCloud) 10.255.0.3
server-4(MicroCloud) 10.255.0.4
server-5(MicroCloud) 10.255.0.5
server-6(MicroCloud) 10.255.0.6
server-7(MicroCloud) 10.255.0.7
server-8(MicroCloud) 10.255.0.8
server-9(MicroCloud) 10.255.0.9
server-10(MicroCloud) 10.255.0.10
server-11(MicroCloud) 10.255.0.11
server-12(MicroCloud) 10.255.0.12
server-13(MicroCloud) 10.255.0.13
server-14(MicroClou 10.255.0.14
server-15(MicroCloud) 10.255.0.15
server-16(MicroCloud) 10.255.0.16
server-17(MicroCloud) 10.255.0.17
server-18(MicroCloud) 10.255.0.18
server-19(MicroCloud) 10.255.0.19
server-20(MicroCloud) 10.255.0.20
1.0Gb
5.0msec



Control Plane: Drawbacks

- Control Plane is global
- Epoch transition requires resources
- Communications

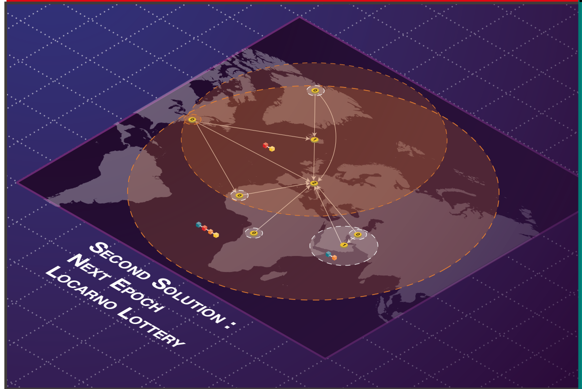


Control Plane: Improvements

Locarno Treaties

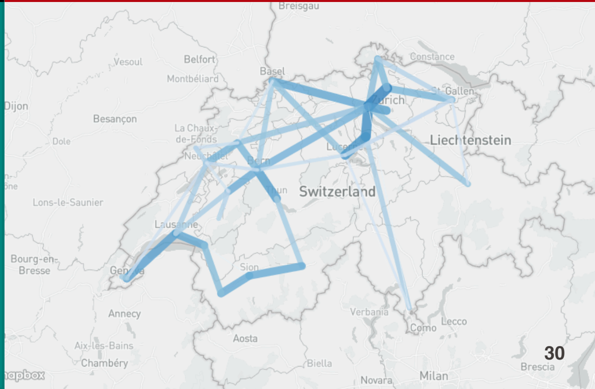
reduce the differences from one epoch to the next

Space Time Interaction distance

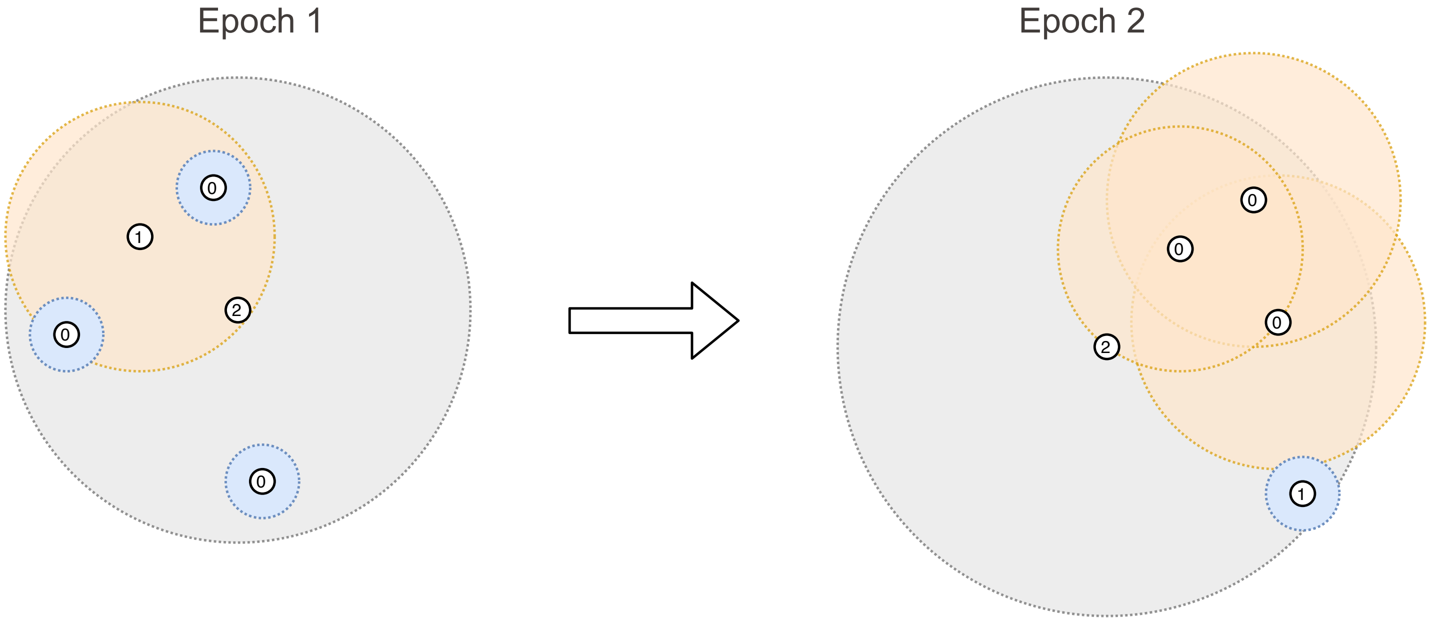


Fog of the war

reduces the amount of information one node needs to know



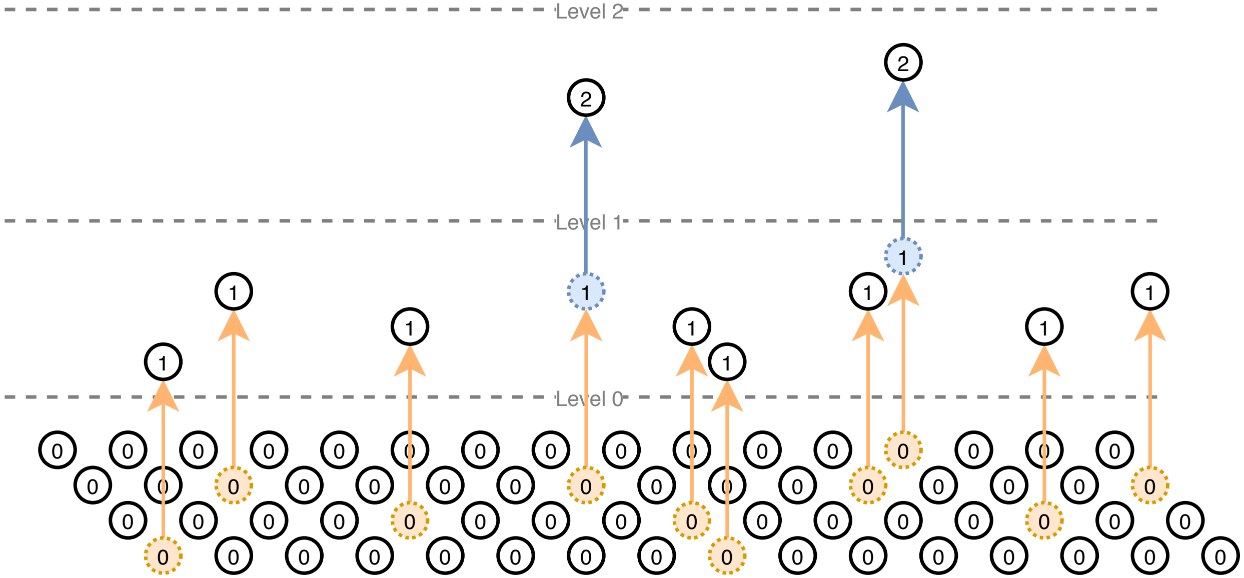
Locarno Treaties : Purpose



Random Lottery implies that regions change a lot from one epoch to the next

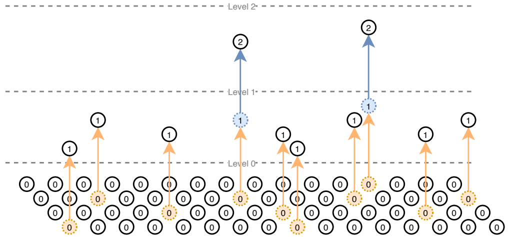
Locarno Treaties : Idea

Nyle - Random Lottery



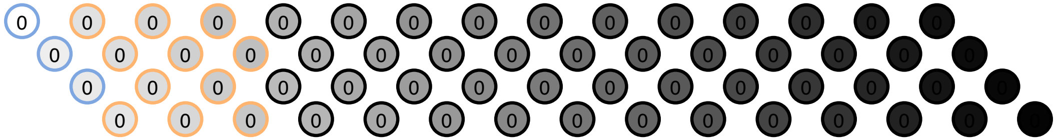
Locarno Treaties : Idea

Random Lottery



Locarno Lottery

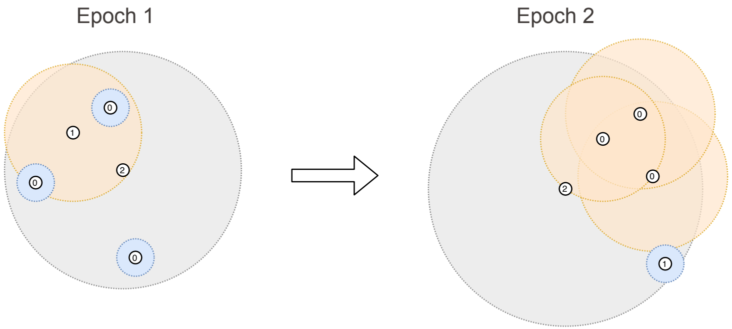
Total : 60	Level 2	Level 1	Level 0
	3	**11**	**46**



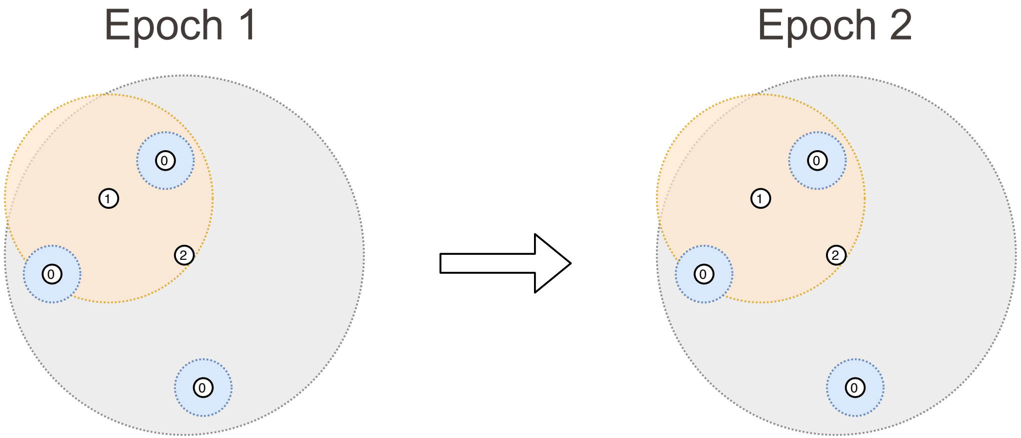
Change the lottery to allow nodes to keep their levels

Locarno Treaties : Comparison

Before : random lottery



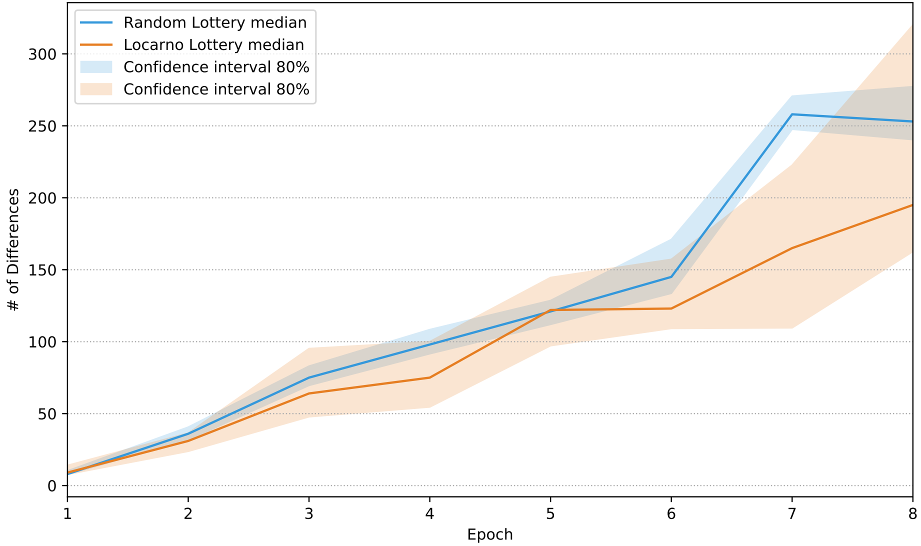
After : Locarno Lottery



If nodes keep their level, the regions do not need to be changed that much

Locarno Treaties : Evaluation

- 10 different experiments using both lotteries
- System starts with 4 nodes, 4 are added at each epoch
- Same evolution for both lotteries
- Locarno Lottery reduces the number of differences
- Variance comes from teleportation

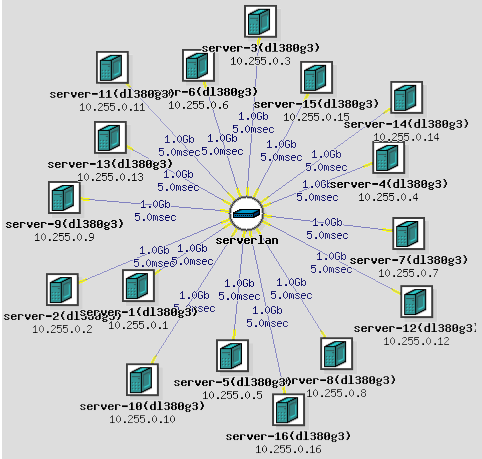


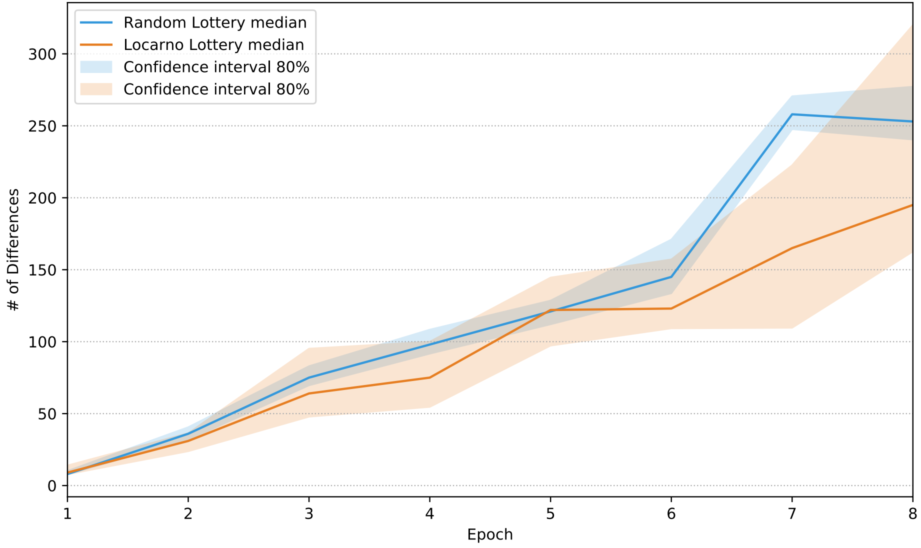
Locarno Treaties : Evaluation

Parameters of the experiment

Hardware
16 dl380g3 nodes
Linked to a central LAN
Delay of links : 5 ms
Throughput of link 1.0Go
Total of 41 processes

Experiment
Registration period : 6 sec
Epoch duration : 4 sec
Number of Epoch 8





Possible improvements

- Replace synchronized clocks by *Threshold Logical Clocks (TLC)*

- Allow the creation of regions with special meaning (for example Switzerland, Europe, …)

- Protect against possible attacks on level by checking at the beginning of one epoch the density of a levels is constant across the whole system

Conclusion

- A protocol for a **control plane in time and space for locality-preserving blockchains** was designed

- A security analysis for the control plane, some experiments and an outline of its drawbacks were made

- A solution for each drawback and some of their implementation was done

References

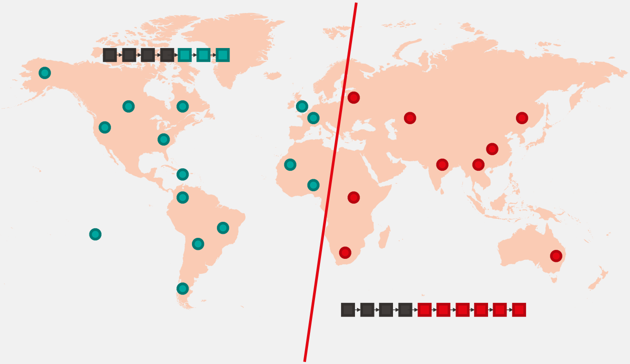
- *Maps of the world came from Free Vector Maps*
- *People Icons made by https://www.flaticon.com/authors/monkik*
- *The video game depicted in Fog of the War is Microsoft. Age of Empires II : The Age of Kings. [CD-ROM]. 1999.*
- *Maps used to display Swiss Federal Railway connection info : MapBox. https:// www.mapbox.com. Accessed: 2020-01-15.*
- *The data for Swiss Federal Railway are accessible at : opendata.swiss. https:// opendata.swiss/en/dataset/fahrplanentwurf-2019-hrdf/ resource/ 32dfd2e1-86a6-4680-9935-b76226dddee1. Accessed: 2020-01-07.*
- ***This works has found inspiration in the following papers :***
- ***Cristina Basescu, Michael F. Nowlan, Kirill Nikitin, Jose M. Faleiro, and Bryan Ford. "Crux: Locality-Preserving Distributed Services". In: (June 2014). arXiv: 1405.0637. U R L : http: /arxiv.org/abs/1405.0637.***
- *Dan Boneh, Manu Drijvers, and Gregory Neven. "Compact Multi-signatures for Smaller Blockchains". In: Advances in Cryptology – ASIACRYPT 2018. Ed. by Thomas Peyrin and Steven Galbraith. Cham: Springer International Publishing, 2018, pp. 435–464. I S B N : 978- 3-030-03329-3.*
- *Miguel Castro and Barbara Liskov. "Practical Byzantine Fault Tolerance". In: February*

  (1999), pp. 1–14.
- *D. Greenhoe. "Properties of distance spaces with power triangle inequalities". In: Carpathian Mathematical Publications 8.1 (2016). I S S N : 2075-9827. D O I : 10 . 15330 / cmp.8.1.51-82.*
- *Eleftherios Kokoris Kogias, Philipp Jovanovic, Nicolas Gailly, Ismail Khoffi, Linus Gasser, Bryan Ford, Eleftherios Kokoris-Kogias, and Bryan Ford Epfl. "Enhancing Bitcoin Security and Performance with Strong Consistency via Collective Signing". In: Proceedings of the 25th USENIX Security Symposium (2016). arXiv: 1602.06997. U R L : https://www.usenix. org/conference/usenixsecurity16/technical-sessions/presentation/kogias.*
- *Leslie Lamport. "The Part-Time Parliament". In: 2.May 1998 (2000*
- *Marta Lokhava, Giuliano Losa, David Mazières, Graydon Hoare, Nicolas Barry, Eli Gafni, Jonathan Jove, Rafał Malinowsky, and Jed McCaleb. "Fast and secure global payments with Stellar". In: (2019), pp. 80–96. D O I : 10.1145/3341301.3359636.*
- *Satoshi Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System". In: (Mar. 2009). U R L : https://bitcoin.org/bitcoin.pdf.*
- *Maxime Sierro, Bryan Ford, Cristina Basescu, and Kelong Cong. "Locality-Preserving Blockchain Implementation". In: (2019). URL: https://github.com/dedis/student% 7B%5C_%7D19%7B%5C_%7Dnylechain/blob/master/report/report.pdf.*
- *Ewa Syta, Philipp Jovanovic, Eleftherios Kokoris Kogias, Nicolas Gailly, Linus Gasser, Ismail Khoffi, Michael J. Fischer, and Bryan Ford. "Scalable Bias-Resistant Distributed Random- ness". In: (2016). https://eprint.iacr.org/2016/1067.*
- *Jiaping Wang and Hao Wang. "Monoxide: Scale out Blockchains with Asynchronous Con- sensus Zones". In: Proceedings of the 16th USENIX Symposium on Networked Systems De- sign and Implementation (NSDI '19) (2019). U R L : https://www.usenix.org/conference/ nsdi19/ presentation/wang-jiaping.*
- *Gavin Wood et al. "Ethereum: A secure decentralised generalised transaction ledger". In: Ethereum project yellow paper 151.2014 (2014), pp. 1–32.*
- *Maofan Yin, Dahlia Malkhi, Michael K. Reiter, Guy Golan Gueta, and Ittai Abraham. "Hot- Stuff: BFT Consensus in the Lens of Blockchain". In: (2018), pp. 1–23. arXiv: 1803.05069. U R L : http://arxiv.org/abs/1803.05069.*

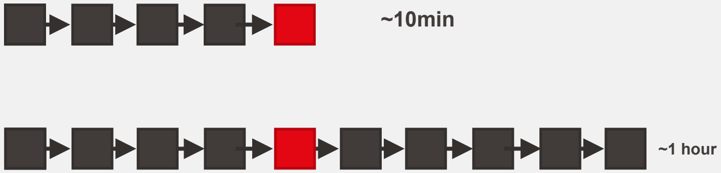
Backup Slides

Problems of traditional blockchains

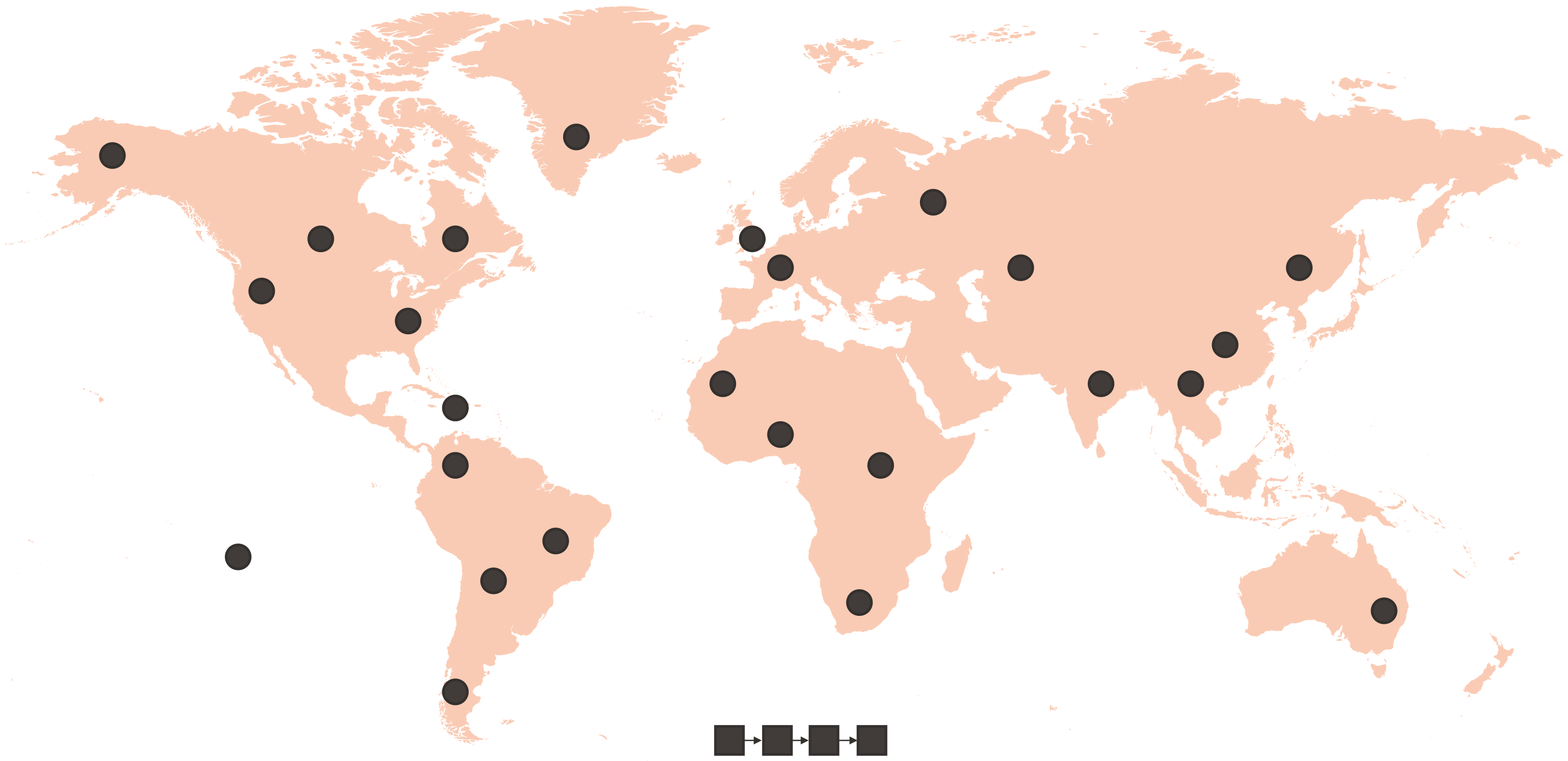
World War III Scenarios



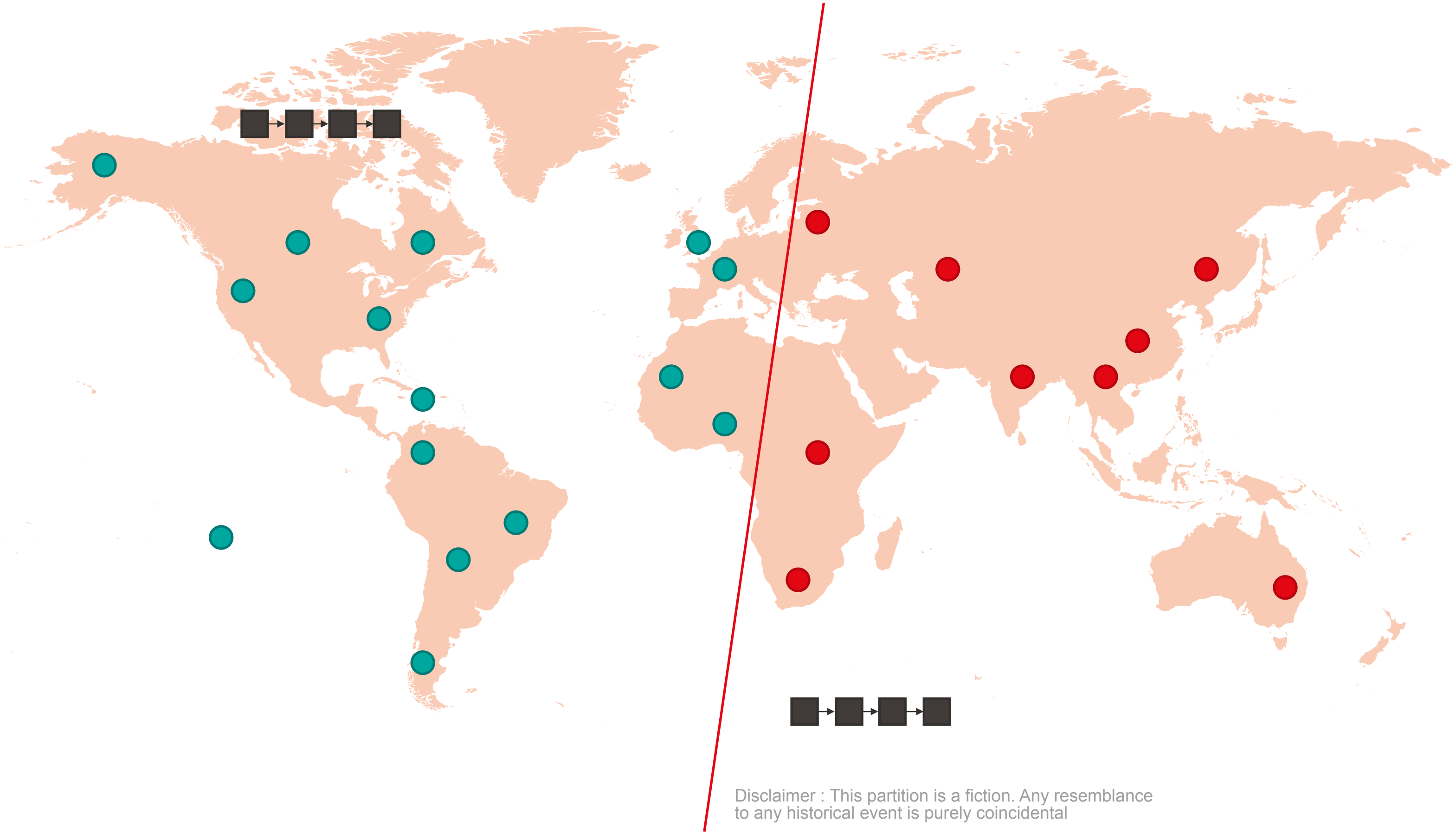
Time for validation



World War III Scenarios



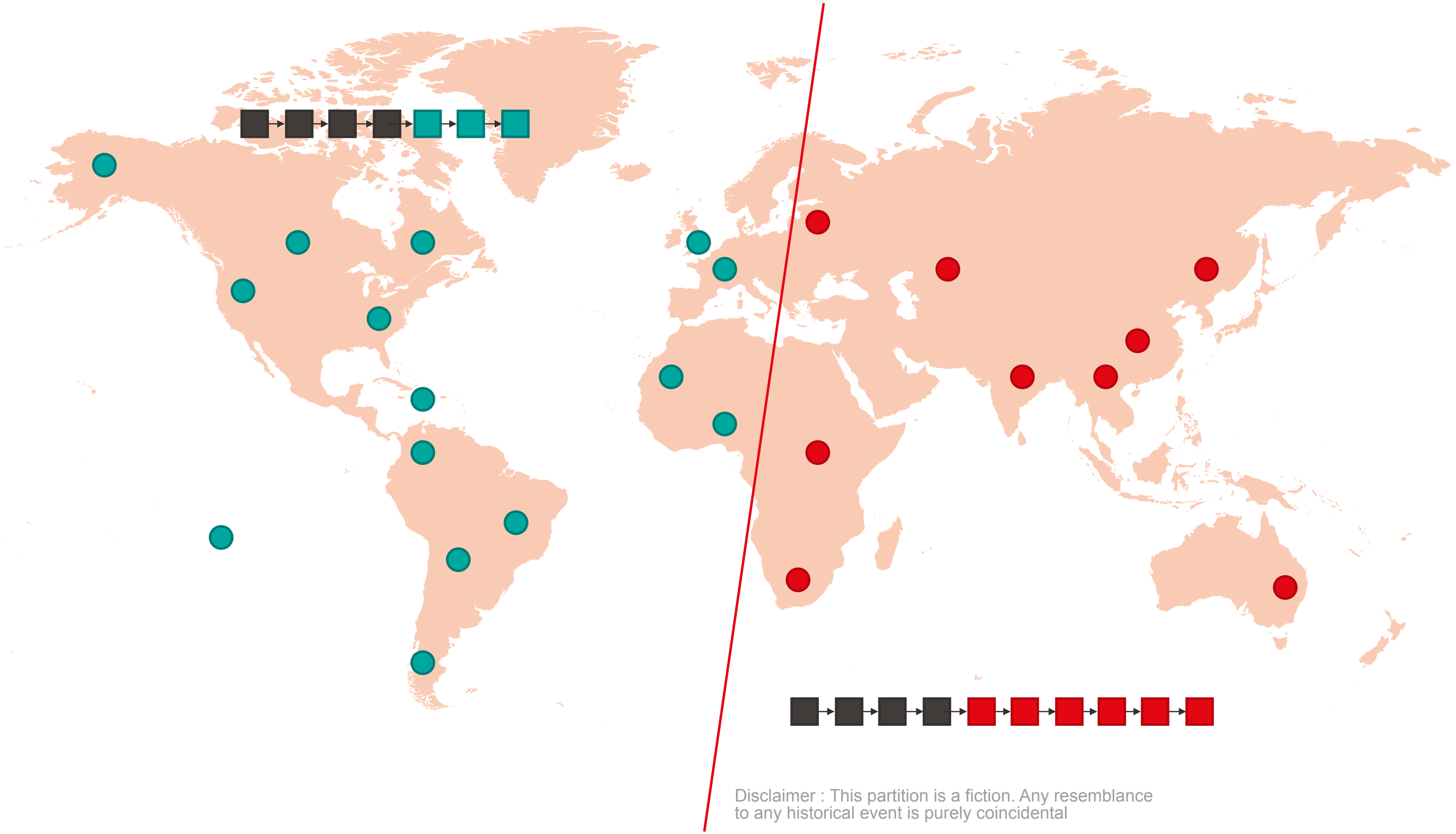
World War III Scenarios



Disclaimer : This partition is a fiction. Any resemblance to any historical event is purely coincidental

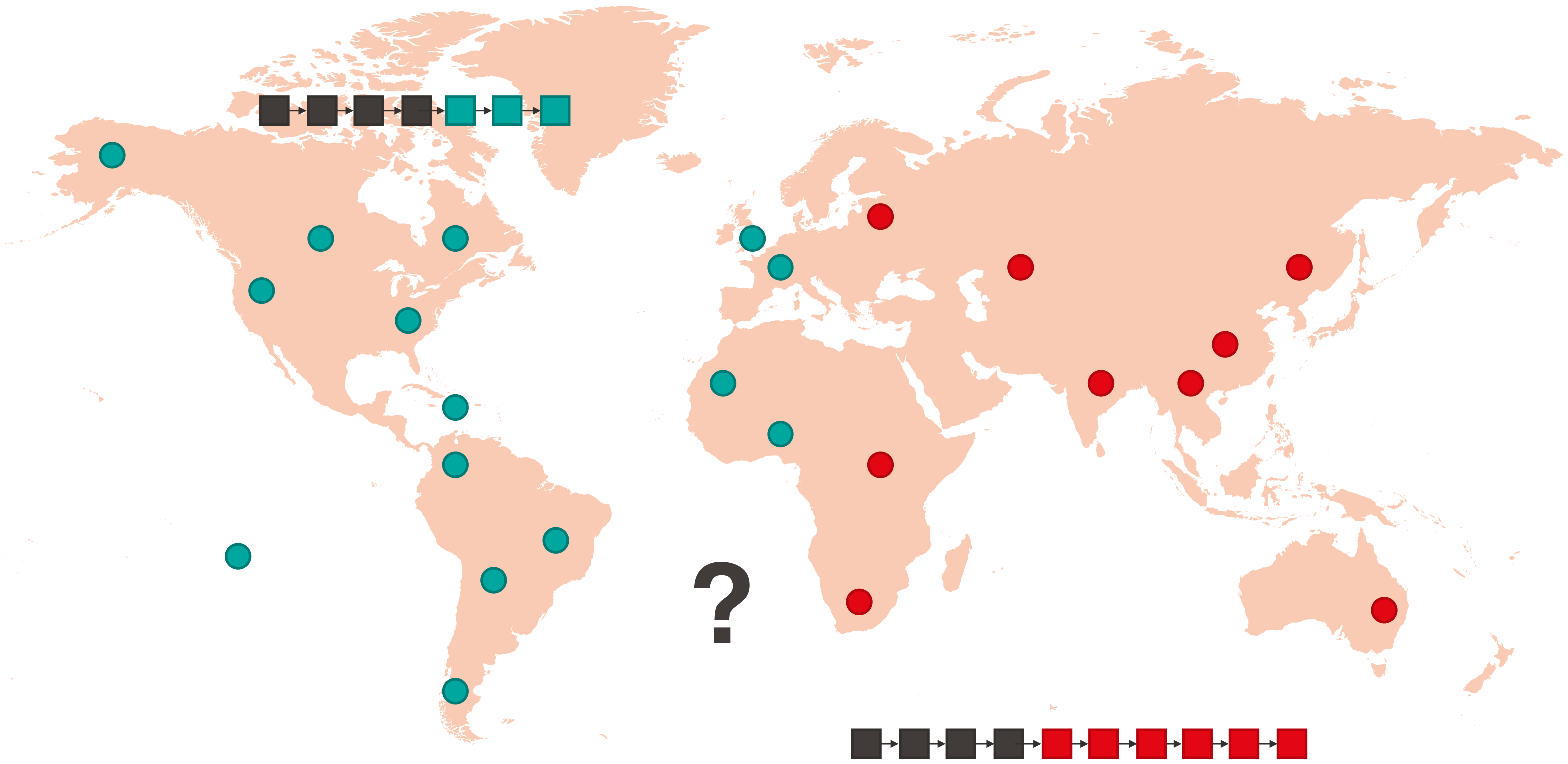
MASTER THESIS DEFENSE

World War III Scenarios



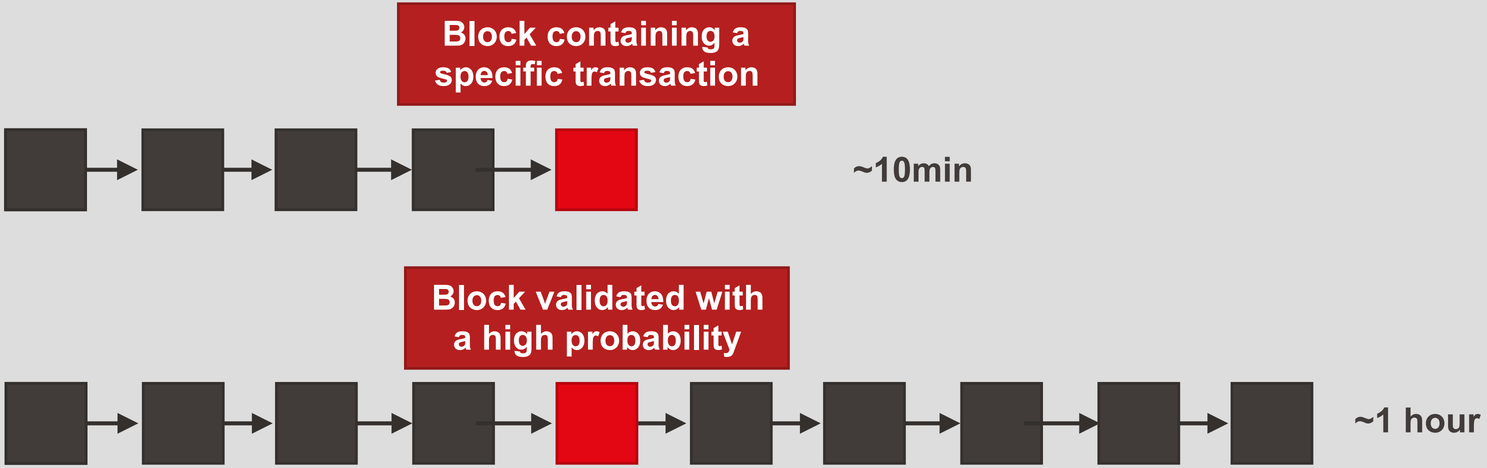
Disclaimer : This partition is a fiction. Any resemblance
to any historical event is purely coincidental

World War III Scenarios



Time for validation

■ Adding a block takes around 10minutes



Block containing a specific transaction

~10min

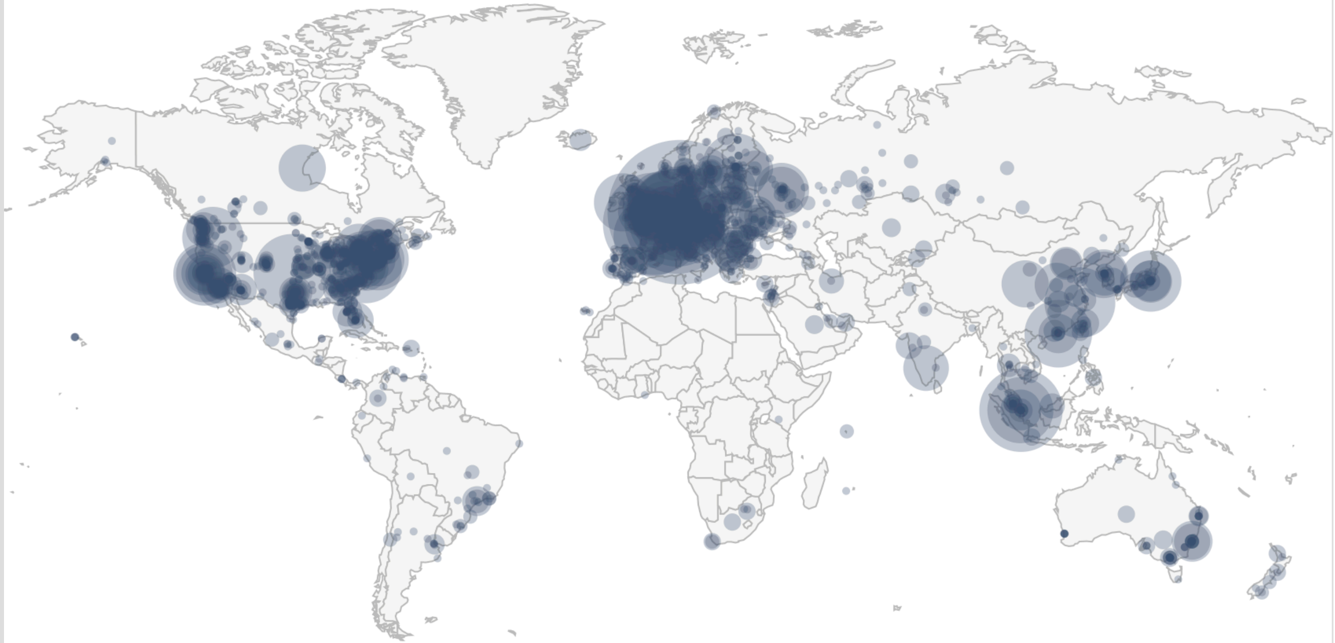
Block validated with a high probability

~1 hour

Control Plane : Purposes

- Nyle only computes the control plane once
- In and Internet-like network nodes comes and go and latencies change

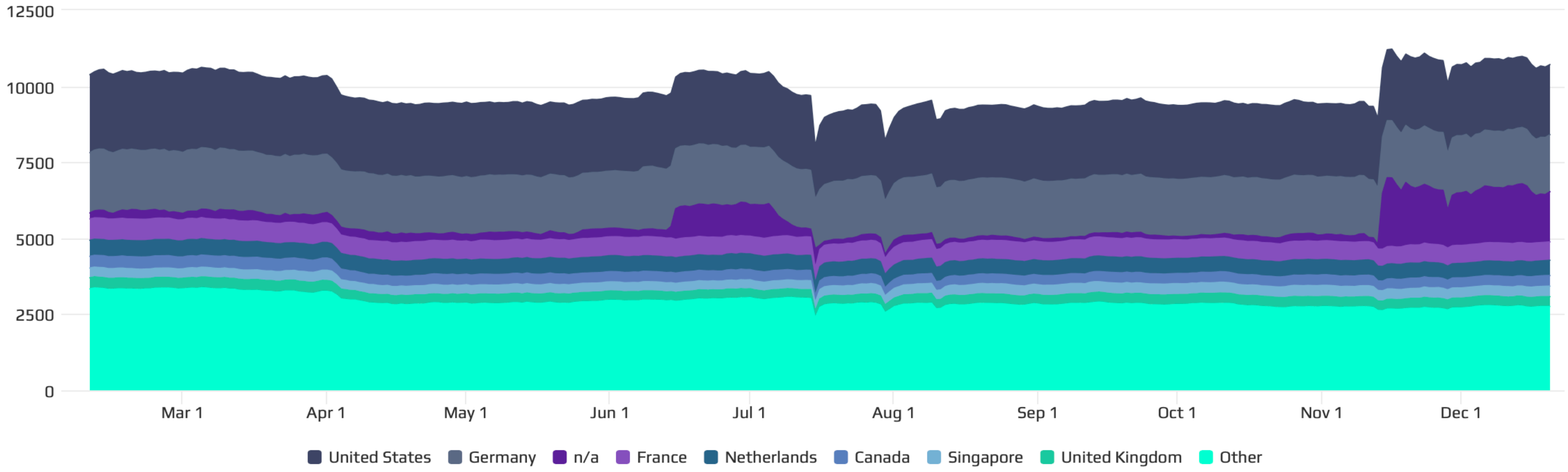
Concentration of reachable Bitcoin nodes found in countries around the world.



Visualization and data from bitnodes.io

Control Plane : Purposes

In and Internet-like network nodes comes and go and latencies change



Number of reachable nodes in the Bitcoin Network during the last 365 days

Visualization and data from bitnodes.io

Context : Nyle

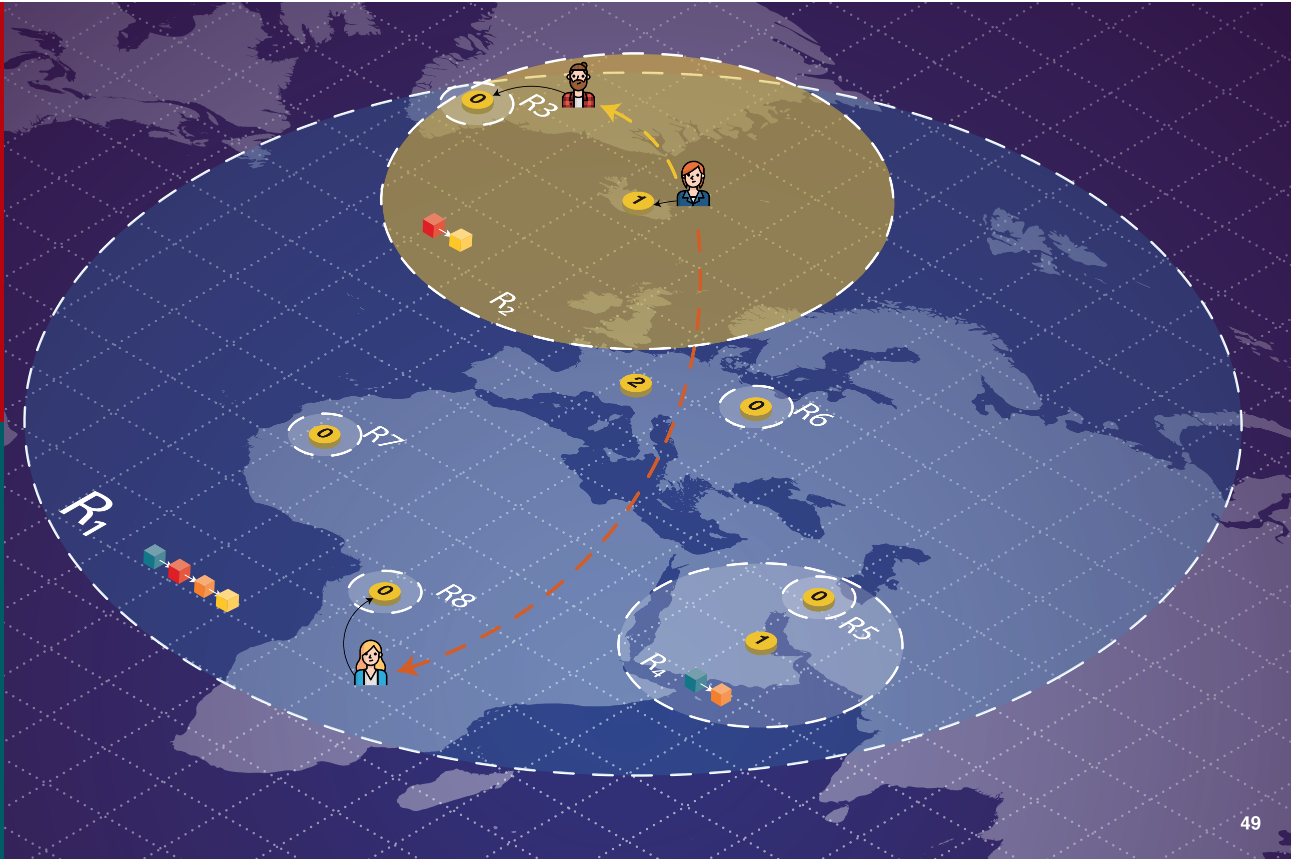
Replicates the system in regions, from local to global

- World War III Scenarios

  If a global partition occurs, the system still works in regions that are not split by a partition

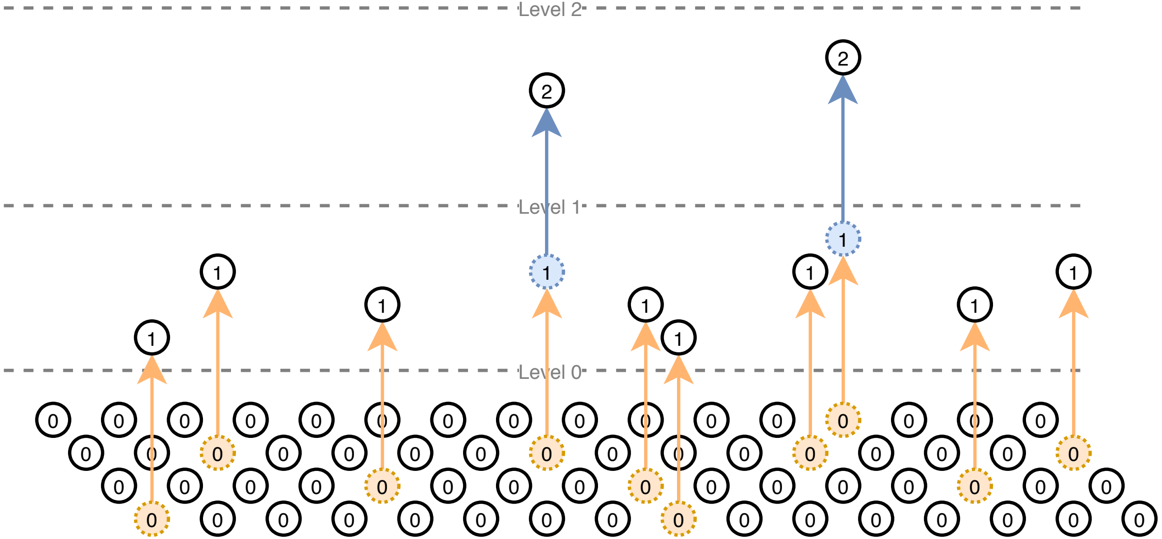
- Time for validation

  Transactions can be validated in regions



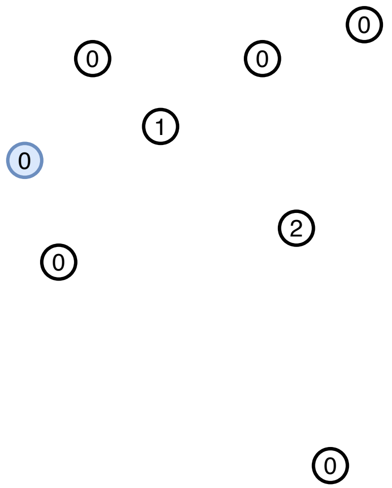
Nyle : Tools

Lottery



Nyle : Tools

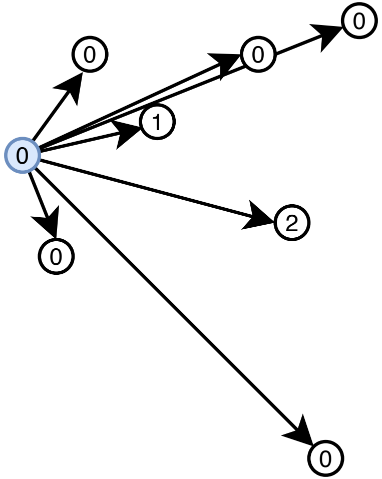
Bunch



A Node considers all other nodes in ascending order of distance. It adds another node in its *bunch* if has not already seen a node of a bigger level.

Nyle : Tools

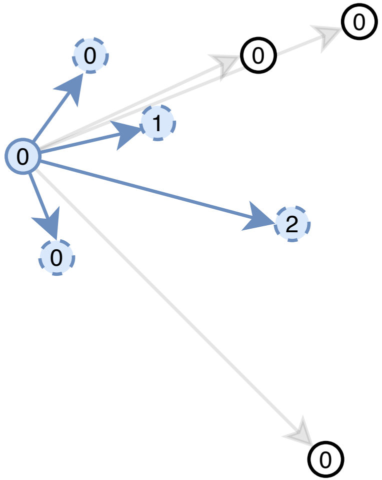
Bunch



A Node considers all other nodes in ascending order of distance. It adds another node in its *bunch* if has not already seen a node of a bigger level.

Nyle : Tools

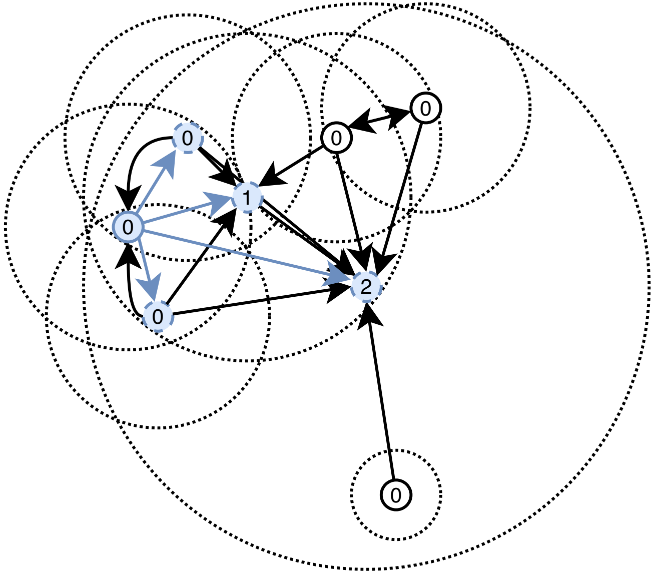
Bunch



A Node considers all other nodes in ascending order of distance. It adds another node in its *bunch* if has not already seen a node of a bigger level.

Nyle : Tools

Bunch



A have B in his bunch

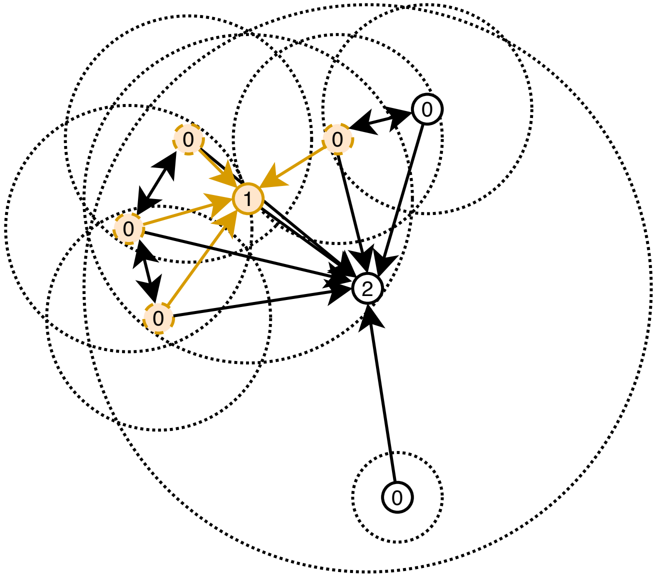
0 node considered

1 node in the bunch

A Node considers all other nodes in ascending order of distance. It adds another node in its *bunch* if has not already seen a node of a bigger level.

Nyle : Tools

Cluster



A have B in his bunch

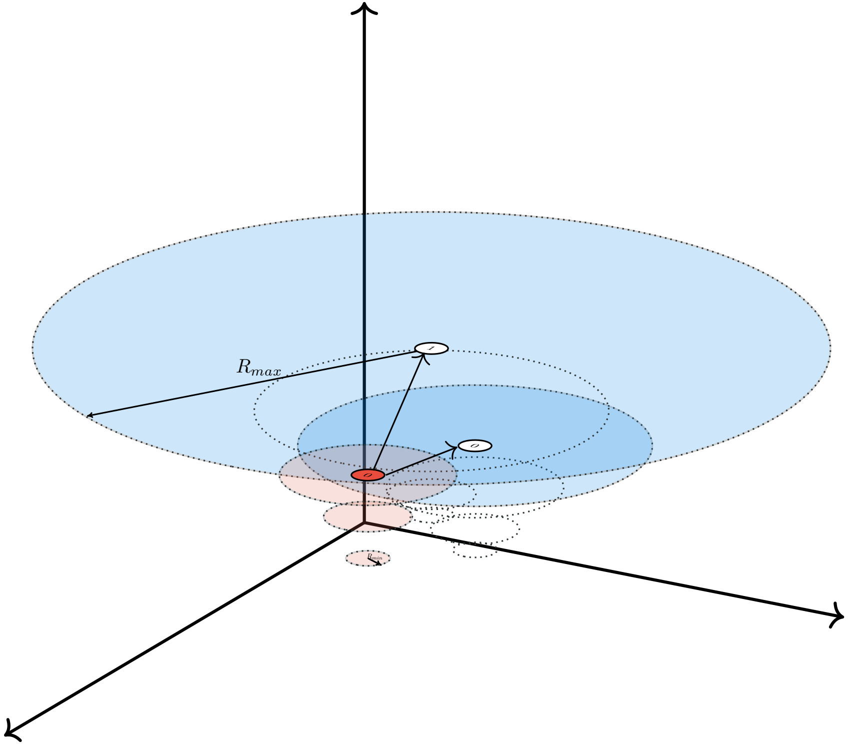
1 node considered

0 node in the cluster

The *cluster* of one node is the set of other nodes that have it in their bunch.

Nyle : Tools

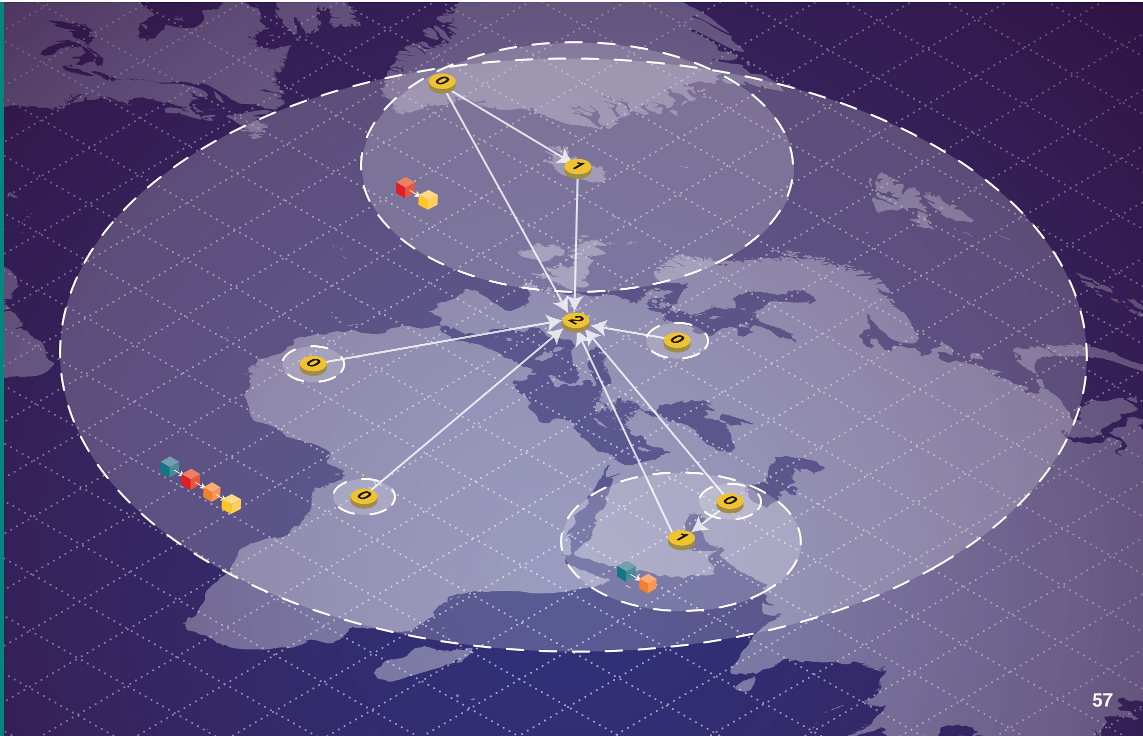
Regions



A Solution : Nyle

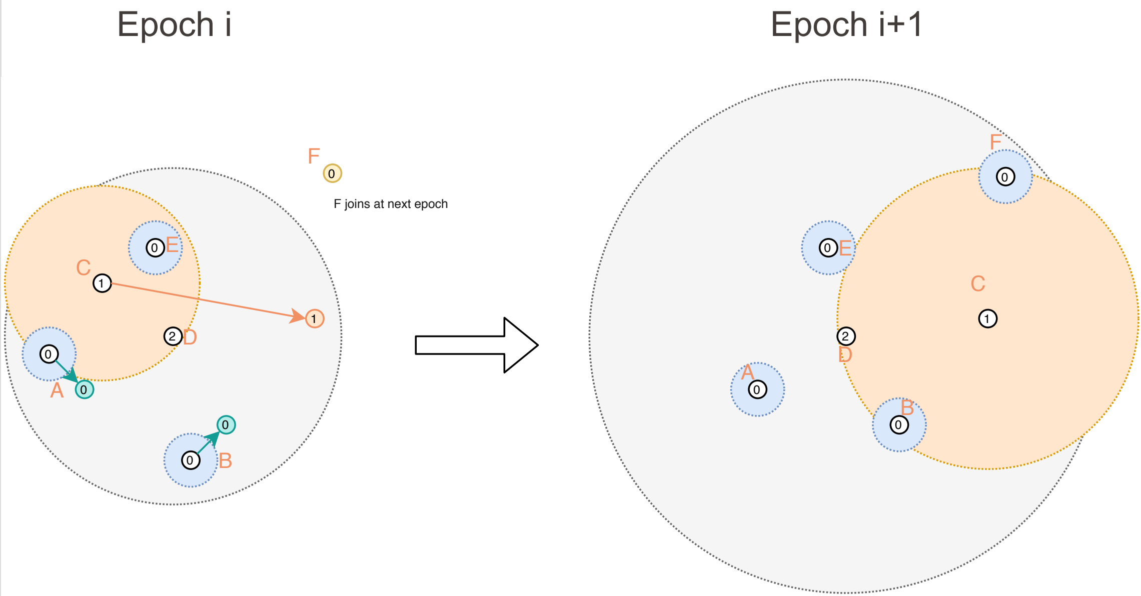
Property

By design : Any two nodes in the system participate within a region with a radius of a small multiple of their *RTT (Round-trip-time)*



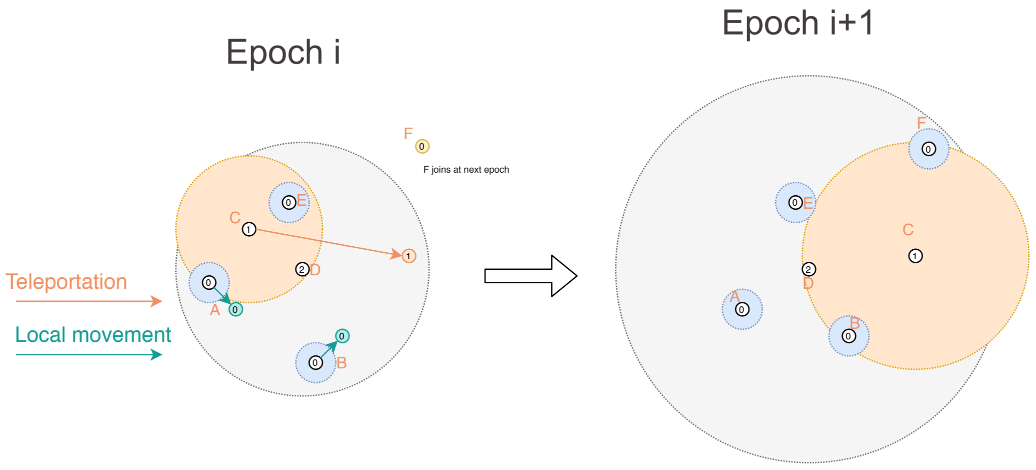
Locarno Treaties : Evaluation - Model

- Nodes are distributed randomly across space
- 10% chance of teleportation at the next epoch
- 20% chance of local movement at the next epoch
- Differences are counted



Locarno Treaties : Evaluation - Model

- Nodes are distributed randomly across space
- 10% chance of teleportation at the next epoch
- 20% chance of local movement at the next epoch
- Differences are counted



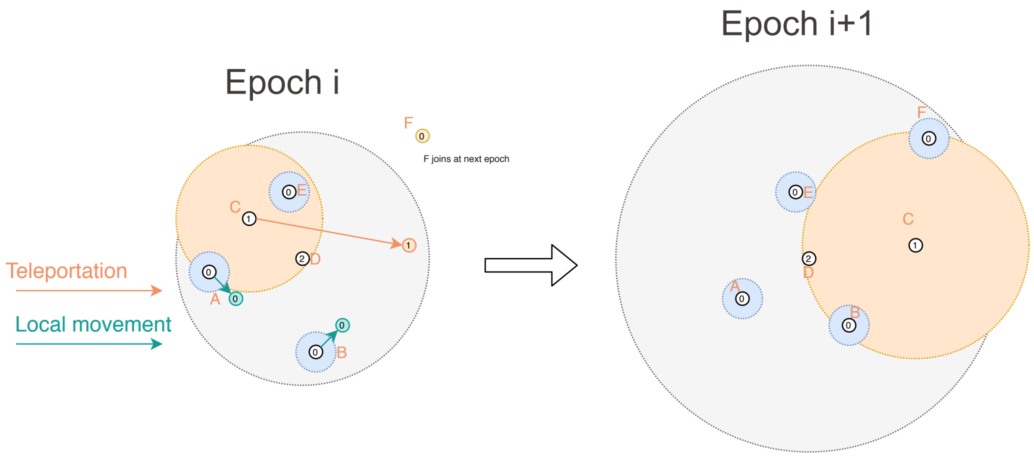
	Bunch	Cluster
A	C, D	-
B	D	-
C	D	A, E
D	-	A, B, C, E
E	C, D	-
F	-	-

8 differences

	Bunch	Cluster
A	~~C~~, D	-
B	D	-
C	D	~~A, E,~~ F
D	-	A, B, C, E, F
E	~~C~~, D	-
F	C,D	-

Locarno Treaties : Evaluation - Model

- Nodes are distributed randomly across space
- 10% chance of teleportation at the next epoch
- 20% chance of local movement at the next epoch
- Differences are counted



Epoch i

	Bunch	Cluster
A	C, D	-
B	D	-
C	D	A, E
D	-	A, B, C, D
E	C, D	-

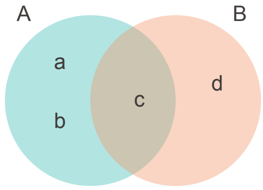
4 differences

Epoch i+1

	Bunch	Cluster
A	~~C~~, D	-
B	D	-
C	D	~~A, E~~
D	-	A, B, C, D
E	~~C~~, D	-

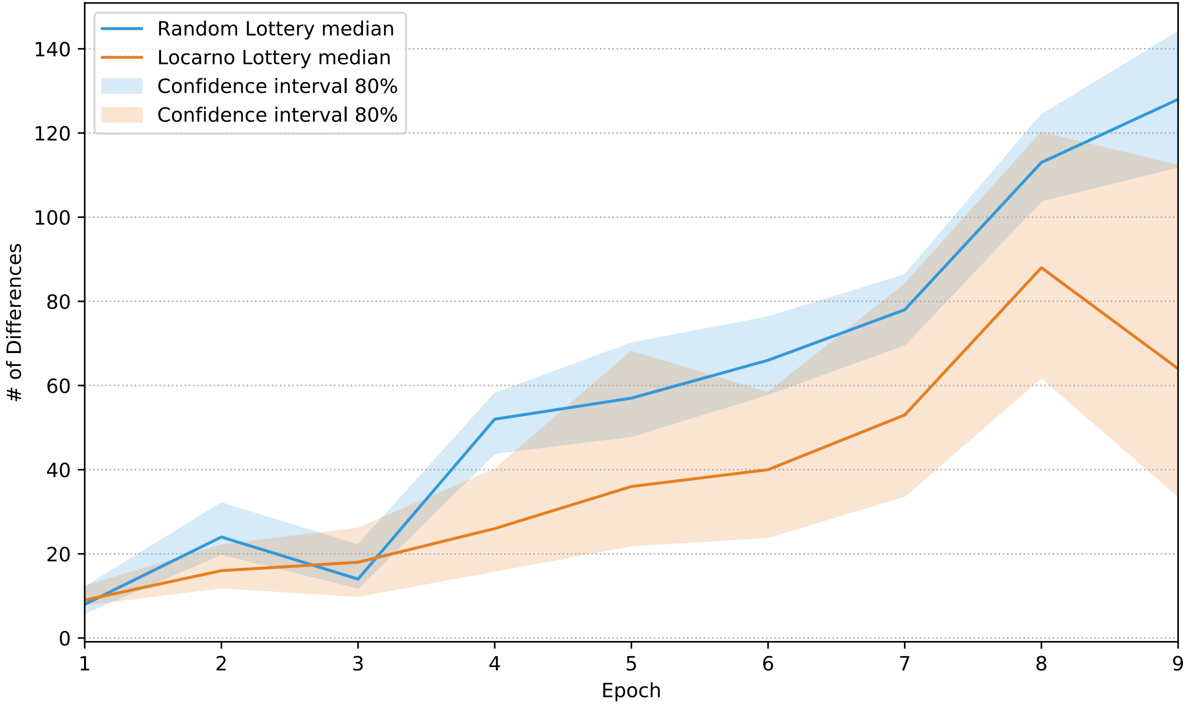
$$\#\text{Diff}(A, B) = \#(A \cup B - A \cap B)$$

A = {a,b,c}



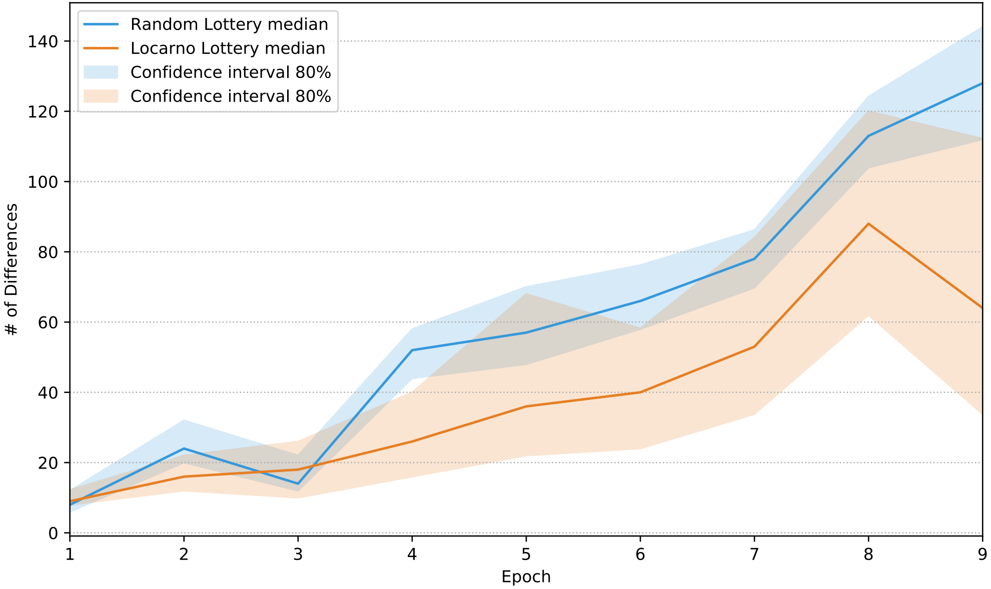
B = {c, d}

Locarno Treaties : Evaluation

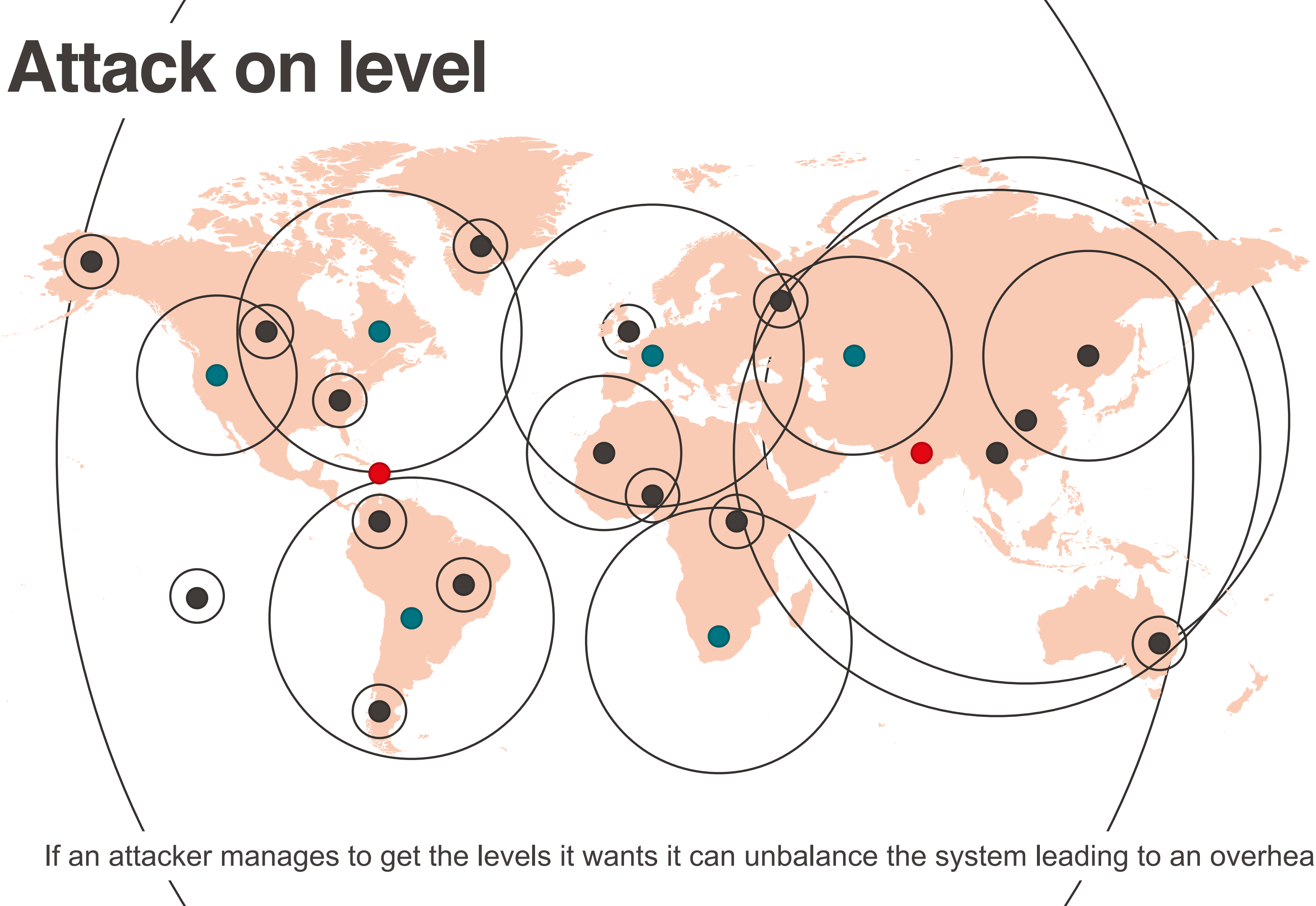


Locarno Treaties : Evaluation

- 100 different experiments using both lotteries
- System starts with 4 nodes, 2 are added at each epoch
- Same evolution for both lotteries
- Locarno Lottery reduces the number of differences
- Variance comes from teleportation

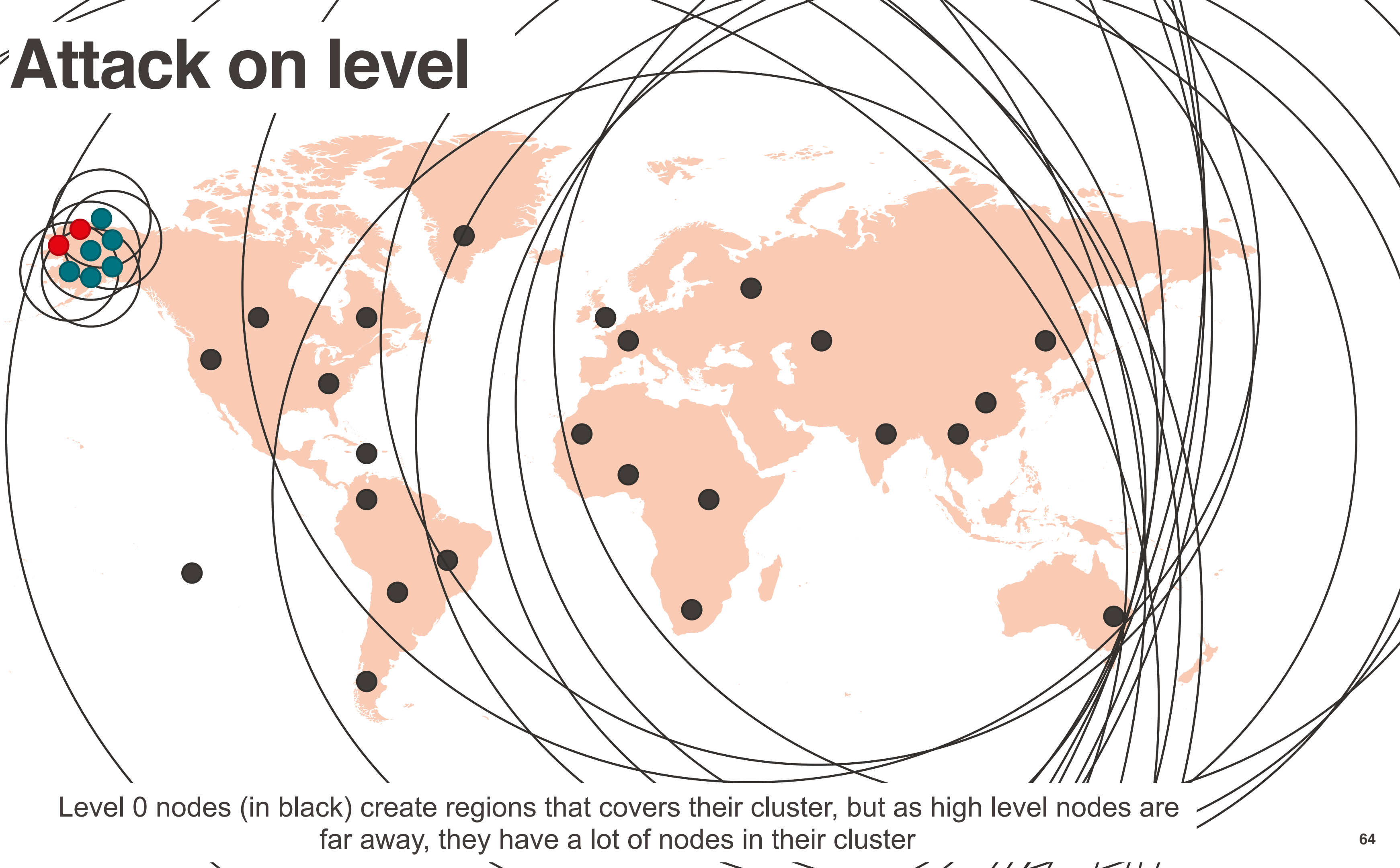


Attack on level



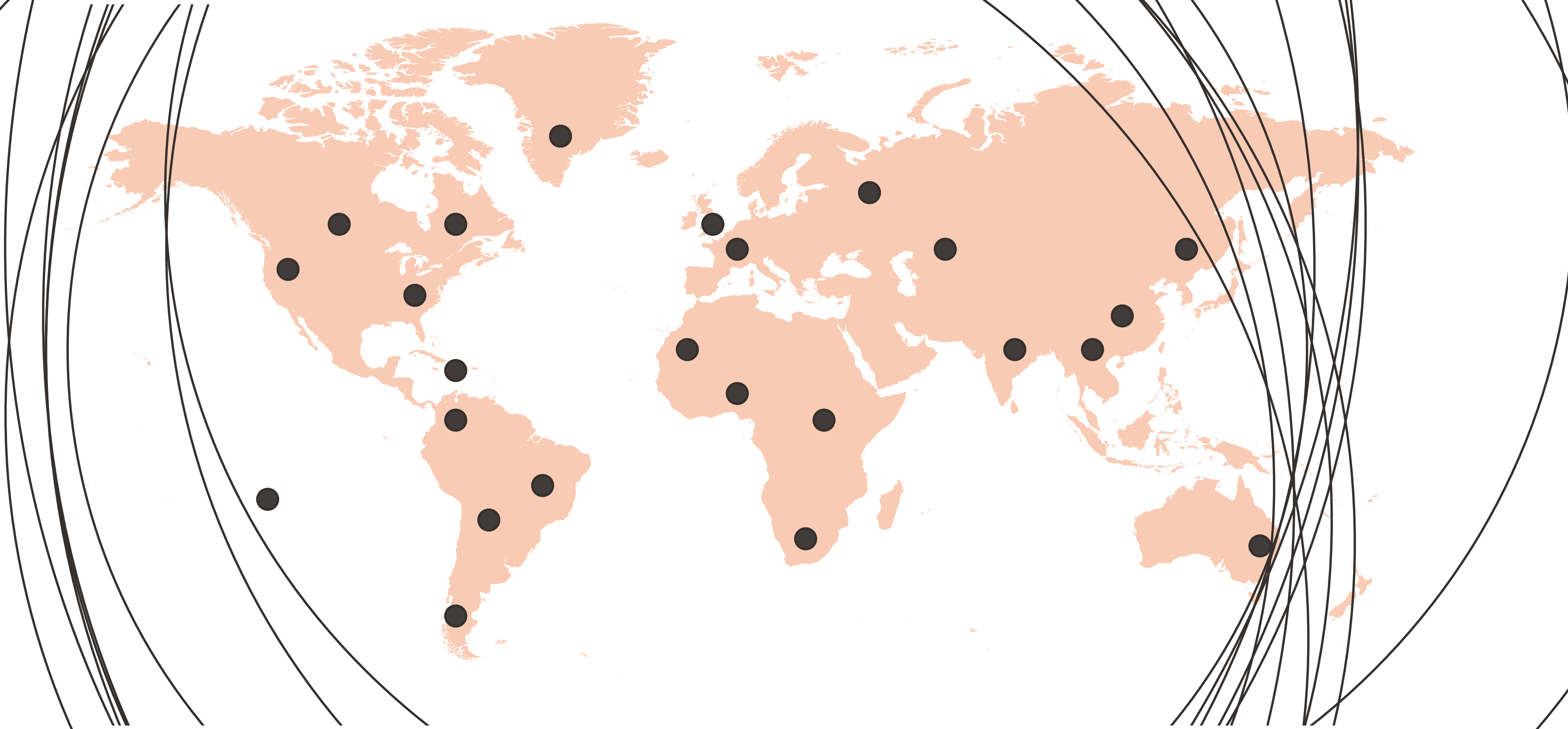
If an attacker manages to get the levels it wants it can unbalance the system leading to an overhead

Attack on level



Level 0 nodes (in black) create regions that covers their cluster, but as high level nodes are far away, they have a lot of nodes in their cluster

Unbalanced levels

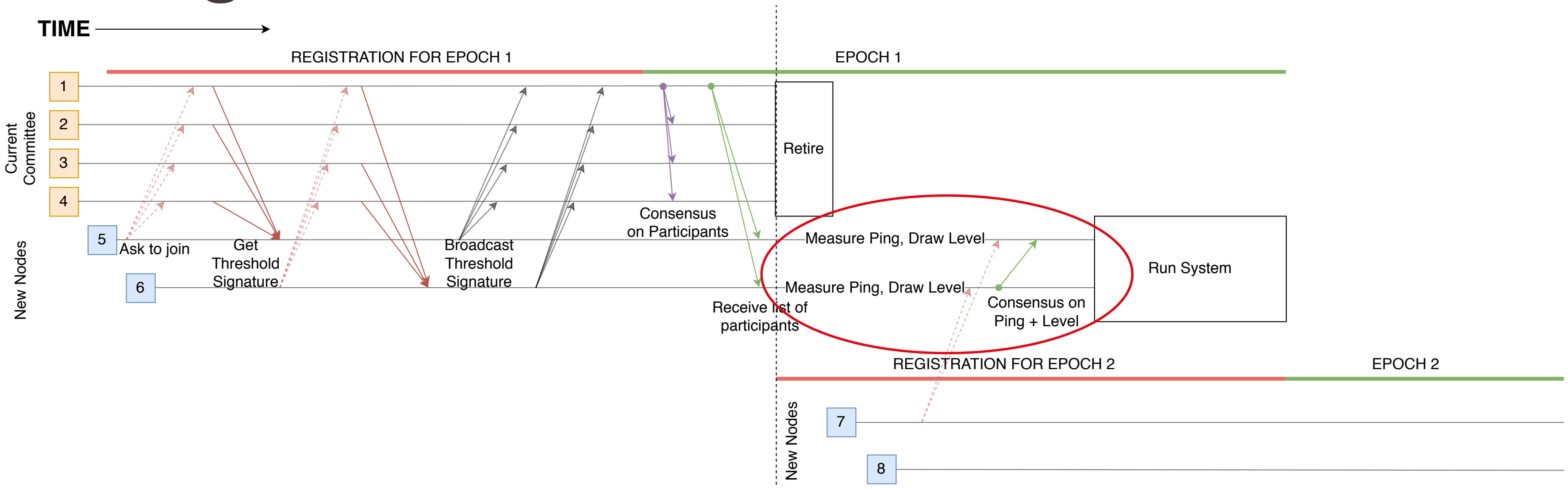


Level 0 nodes (in black) create regions that covers their cluster, but as high level nodes are far away, they have a lot of nodes in their cluster

Fog of the war : Purpose

Nodes do not need to know everything

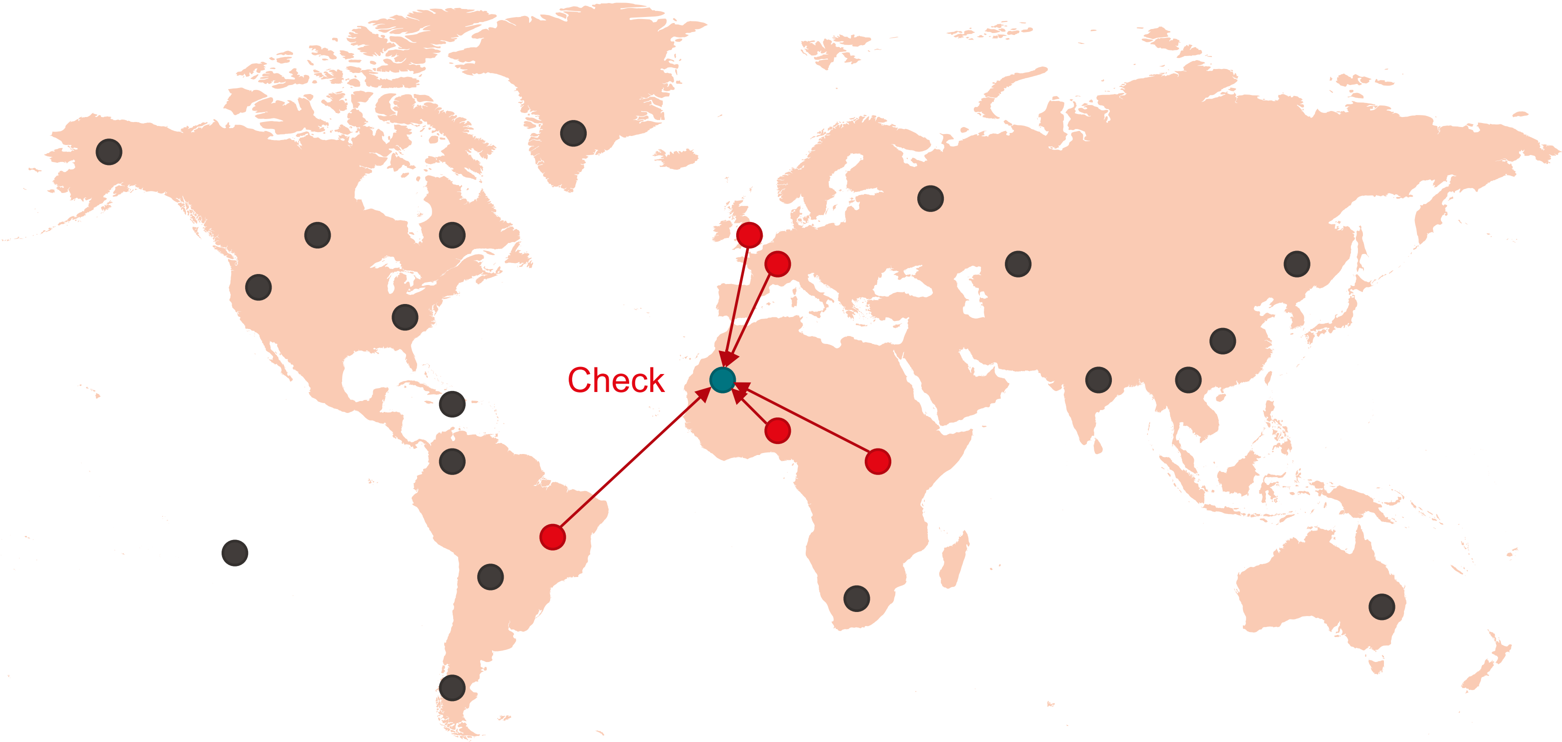
Fog of the war : Idea



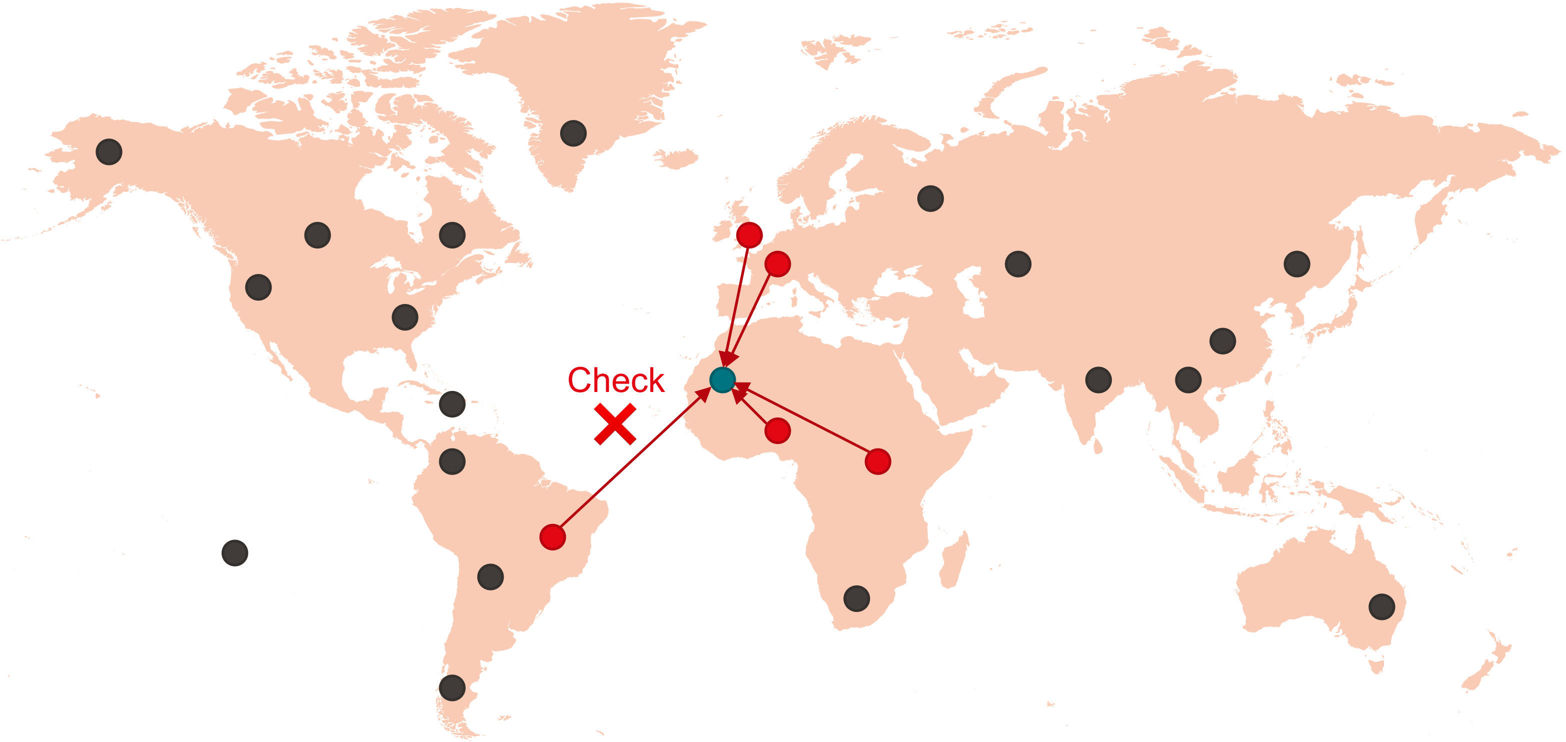
Change measure and consensus on pings with a declared position and a series of checks

Fog of the war : Idea

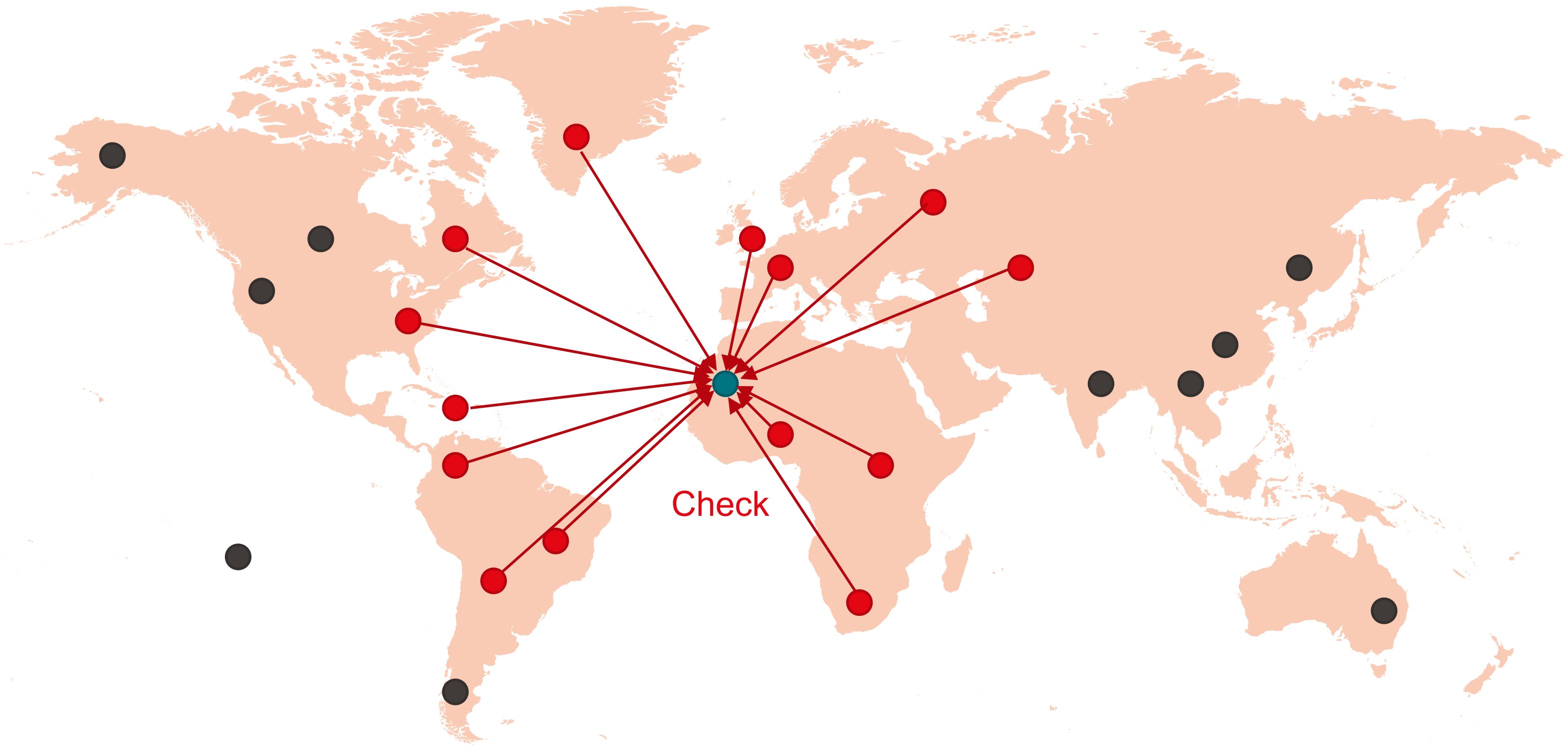
Fog of the war : Idea



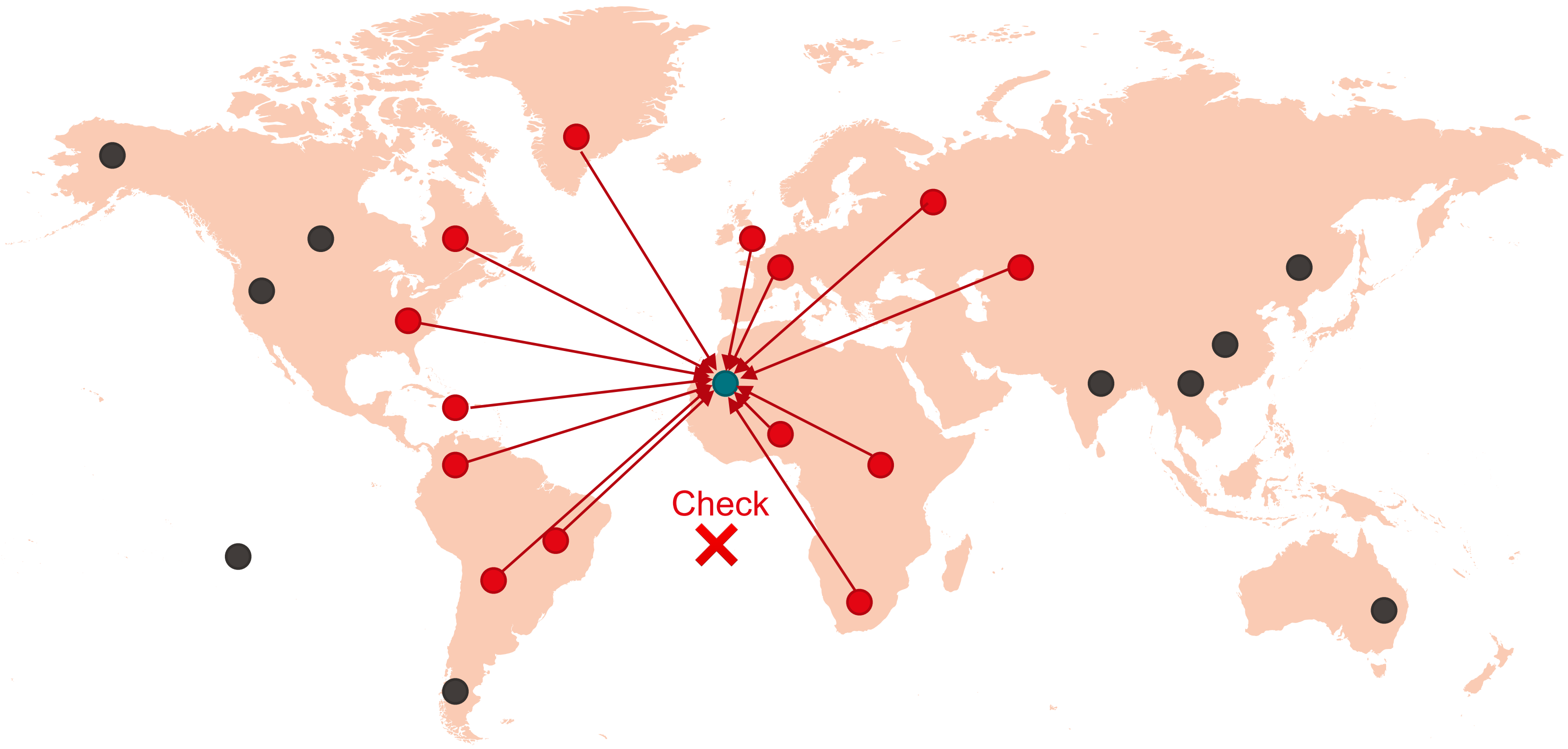
Fog of the war : Idea



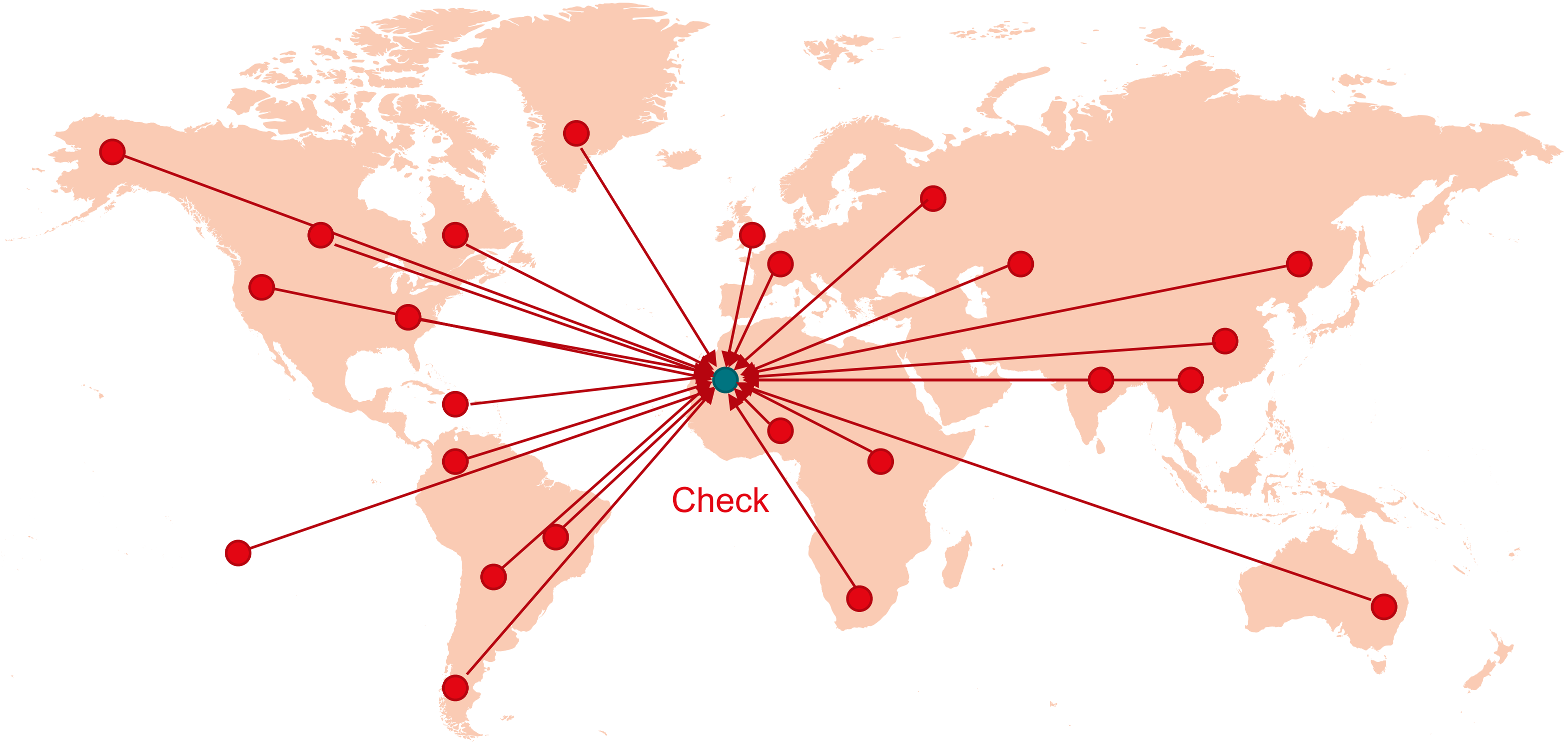
Fog of the war : Idea



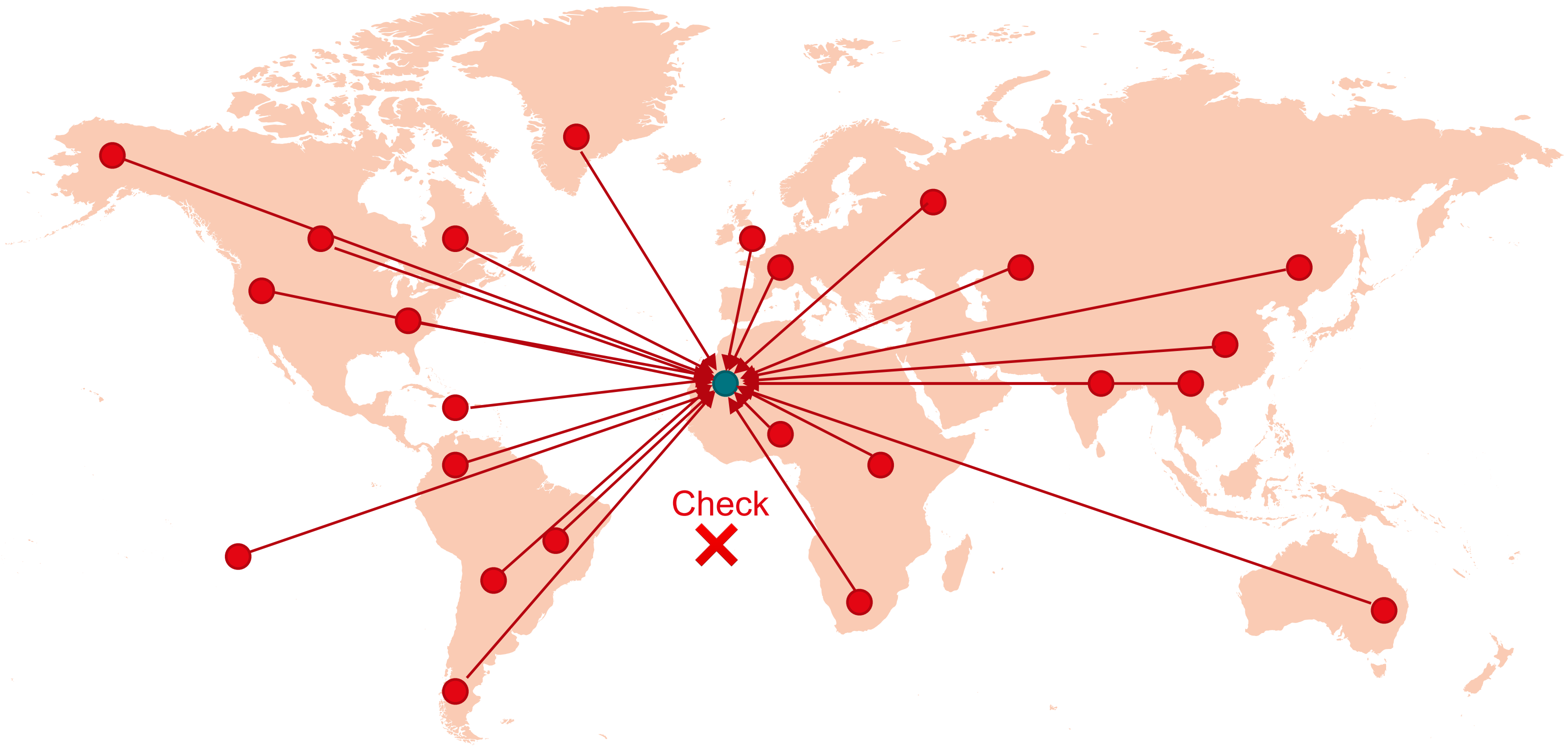
Fog of the war : Idea



Fog of the war : Idea

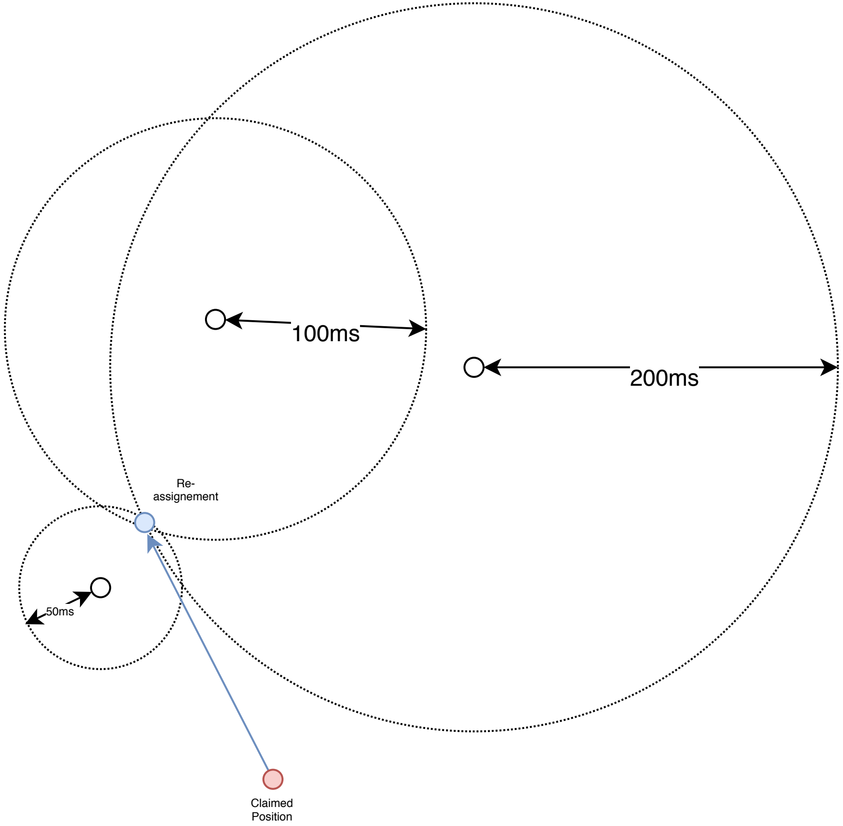


Fog of the war : Idea



If no checks pass

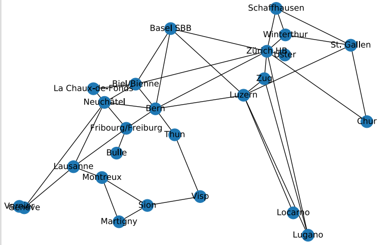
- Assign a new position to the node based on the pings
- A kind of triangulation strategy can be used
- As in Internet-like networks there is triangle inequality violation, this might not be possible
- Could be replaced by the « best candidate » for the position
- Was not implemented

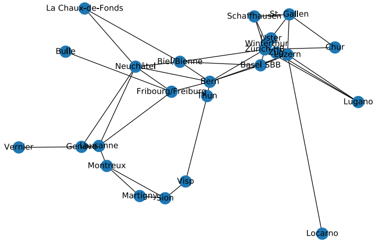


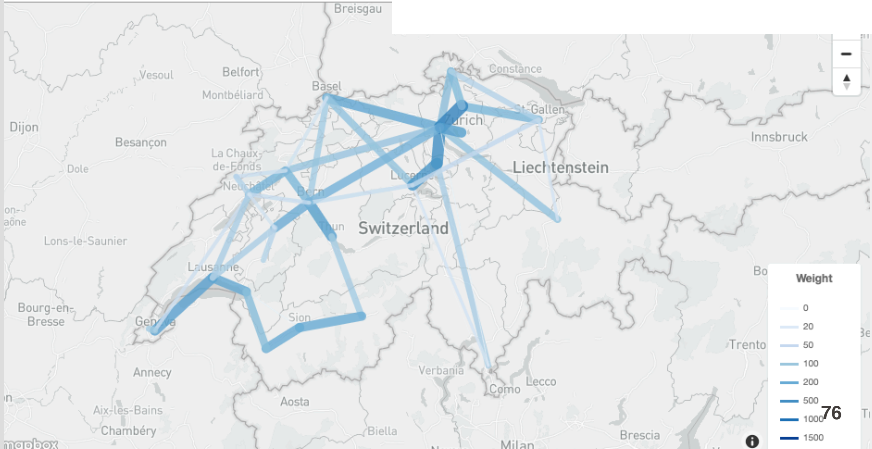
Space Time interaction metric

- Maybe what we want to conserve might not be latency or availability but *interactions* between nodes

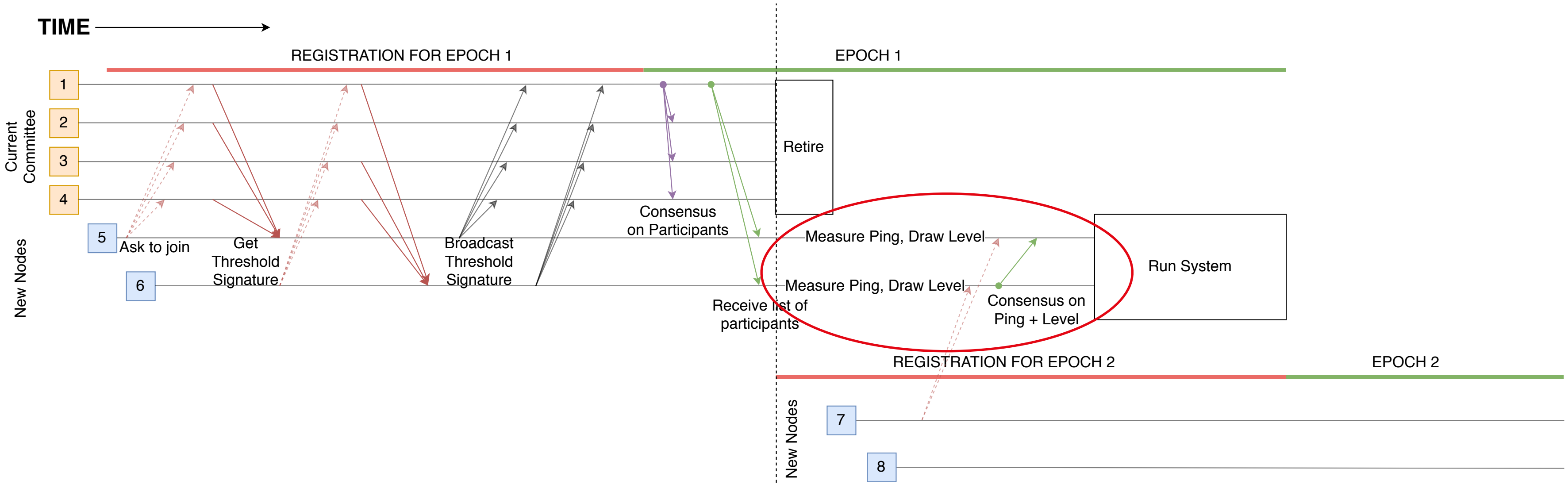
- If there are random partitions, one might want to protect nodes that interact a lot from failing







Space Time interaction metric

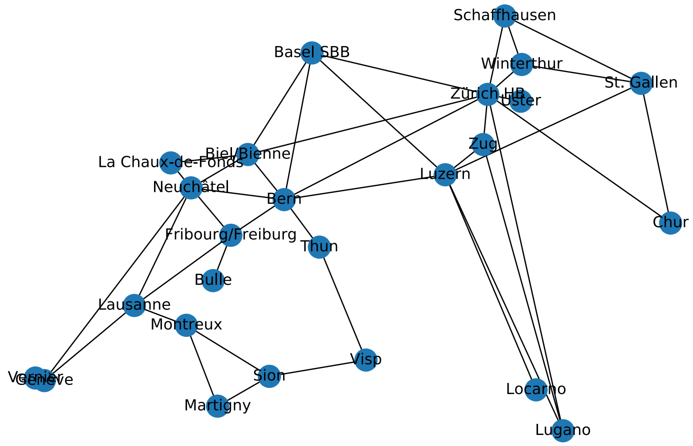


Change ping with a new measure of distance

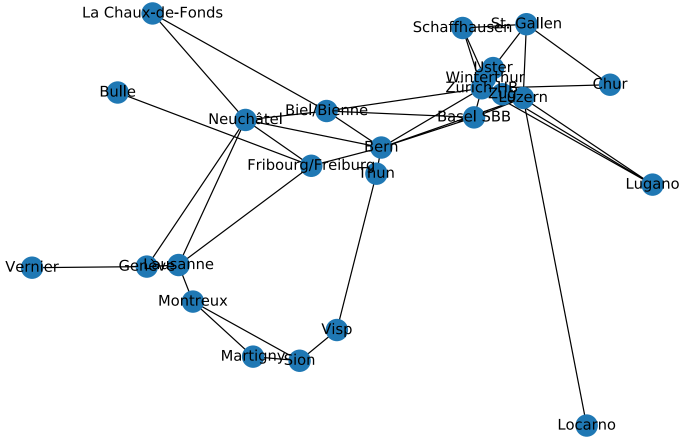
$$d(A, B) = \frac{1}{\text{\# messages between A and B per unit of time}}$$

Each node count each time it interacts with another node during one epoch and publish it at the beginning of the next

Space Time interaction metric explanation



Map using regular distance



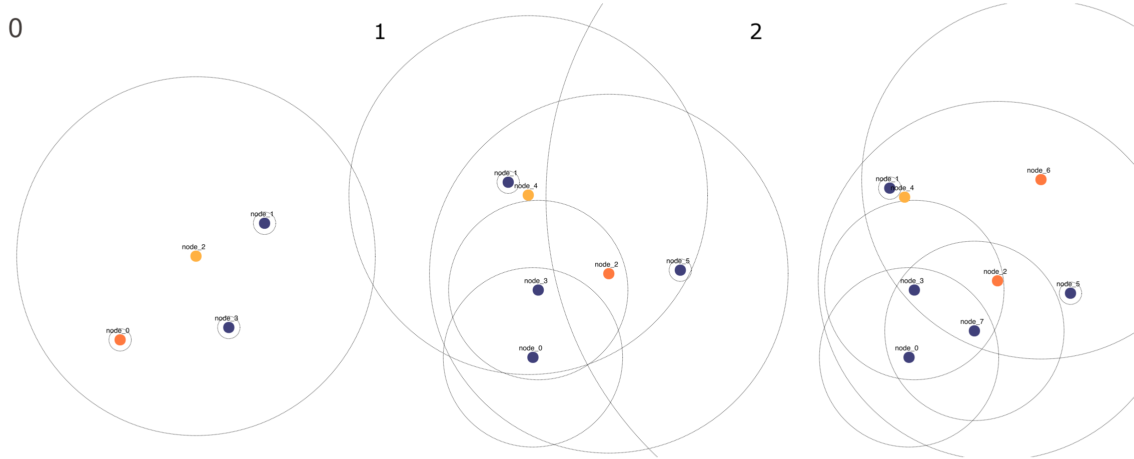
Map using interaction distance

Points are close if there is a lot of connections between them

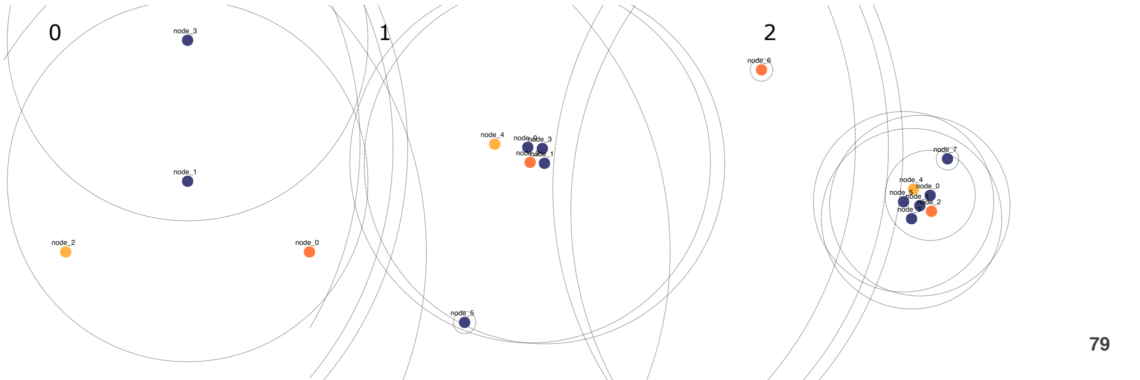
Space Time interaction metric
Drawbacks

- Interactions might change a lot from an epoch to the next
- Might be more complex to conceptualize for an user
- Preserving interactions over availability and latency might be disputable

3 Epochs - Locarno Lottery - Regular distances



3 Epochs - Locarno Lottery - Space Time interaction distances


0
1
2
node_3
node_1
node_2
node_0
node_4
node_0
node_3
node_1
node_5
node_6
node_7
node_4
node_0
node_2