"oFIQUIC: Leveraging QUIC in OSPF for seamless network topology changes"

Rybowski, Nicolas ; Pelsser, Cristel ; Bonaventure, Olivier

ABSTRACT

Link state-routing protocols such as OSPF and ISIS are used in most if not all Internet Service Provider and enterprise networks. They both rely on flooding to distribute the network topology to all routers. Upon topology changes, all routers update their forwarding tables asynchronously which leads to transient events such as micro-loops and packet losses. We propose two improvements to OSPF in an extension called oFIQUIC. First, we use QUIC to exchange routing information between neighboring routers. Second, we revisit the OSPF flooding process. Instead of relying entirely on flooding to distribute topology changes, we establish secure remote QUIC sessions with distant OSPF routers to inform them of topology changes. This enables oFIQUIC to prevent transient loops by ordering the updates of the forwarding tables of all routers after a topology change. We add oFIQUIC to the BIRD implementation of OSPF. Our evaluation demonstrates that oFIQUIC prevents loops and converges quickly in different topologies.

CITE THIS VERSION

Rybowski, Nicolas ; Pelsser, Cristel ; Bonaventure, Olivier. *oFIQUIC: Leveraging QUIC in OSPF for seamless network topology changes.*IFIP Networking (du 03/06/2024 au 06/06/2024). In: *IFIP Networking Conference*, (2024) http://hdl.handle.net/2078.1/286860 -- DOI : 10.23919/IFIPNetworking62109.2024.10619718

oFIQUIC: Leveraging QUIC in OSPF for seamless network topology changes

Nicolas Rybowski, Cristel Pelsser, Olivier Bonaventure
firstname.lastname@uclouvain.be

UCLouvain icteam

wel Research Institute

Agenda

- **Topology changes in IGPs**
- Ordered FIB updates (oFIB)
- oFIQUIC design
- oFIQUIC performances

Seamless topology changes is necessary
Topology changes in IGPs

Planned modification

Unplanned modification

Seamless topology changes is necessary
Topology changes in IGPs

Planned modification

- Adding / removing links / nodes
  - Physically upgrading backbone network
  - Logically reconfiguring optical underlay
    - e.g. add virtual link, disable link, modify available bandwidth

Unplanned modification

Seamless topology changes is necessary

Topology changes in IGPs

Planned modification

- Adding / removing links / nodes
  - Physically upgrading backbone network
  - Logically reconfiguring optical underlay
    - e.g. add virtual link, disable link, modify available bandwidth
- Modifying links metric
  - Reflect link bandwidth
  - Traffic engineering

Unplanned modification

Seamless topology changes is necessary
Topology changes in IGPs

Planned modification

- Adding / removing links / nodes
  - Physically upgrading backbone network
  - Logically reconfiguring optical underlay
    - e.g. add virtual link, disable link, modify available bandwidth
- Modifying links metric
  - Reflect link bandwidth
  - Traffic engineering

Unplanned modification

- Link failures
  - e.g. Fiber cut, failed NIC
- Node failure

Seamless topology changes is necessary
Topology changes in IGPs

Planned modification

- Adding / removing links / nodes
  - Physically upgrading backbone network
  - Logi
    -
- Modifyi
  - Refle
  - Traff

Unplanned modification

- Link failures
  - e.g. Fiber cut, failed NIC

Table 2
About topological changes, a spatial aggregation attempt of listener events (per day statistics).

Types	Average	Median	Max	Total
Link down/up	37.35	6	2074	16,848
Router down/up	0.3	0	8	136
Weight change	0.12	0	4	57

Merindol, P., David, P., Pansiot, J. J., Clad, F., & Vissicchio, S. (2018). A fine-grained multi-source measurement platform correlating routing transitions with packet losses. *Computer Communications*, *129*, 166-183.

Seamless topology changes is necessary

Topology changes in IGPs

Planned modification

- Adding / removing links / nodes
  - Physically upgrading backbone network
  - Logi
    - ▪
- Modifyi
  - Refle
  - Traff

Unplanned modification

- Link failures
  - e.g. Fiber cut, failed NIC
- Node failure

Table 2
About topological changes, a spatial aggregation attempt of listener events (per day statistics).

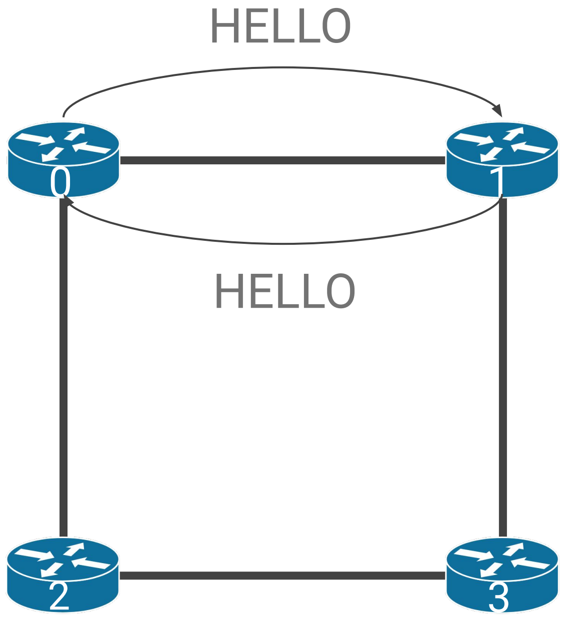
Types	Average	Median	Max	Total
Link down/up	37.35	6	2074	16,848
Router down/up	0.3	0	8	136
Weight change	0.12	0	4	57

Issue: IGPs not designed for frequent changes

Merindol, P., David, P., Pansiot, J. J., Clad, F., & Vissicchio, S. (2018). A fine-grained multi-source measurement platform correlating routing transitions with packet losses. *Computer Communications*, *129*, 166-183.

OSPF discovers neighbors
IGP remainder

1. Neighbor discovery
   (2-way connectivity check)

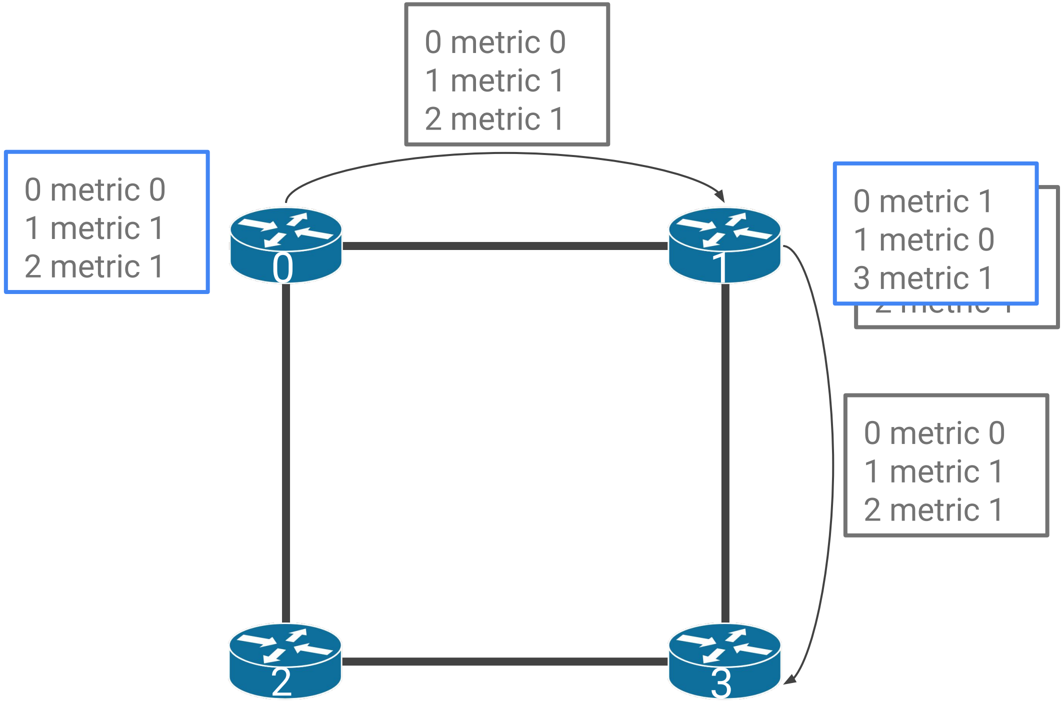


All metrics are unitary

OSPF floods Link State Advertisments (LSAs)

IGP remainder

1. Neighbor discovery (2-way connectivity check)
2. Flood Link State Advertisement (LSAs)

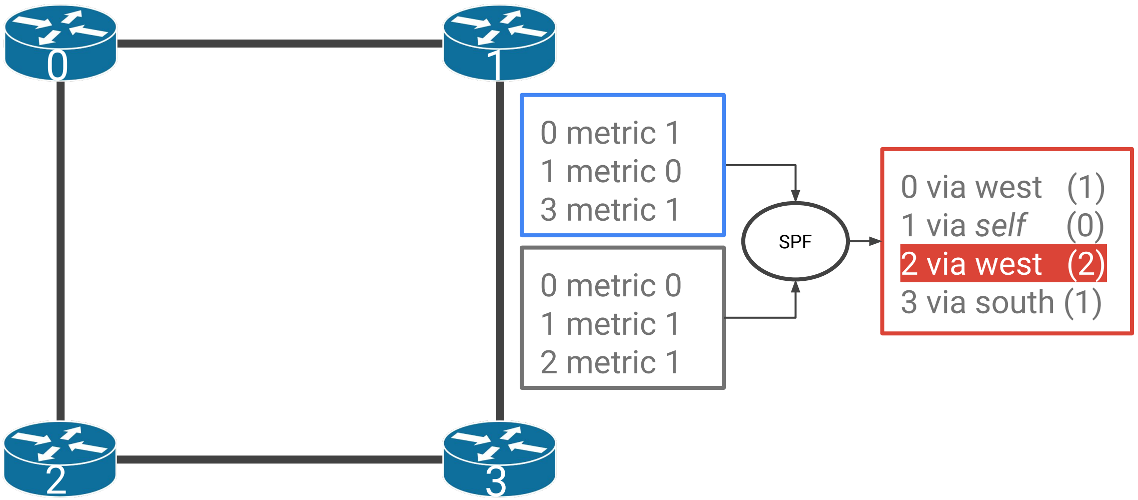


All metrics are unitary

OSPF computes the shortest paths

IGP remainder

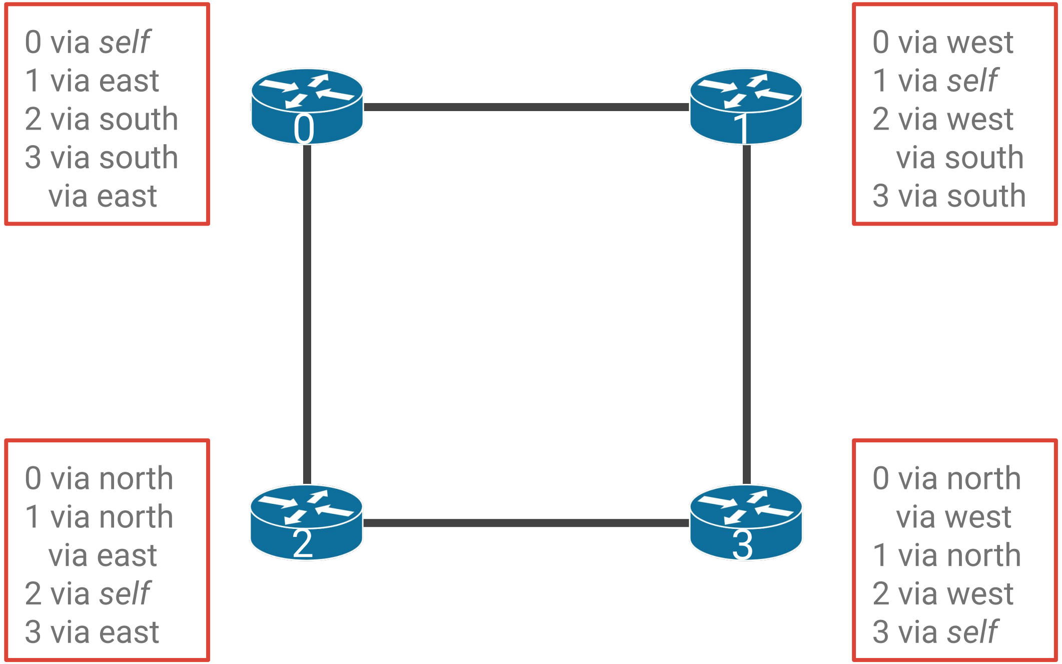
1. Neighbor discovery (2-way connectivity check)
2. Flood Link State Advertisement (LSAs)
3. Shortest paths computation



All metrics are unitary

OSPF converges

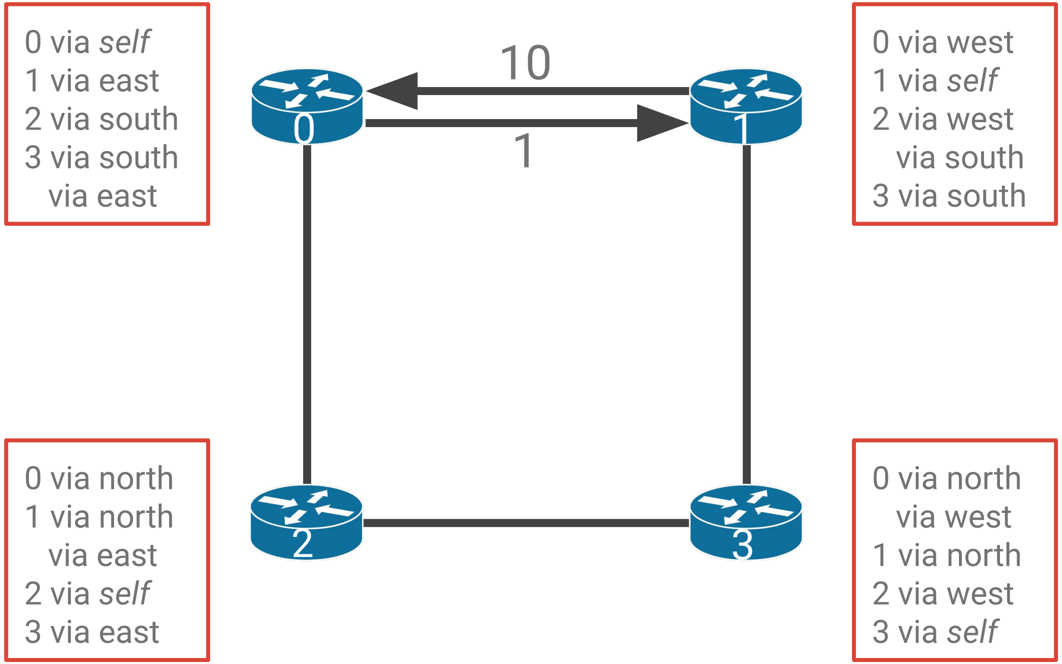
IGP remainder



Increasing link metric

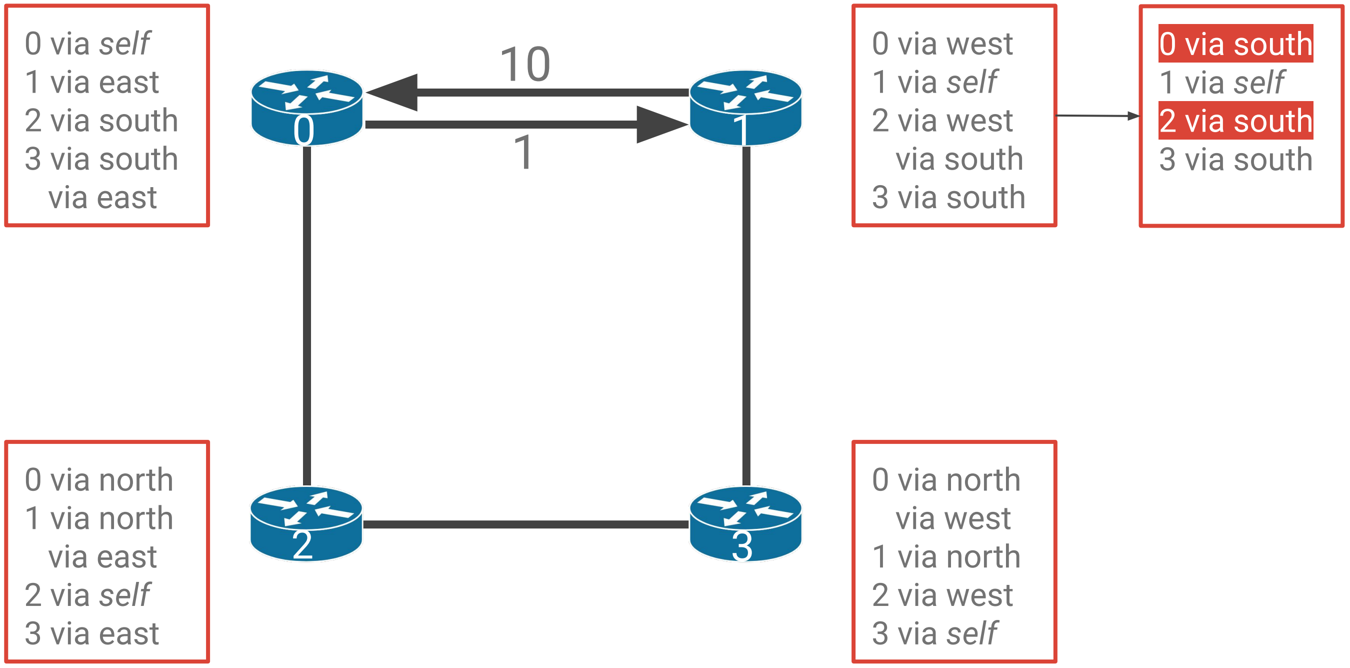
Topology changes in IGPs

Virtually remove a link with max metric - 1



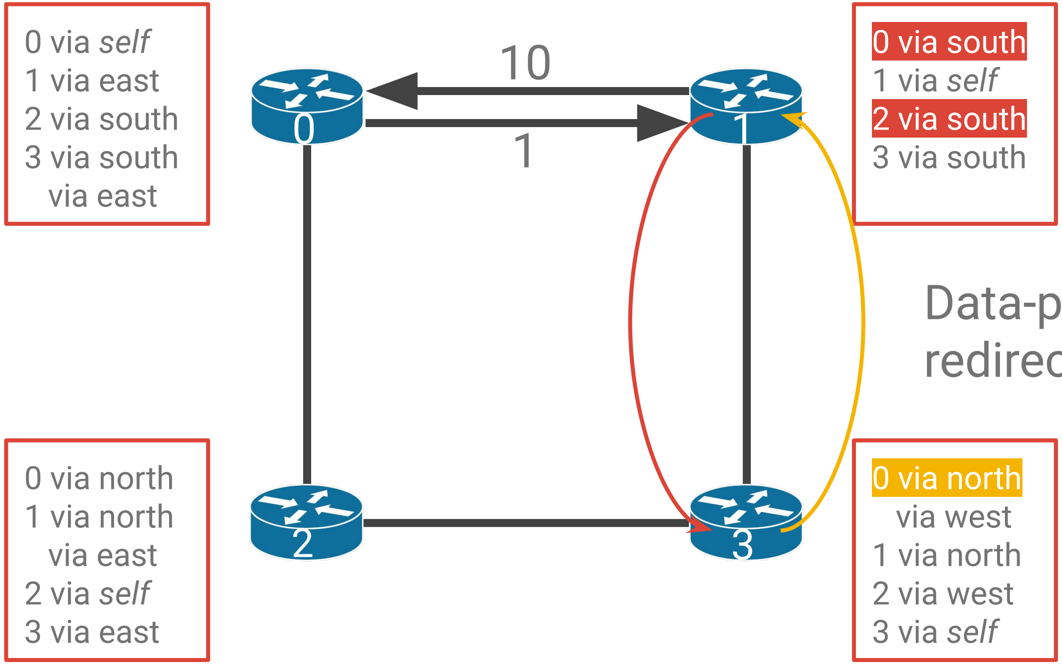
Increasing link metric

Topology changes in IGPs



Increasing link metric triggers micro-loops

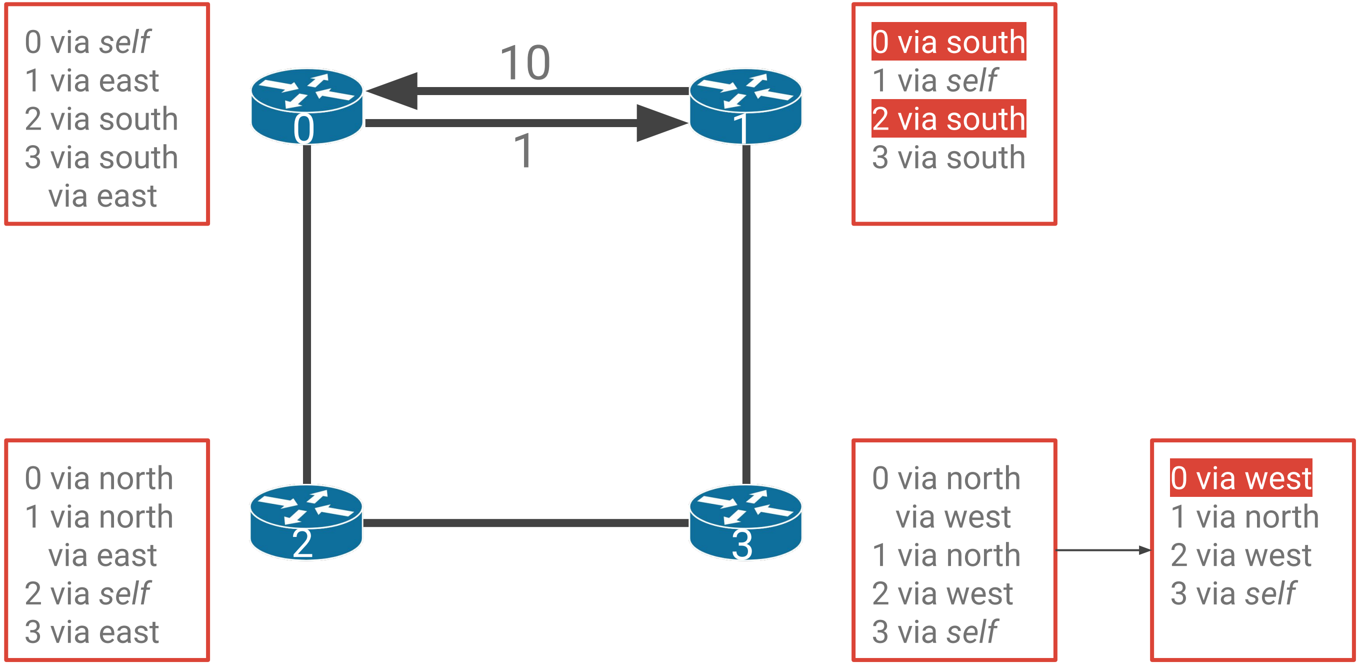
Topology changes in IGPs



Data-plane traffic is redirected

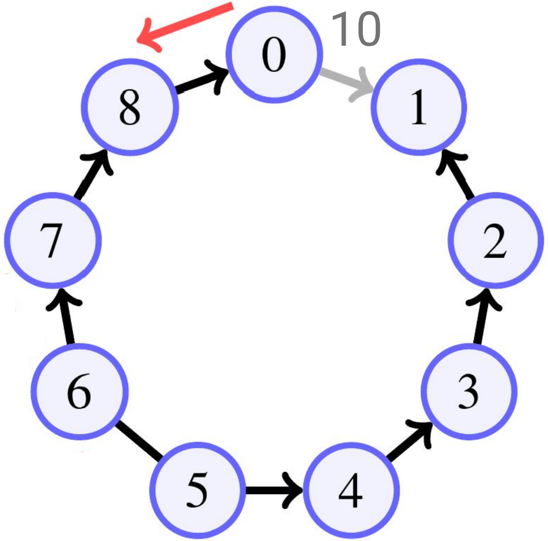
Increasing link metric triggers micro-loops

Topology changes in IGPs



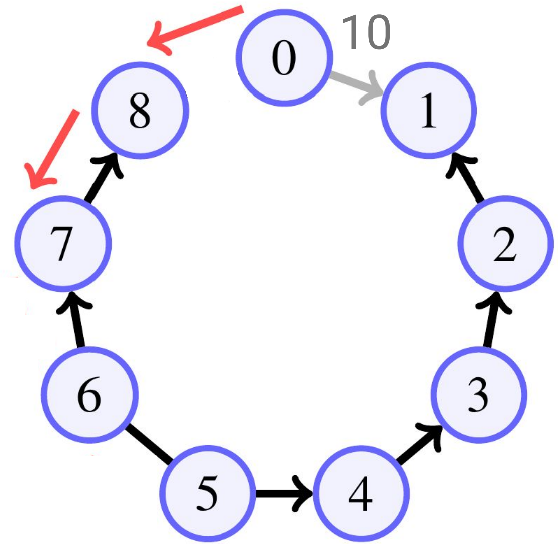
Micro-loops propagate
Topology changes in IGPs

- Unitary link metric



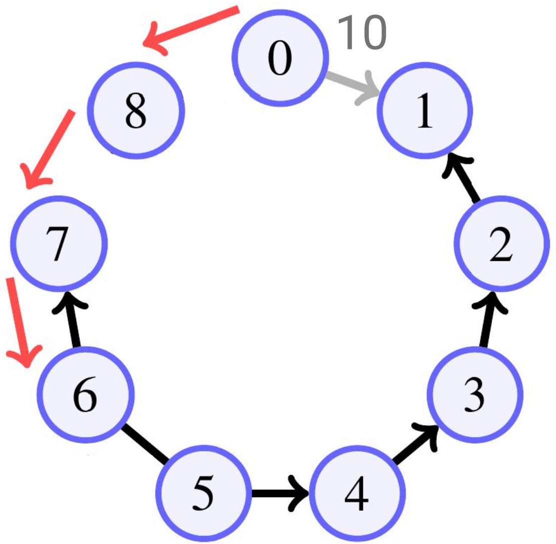
Micro-loops propagate
Topology changes in IGPs

- Unitary link metric



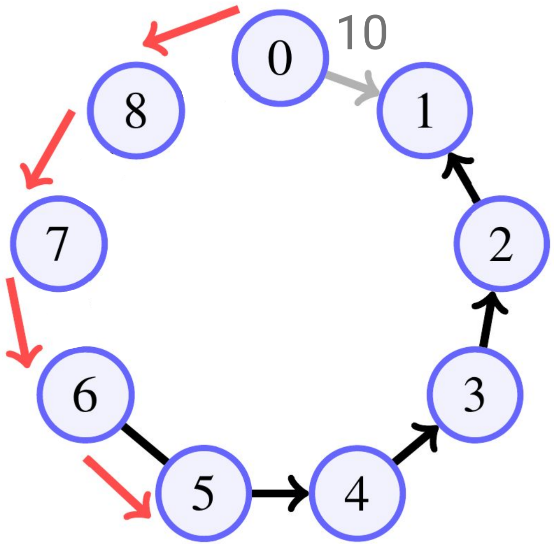
Micro-loops propagate
Topology changes in IGPs

- Unitary link metric



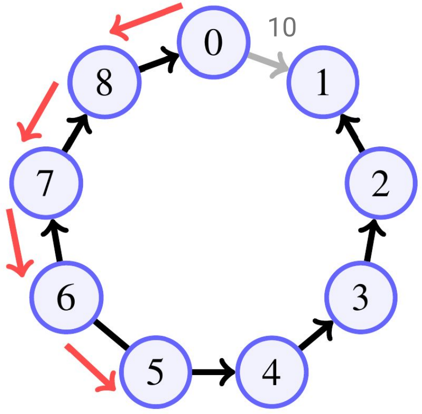
Micro-loops propagate
Topology changes in IGPs

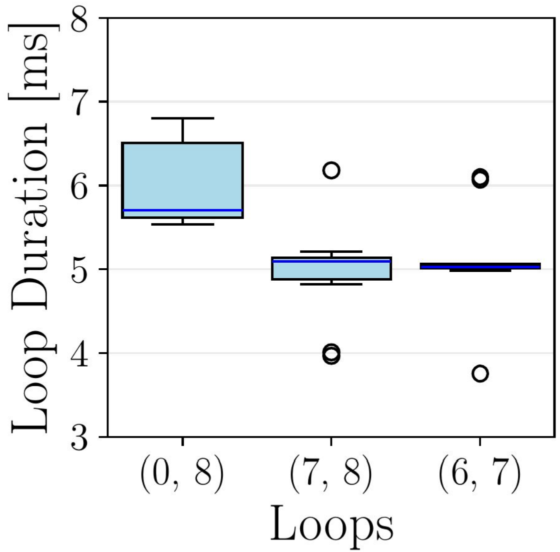
- Unitary link metric



Measuring micro-loops duration
Topology changes in IGPs

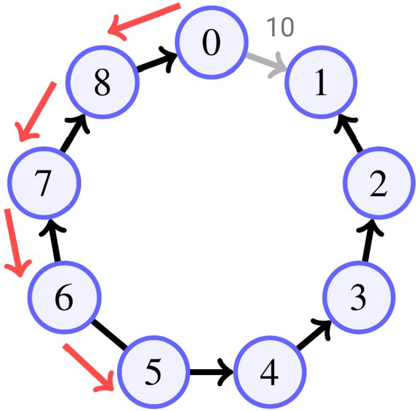
- Unitary link metric
- 5ms link delay

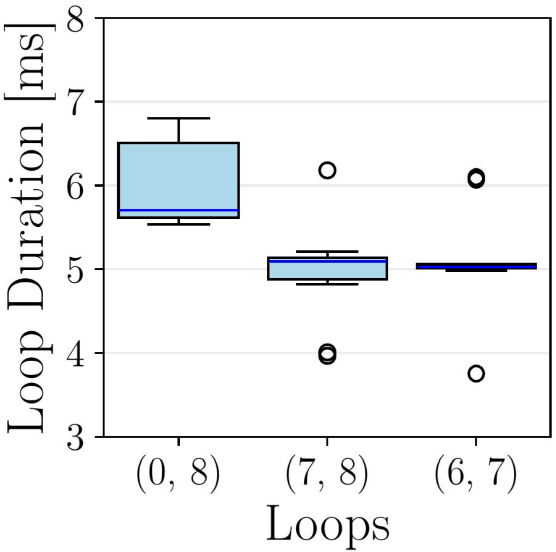




Measuring micro-loops duration
Topology changes in IGPs

- Unitary link metric
- 5ms link delay

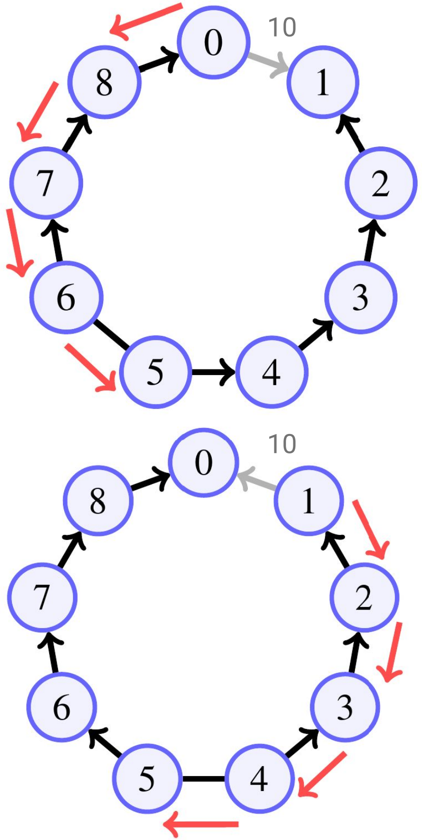


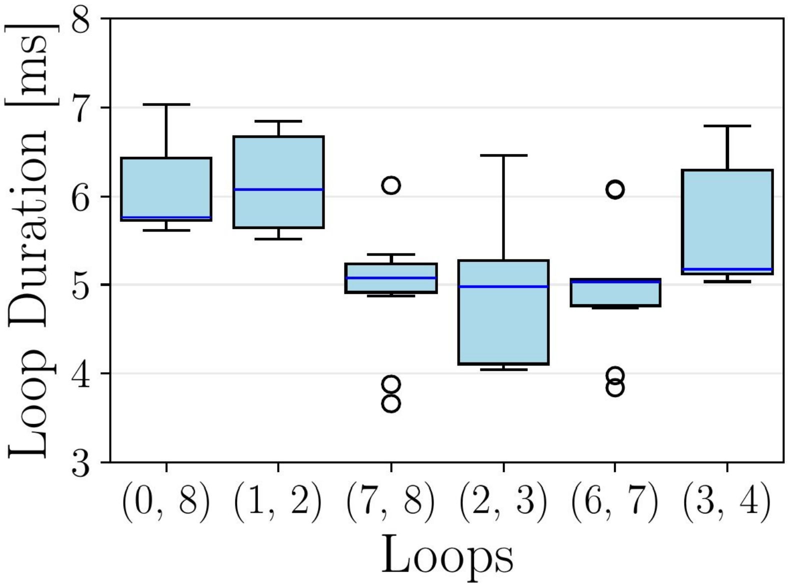


- Micro-loop duration factors
  - Link delay
  - # prefixes
    - SPF computation
    - FIB injection duration (injection rate)
  - …

Symmetric update increases the number of loops
Topology changes in IGPs

- Unitary link metric
- 5ms link delay





Impact of micro-loops

Topology changes in IGPs

- Packet reordering

Impact of micro-loops
Topology changes in IGPs

- Packet reordering
- TTL reaches 0 if micro-loop lasts long enough => transient black-hole

Impact of micro-loops

Topology changes in IGPs

- Packet reordering
- TTL reaches 0 if micro-loop lasts long enough => transient black-hole

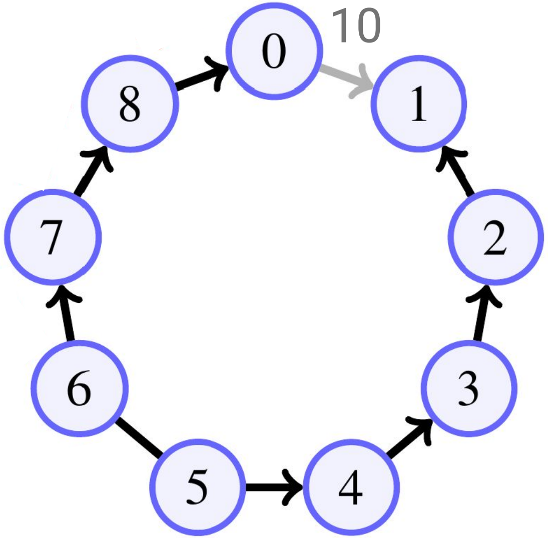
IGPs trigger traffic disruption by design

Agenda

- Topology changes in IGPs
- **Ordered FIB updates (oFIB)**
- oFIQUIC design
- oFIQUIC performances

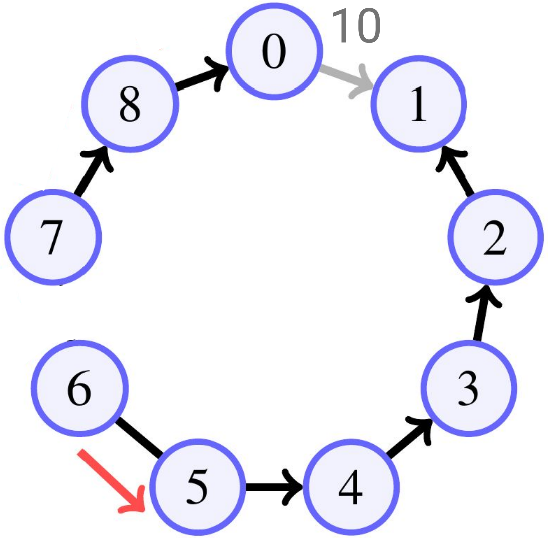
Ordering FIB updates suppresses micro-loops

Ordered FIB updates (oFIB)



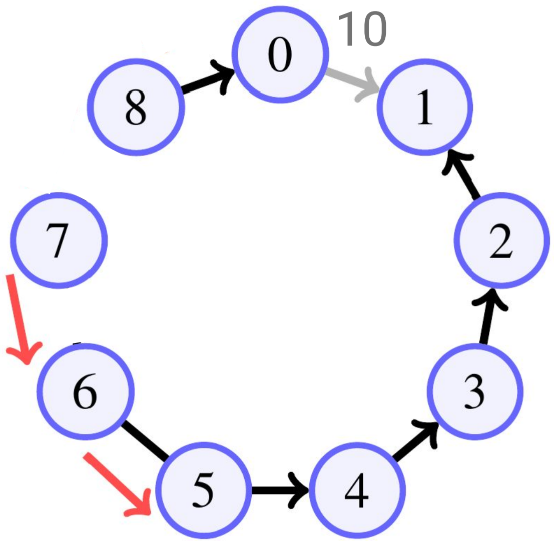
Ordering FIB updates suppresses micro-loops

Ordered FIB updates (oFIB)



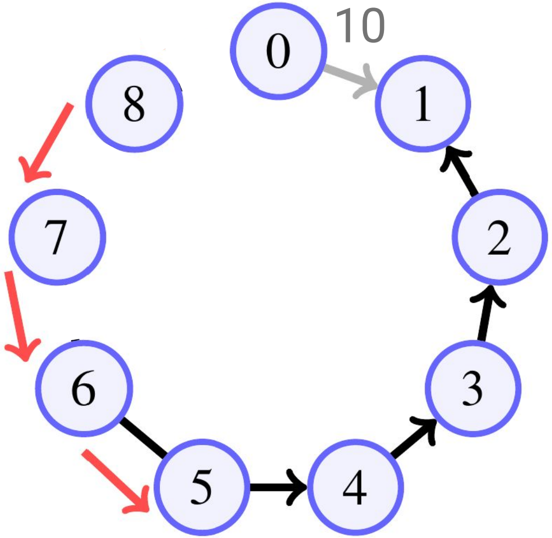
Ordering FIB updates suppresses micro-loops

Ordered FIB updates (oFIB)



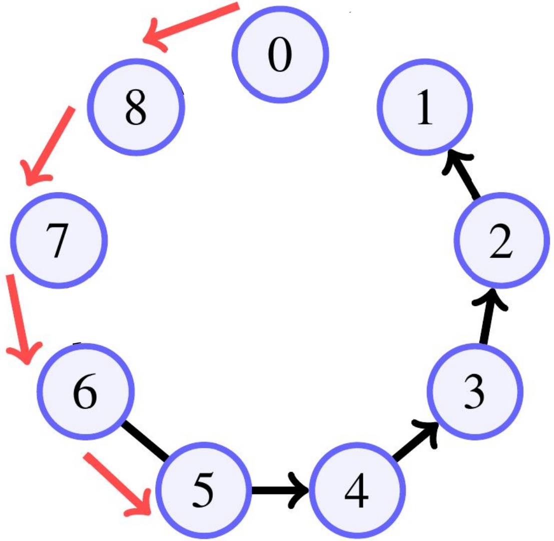
Ordering FIB updates suppresses micro-loops

Ordered FIB updates (oFIB)



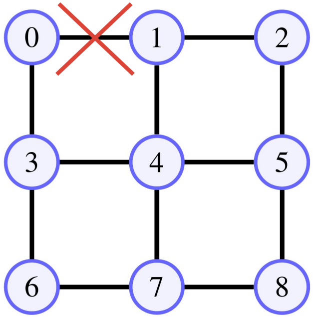
Ordering FIB updates suppresses micro-loops

Ordered FIB updates (oFIB)



Defining FIB update ordering

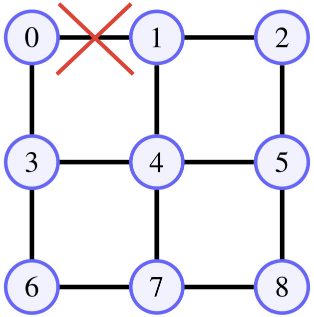
Ordered FIB updates (oFIB)

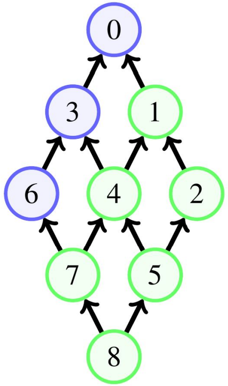


[1] François, P., & Bonaventure, O. (2007). Avoiding transient loops during the convergence of link-state routing protocols. *IEEE/ACM Transactions On Networking*, *15*(6), 1280-1292.

Defining FIB update ordering
Ordered FIB updates (oFIB)

i. Compute Reverse Shortest Path Directed Acyclic Graph (rspDAG) rooted at R0 [1]



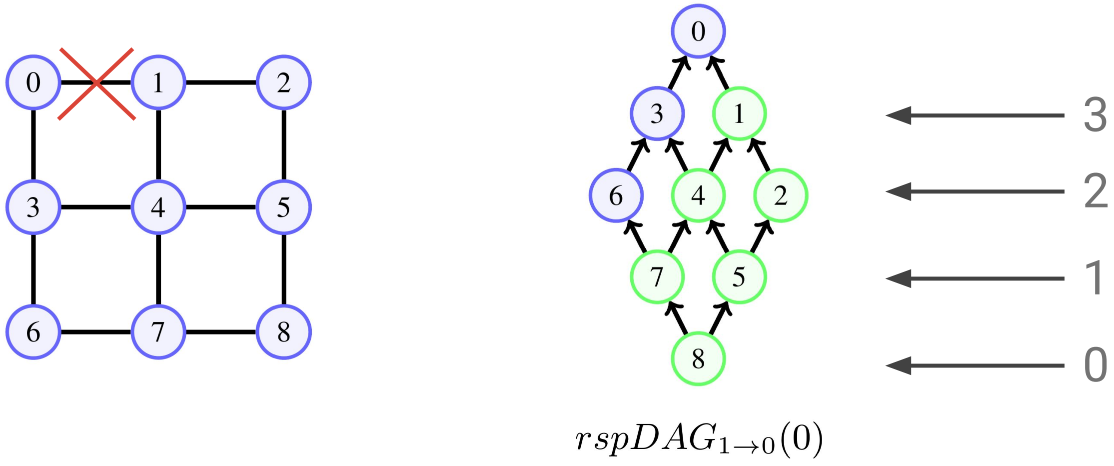


$rspDAG_{1 \to 0}(0)$

[1] François, P., & Bonaventure, O. (2007). Avoiding transient loops during the convergence of link-state routing protocols. *IEEE/ACM Transactions On Networking*, *15*(6), 1280-1292.

Defining FIB update ordering
Ordered FIB updates (oFIB)

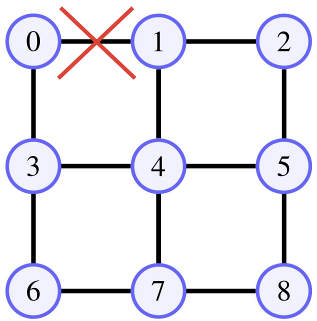
ii. Attribute a rank to each node in the sub-tree [1]

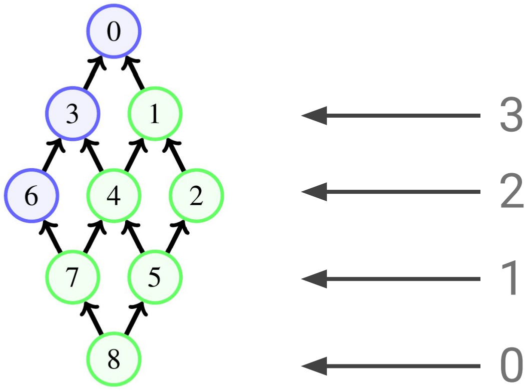


[1] François, P., & Bonaventure, O. (2007). Avoiding transient loops during the convergence of link-state routing protocols. *IEEE/ACM Transactions On Networking*, *15*(6), 1280-1292.

Defining FIB update ordering
Ordered FIB updates (oFIB)

iii. Associate a timer to each node in the sub-tree [1]
T(R) = Rank(R) x MAX_FIB_TIME
=> Update FIB at timer expiration





$rspDAG_{1 \to 0}(0)$

[1] François, P., & Bonaventure, O. (2007). Avoiding transient loops during the convergence of link-state routing protocols. *IEEE/ACM Transactions On Networking*, *15*(6), 1280-1292.

Drawbacks of time-based ordering
Ordered FIB updates (oFIB)

- Worst-case timer delays whole convergence
  - Possible shortcut with completion messages [1]
- Distributed computation
- Asynchronous update

=> not suitable for frequent changes

[1] François, P., & Bonaventure, O. (2007). Avoiding transient loops during the convergence of link-state routing protocols. *IEEE/ACM Transactions On Networking*, *15*(6), 1280-1292.

Agenda

- Topology changes in IGPs
- Ordered FIB updates (oFIB)
- **oFIQUIC design**
- oFIQUIC performances

Connecting remote IGP peers

oFIQUIC design

- Scheduling node
  - Ordering computed once
  - Timer-less synchronous updates

Connecting remote IGP peers
oFIQUIC design

- Scheduling node
  - Ordering computed once
  - Timer-less synchronous updates
- Requirements
  - Multi-hop transport layer
  - Authentication (Authorization)

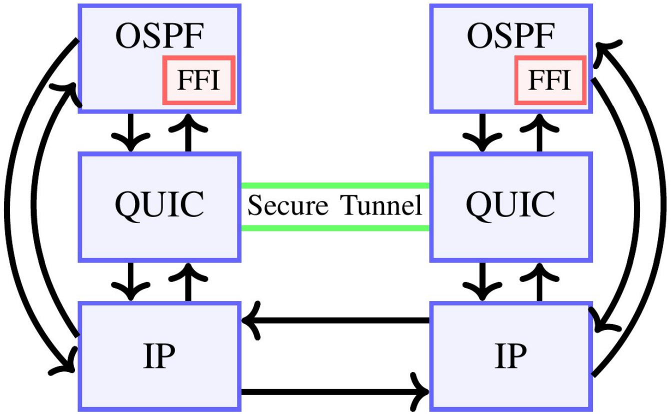
Connecting remote IGP peers

oFIQUIC design

- Scheduling node
  - Ordering computed once
  - Timer-less synchronous updates
- Requirements
  - Multi-hop transport layer
  - Authentication (Authorization)
- Leverage QUIC transport
  - Complete transport protocol based on UDP
  - Native support for TLS (mutual TLS)

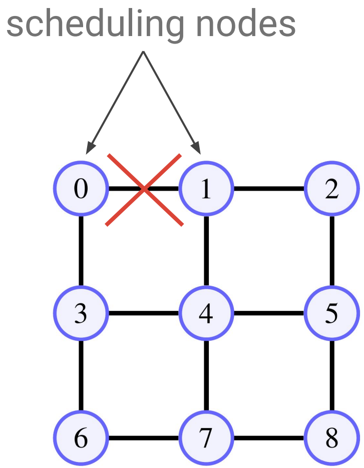
Connecting remote IGP peers
oFIQUIC design

- Scheduling node
    - Ordering computed once
    - Timer-less synchronous updates
- Requirements
    - Multi-hop transport layer
    - Authentication (Authorization)
- Leverage QUIC transport
    - Complete transport protocol based on UDP
    - Native support for TLS (mutual TLS)



Ordering FIB updates with oFIQUIC

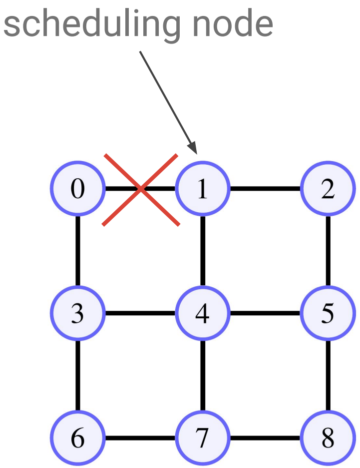
oFIQUIC design

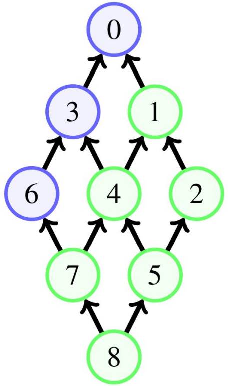


Ordering FIB updates with oFIQUIC

oFIQUIC design

1. Compute rspDAG



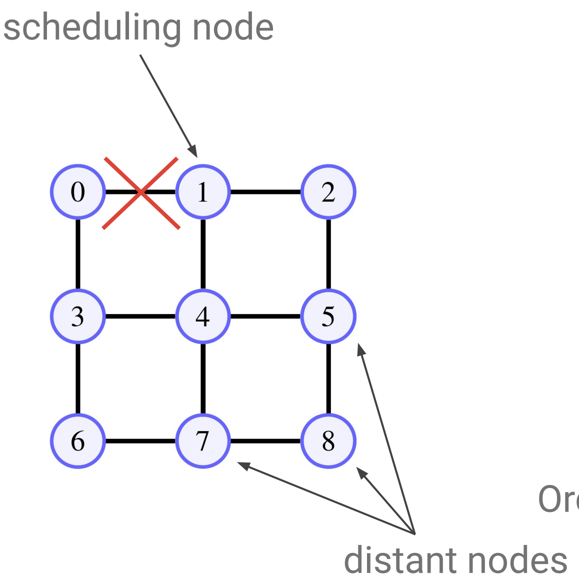


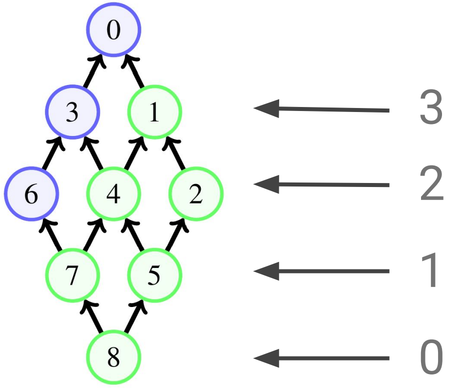
$rspDAG_{1\to 0}(0)$

Ordering FIB updates with oFIQUIC

oFIQUIC design

1. Compute rspDAG
2. Derive an ordered list of (distant) nodes



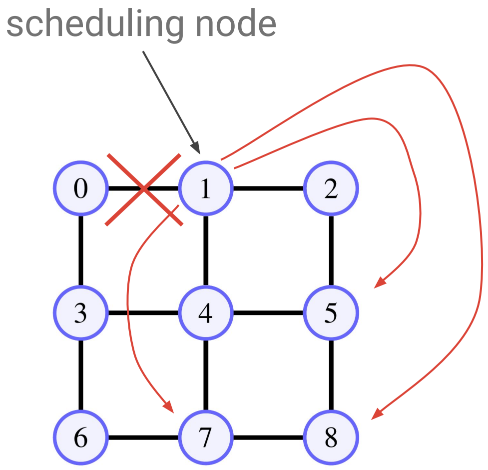


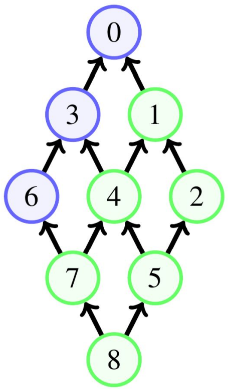
$rspDAG_{1 \to 0}(0)$

Ordered list: [R8, R5, R7, R2, R4]

Ordering FIB updates with oFIQUIC
oFIQUIC design

1. Compute rspDAG
2. Derive an ordered list of (distant) nodes
3. Initiate QUIC tunnels





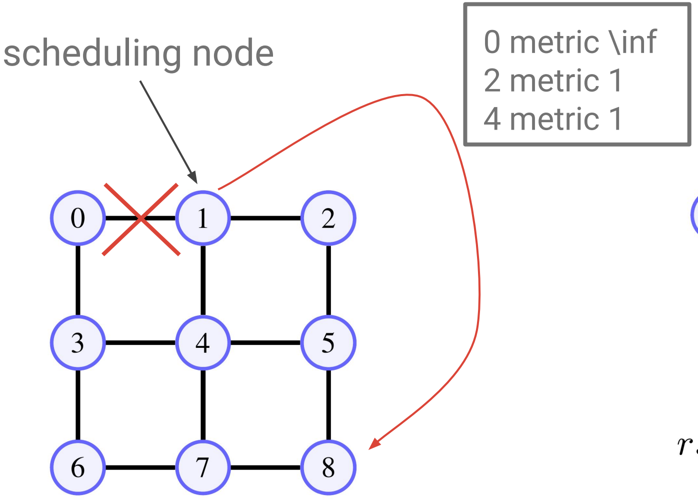
$rspDAG_{1\to 0}(0)$

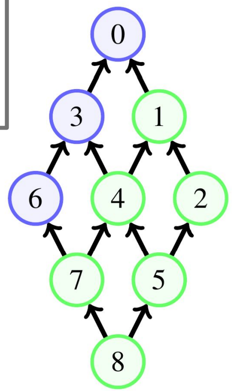
Ordered list: [R8, R5, R7, R2, R4]

Ordering FIB updates with oFIQUIC

oFIQUIC design

1. Compute rspDAG
2. Derive an ordered list of (distant) nodes
3. Initiate QUIC tunnels
4. Send LSA to first node



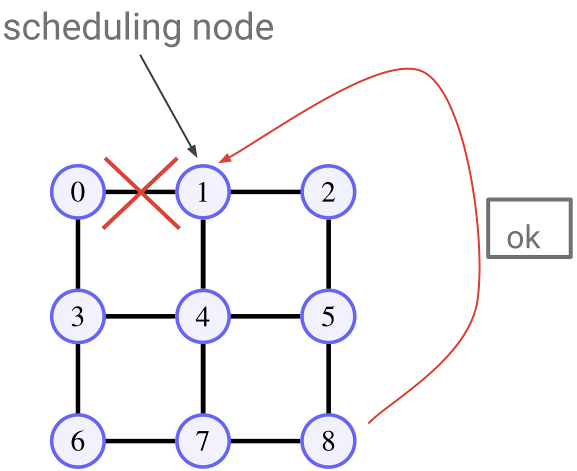


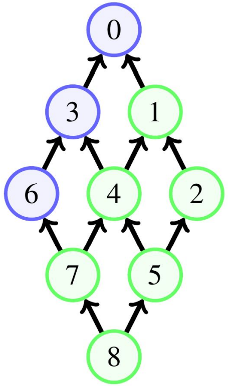
$rspDAG_{1\to 0}(0)$

Ordered list: [R8, R5, R7, R2, R4]

Ordering FIB updates with oFIQUIC
oFIQUIC design

1. Compute rspDAG
2. Derive an ordered list of (distant) nodes
3. Initiate QUIC tunnels
4. Send LSA to first node
5. Wait for distant node confirmation





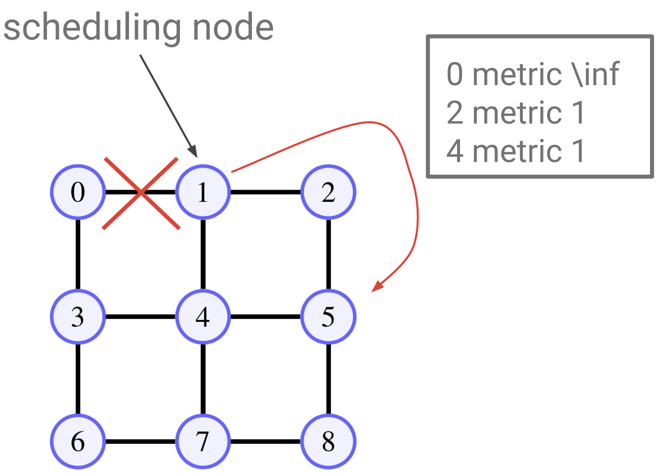
$rspDAG_{1\to 0}(0)$

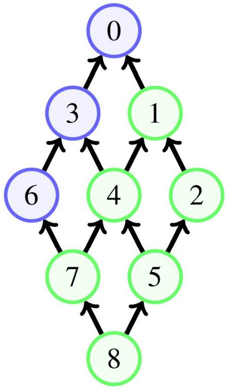
Ordered list: [R8, R5, R7, R2, R4]

Ordering FIB updates with oFIQUIC

oFIQUIC design

1. Compute rspDAG
2. Derive an ordered list of (distant) nodes
3. Initiate QUIC tunnels
4. Send LSA to first node
5. Wait for distant node confirmation
6. Repeat steps 4 - 5 until list is empty

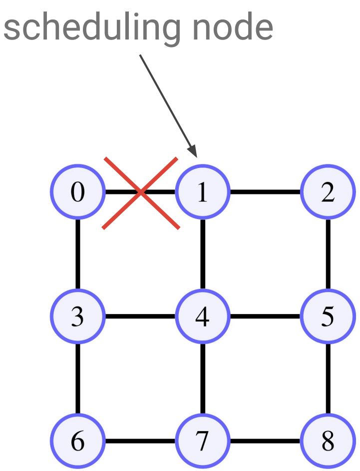


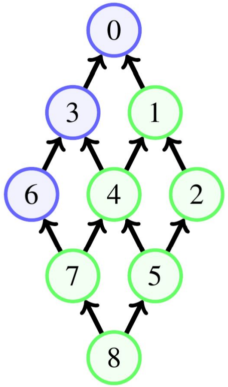


$rspDAG_{1\to 0}(0)$

Ordered list: [R8, R5, R7, R2, R4]

1. Compute rspDAG
2. Derive an ordered list of (distant) nodes
3. Initiate QUIC tunnels
4. Send LSA to first node
5. Wait for distant node confirmation
6. Repeat steps 4 - 5 until list is empty
7. Final flooding

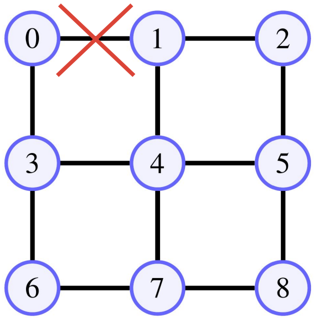


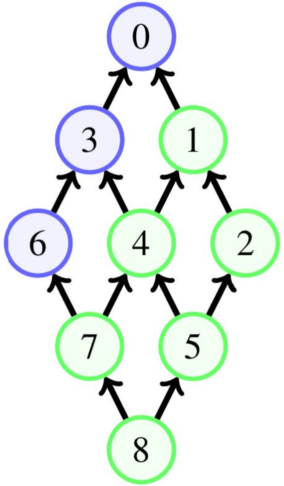


$rspDAG_{1\to 0}(0)$

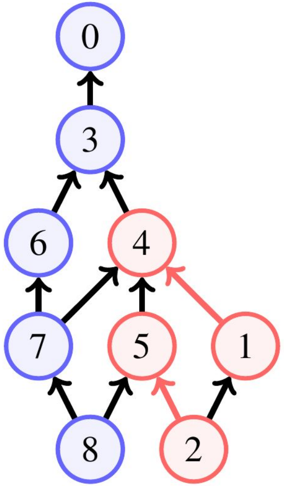
Ordered list: [R8, R5, R7, R2, R4]

Partial order vs Total order
oFIQUIC design





pre-convergence



post-convergence

Total order list: [R8, R5, R7, R2, R4]

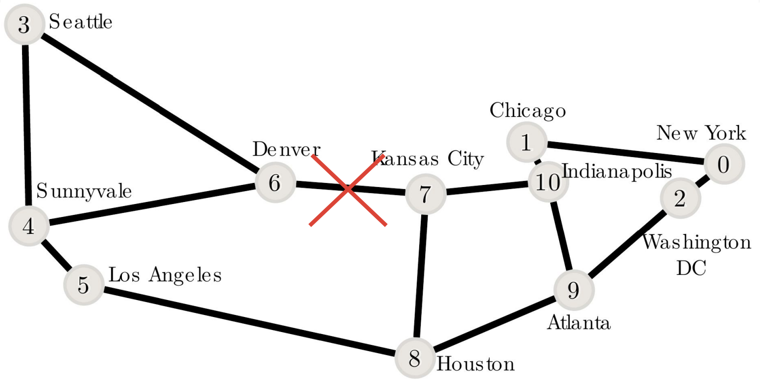
Partial order list: [R5, R2, R4]

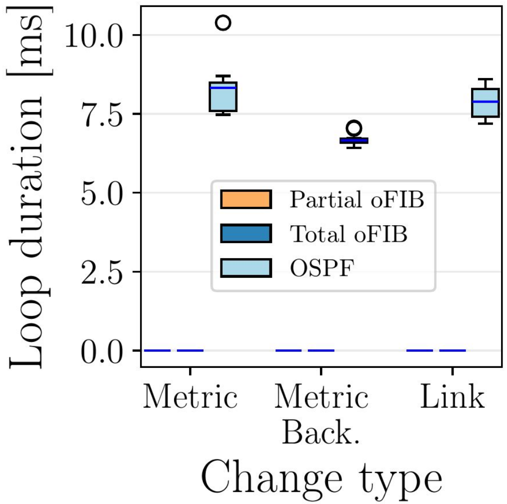
Agenda

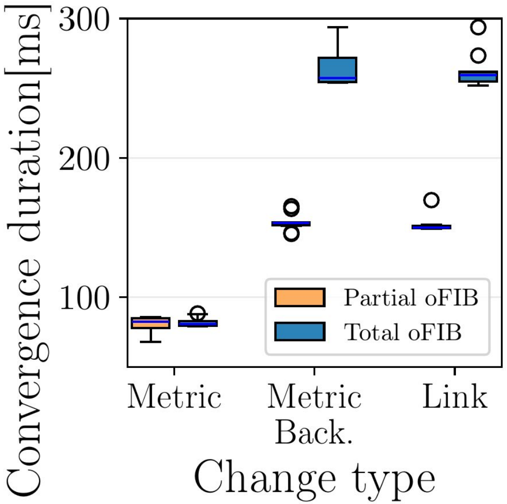
- Topology changes in IGPs
- oFIB: Ordered FIB updates
- oFIQUIC design
- **oFIQUIC performances**

oFIQUIC convergence duration in Abilene

oFIQUIC performances

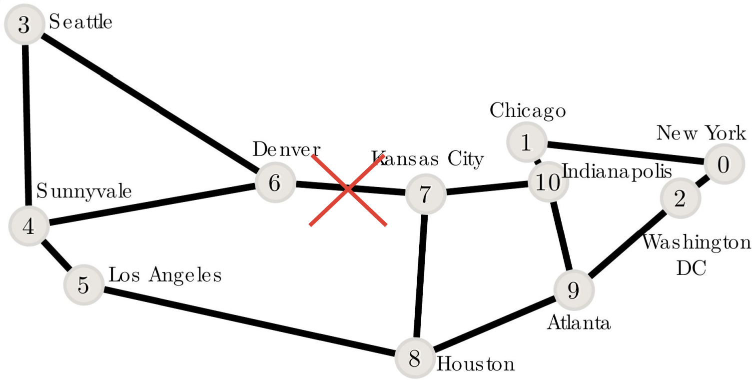


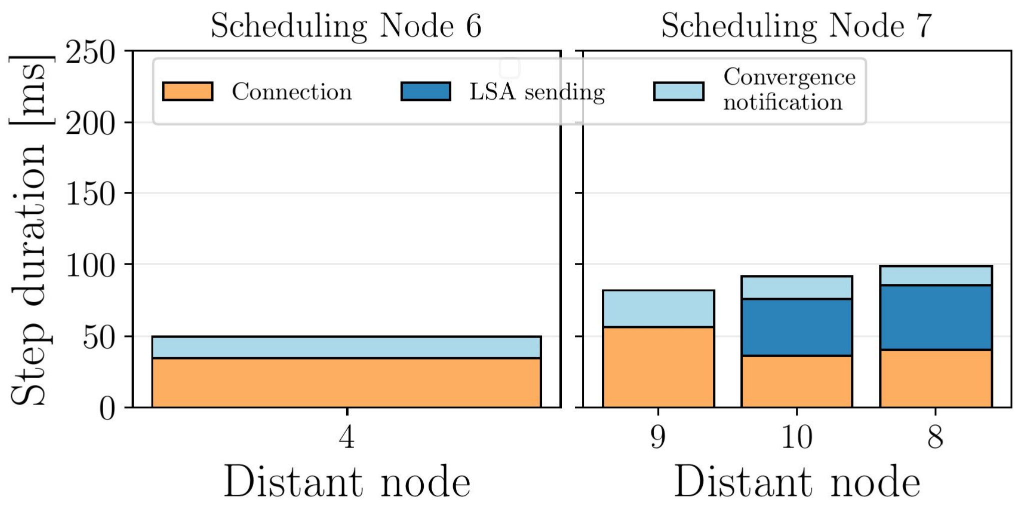


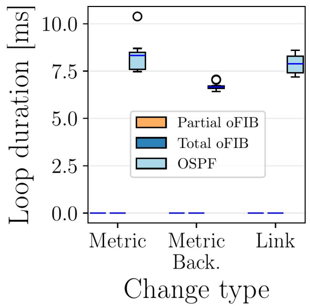


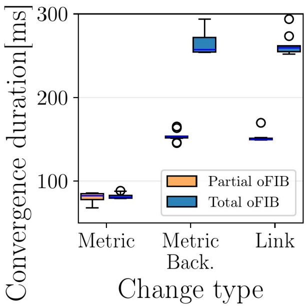
oFIQUIC contributions in Abilene

oFIQUIC performances



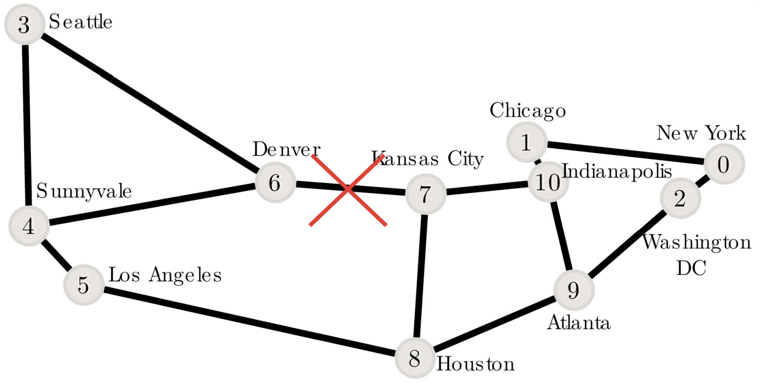


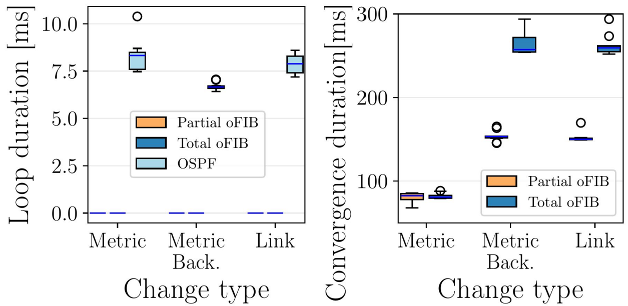


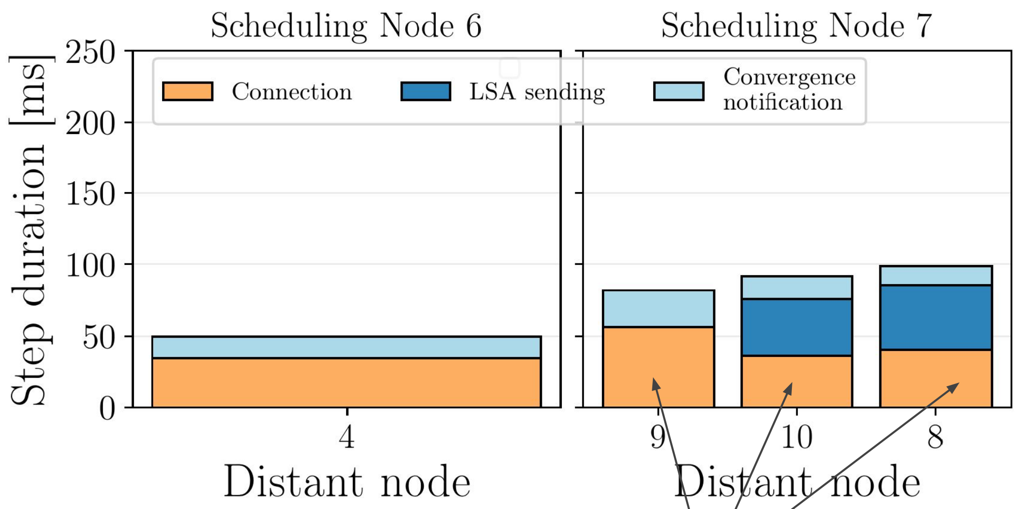


oFIQUIC contributions in Abilene

oFIQUIC performances



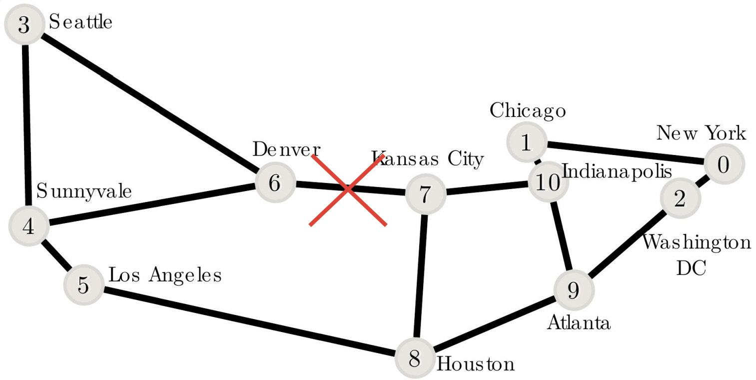


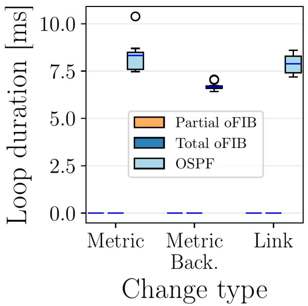


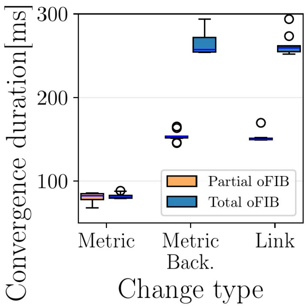
Related to path length

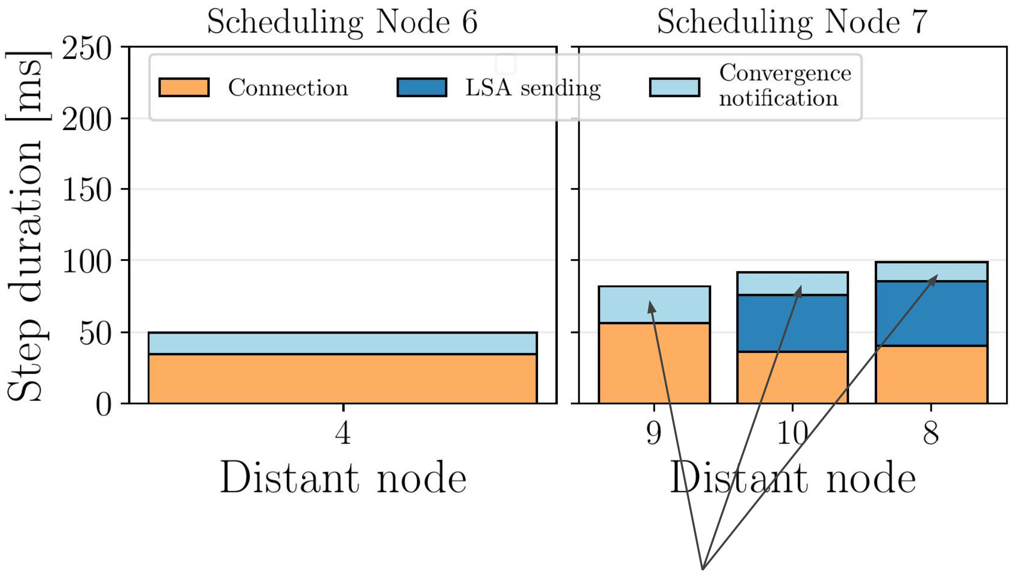
oFIQUIC contributions in Abilene

oFIQUIC performances





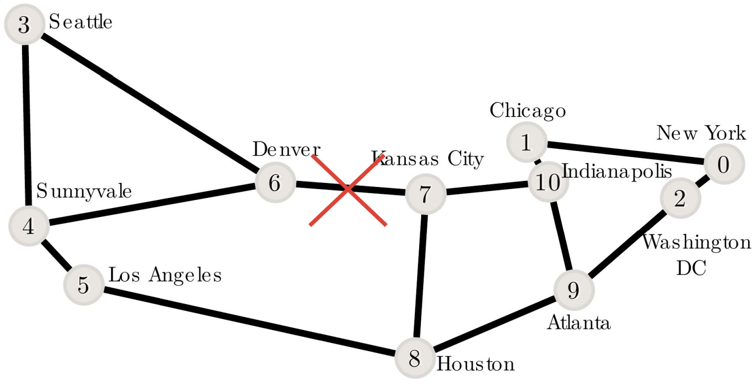


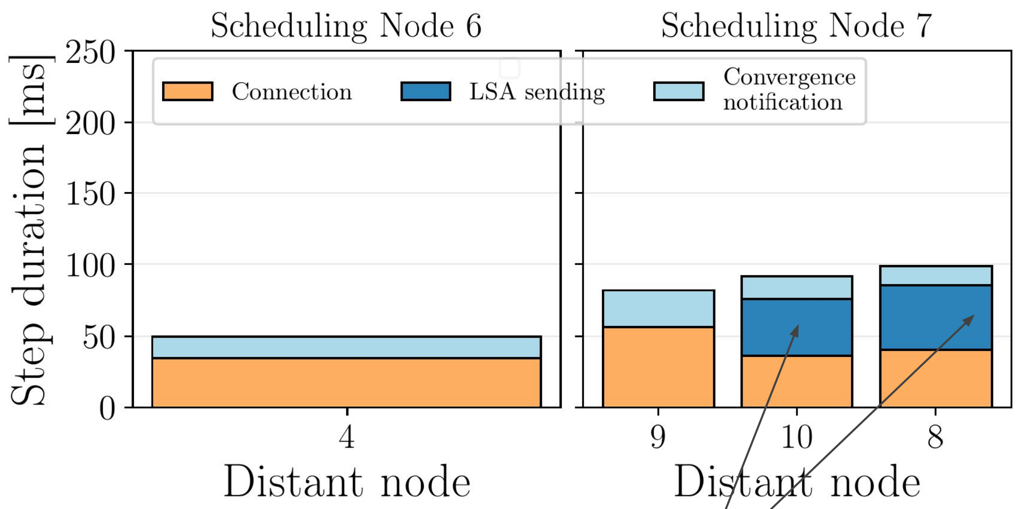


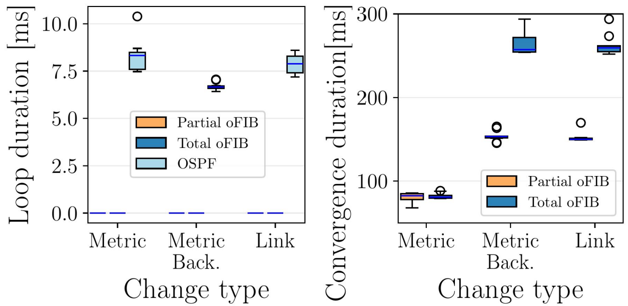
Related to path length

oFIQUIC contributions in Abilene

oFIQUIC performances



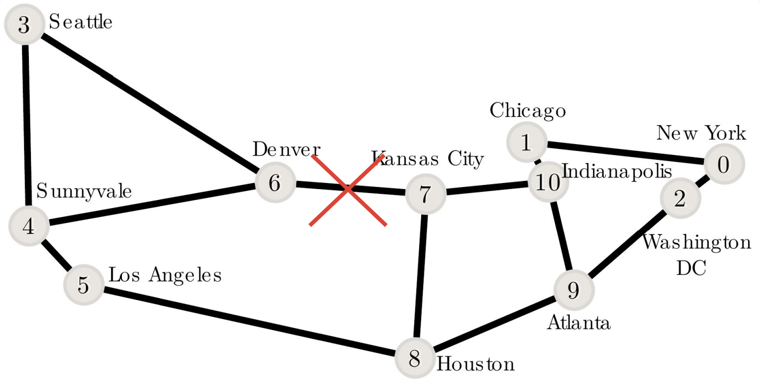


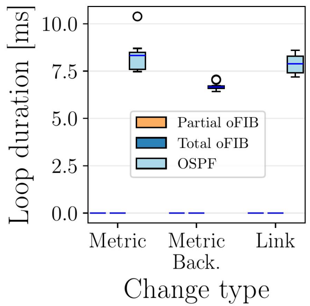


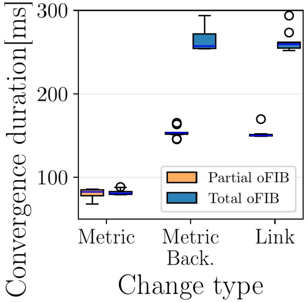
Related to previous step

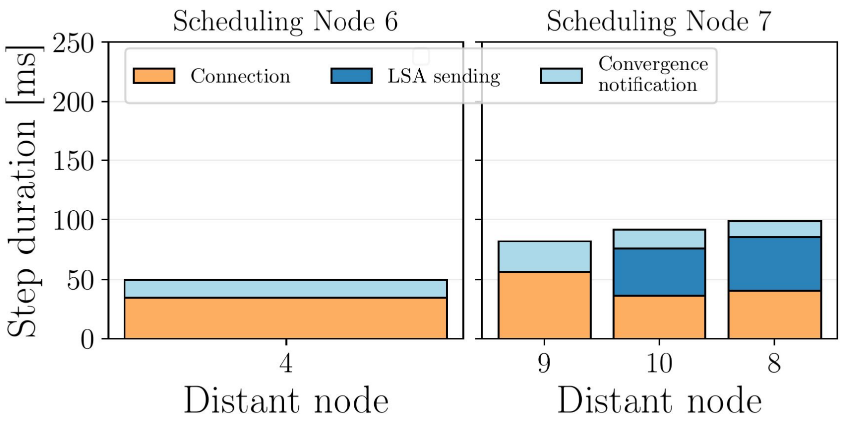
oFIQUIC contributions in Abilene

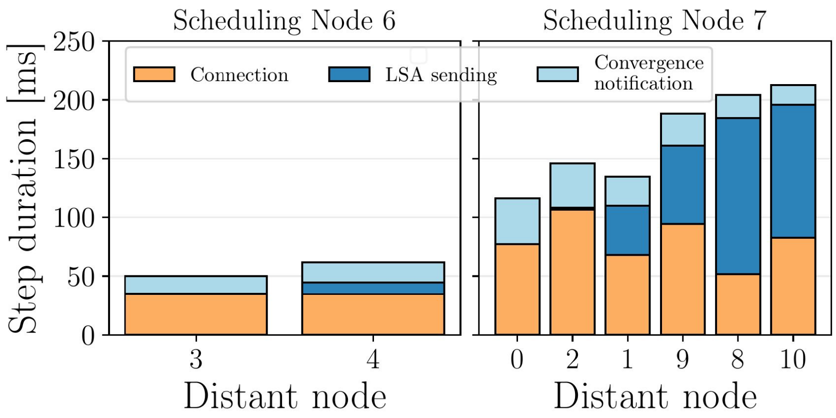
oFIQUIC performances





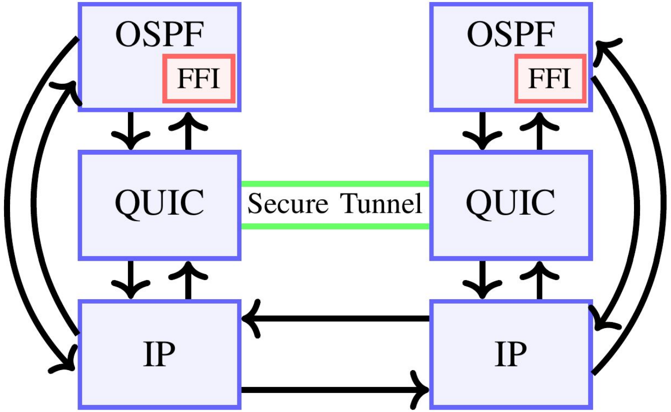






Summary

paper



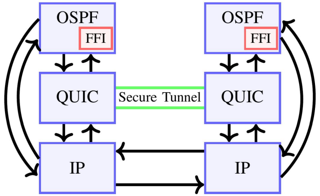
artefacts

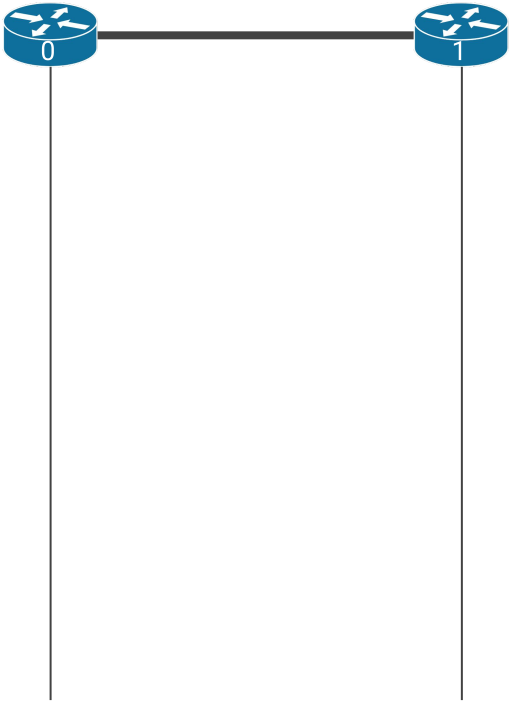
nicolas.rybowski@uclouvain.be

Backup slides

Establishing QUIC tunnels with direct IGP peers

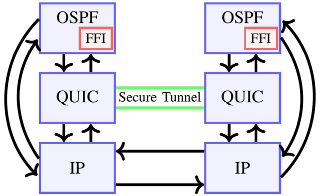
oFIQUIC design



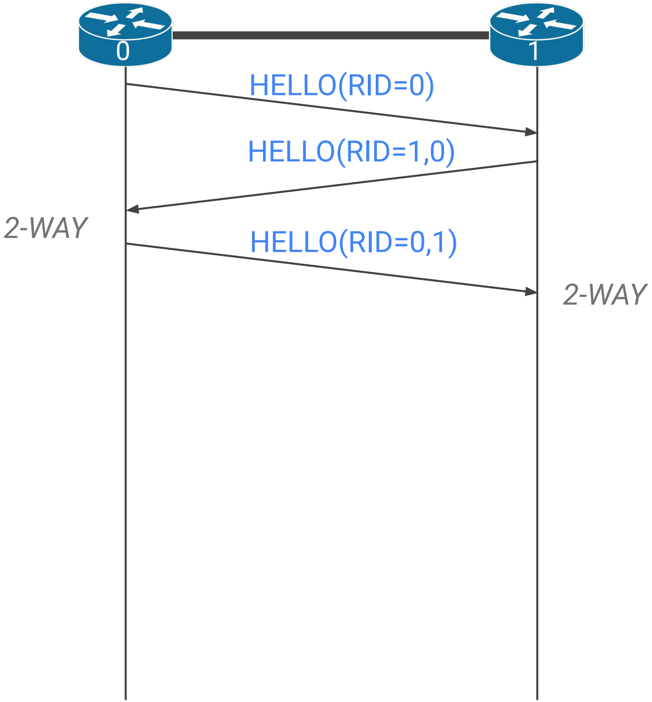


Establishing QUIC tunnels with direct IGP peers

oFIQUIC design

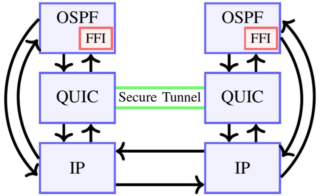


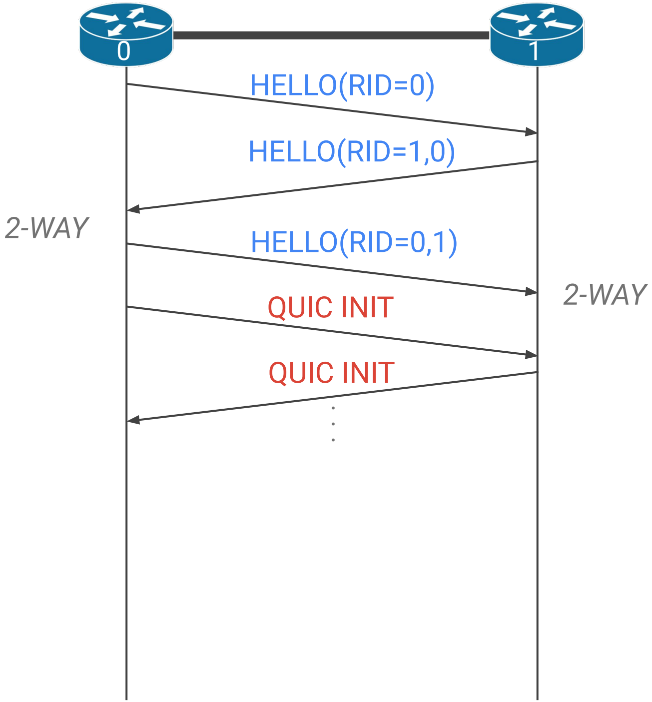
OSPF packet



Establishing QUIC tunnels with direct IGP peers

oFIQUIC design

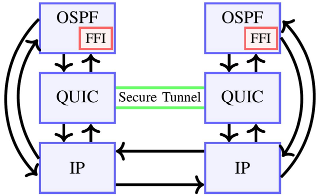




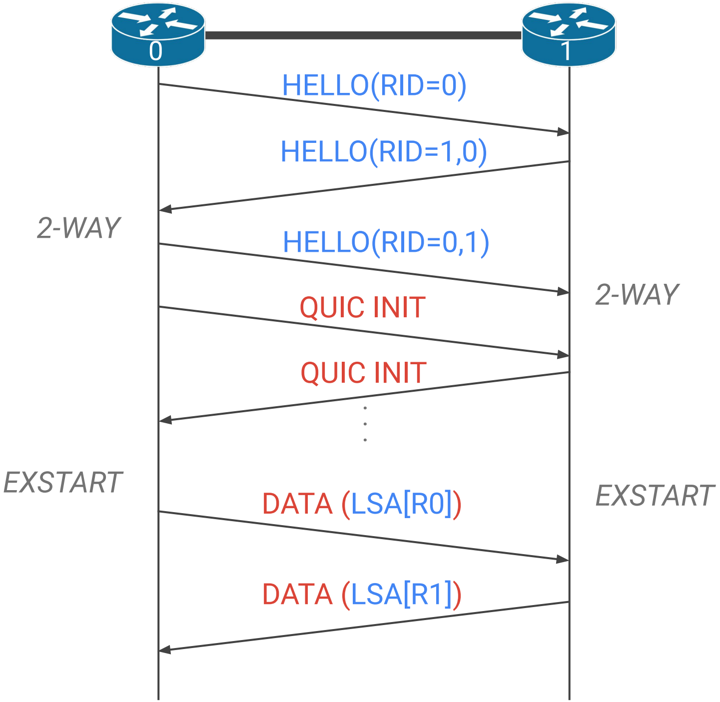
OSPF packet
QUIC packet

Establishing QUIC tunnels with direct IGP peers

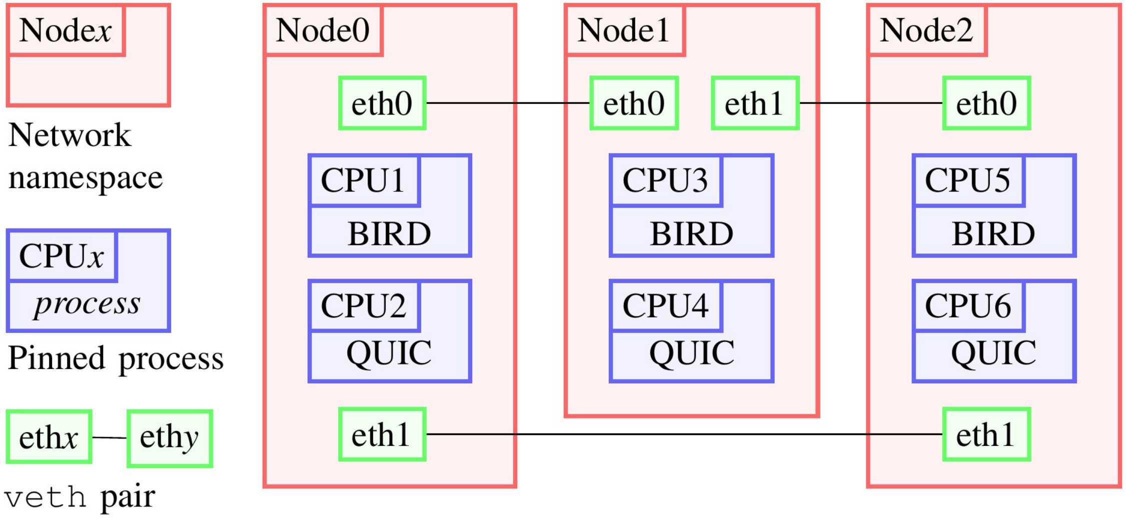
oFIQUIC design



OSPF packet
QUIC packet

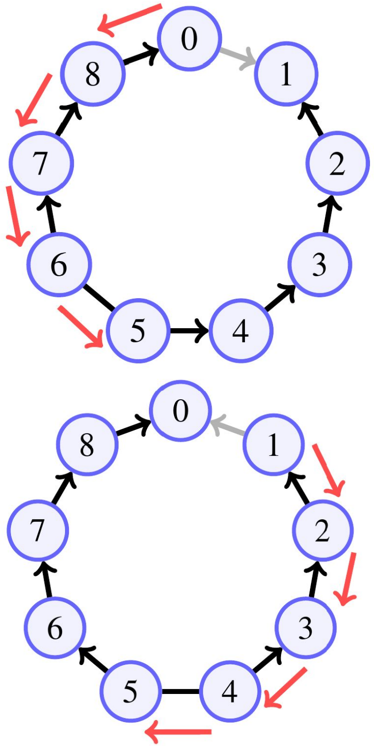


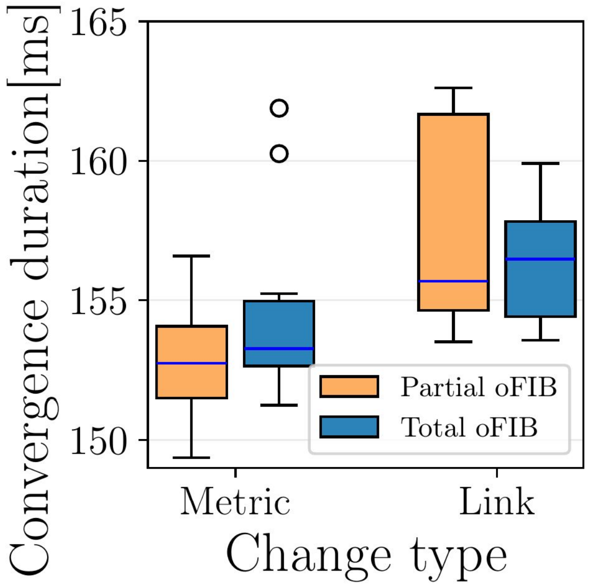
Distributed Micro Network Emulation (DUNE) framework



oFIQUIC convergence duration in a ring

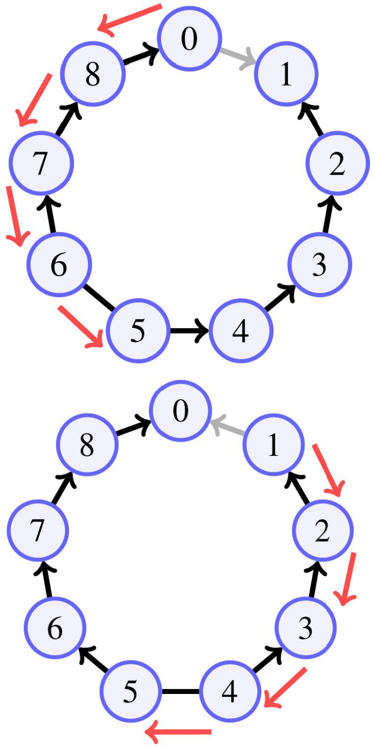
oFIQUIC performances

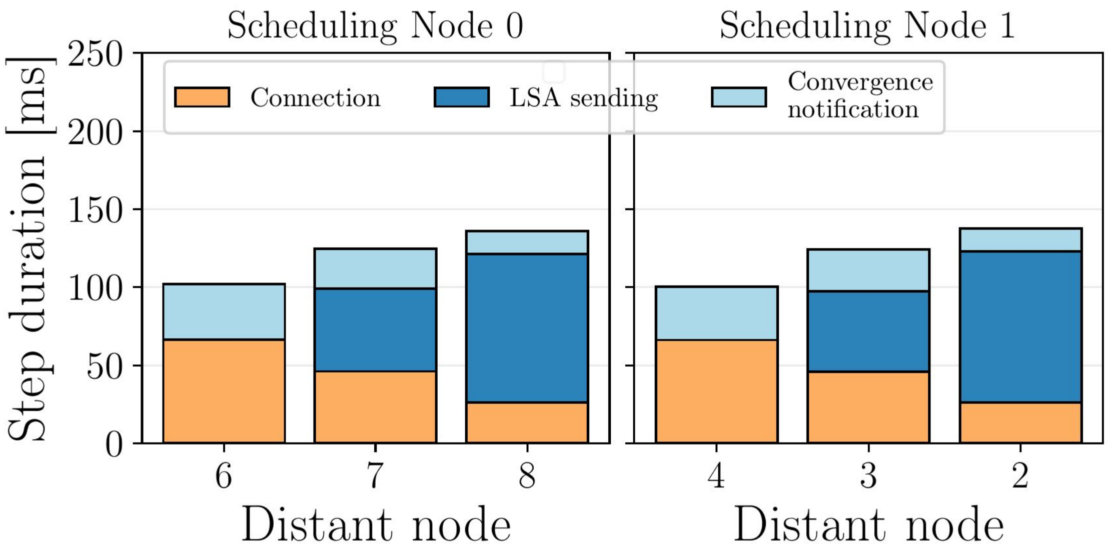




oFIQUIC contributions in a ring

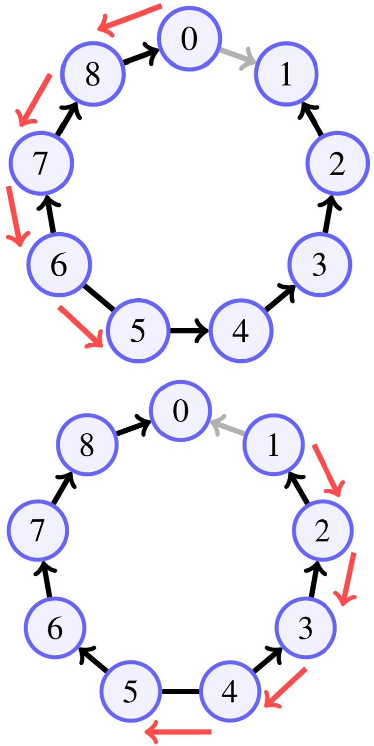
oFIQUIC performances

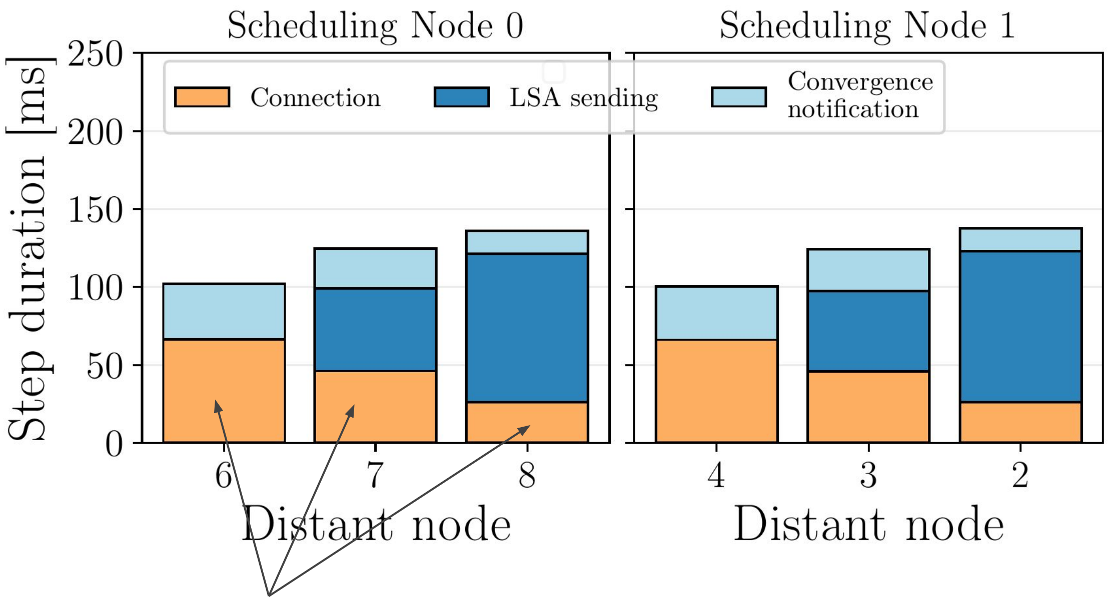




oFIQUIC contributions in a ring

oFIQUIC performances

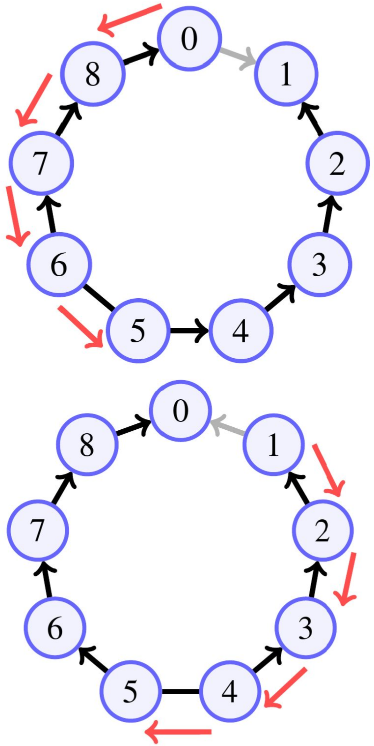


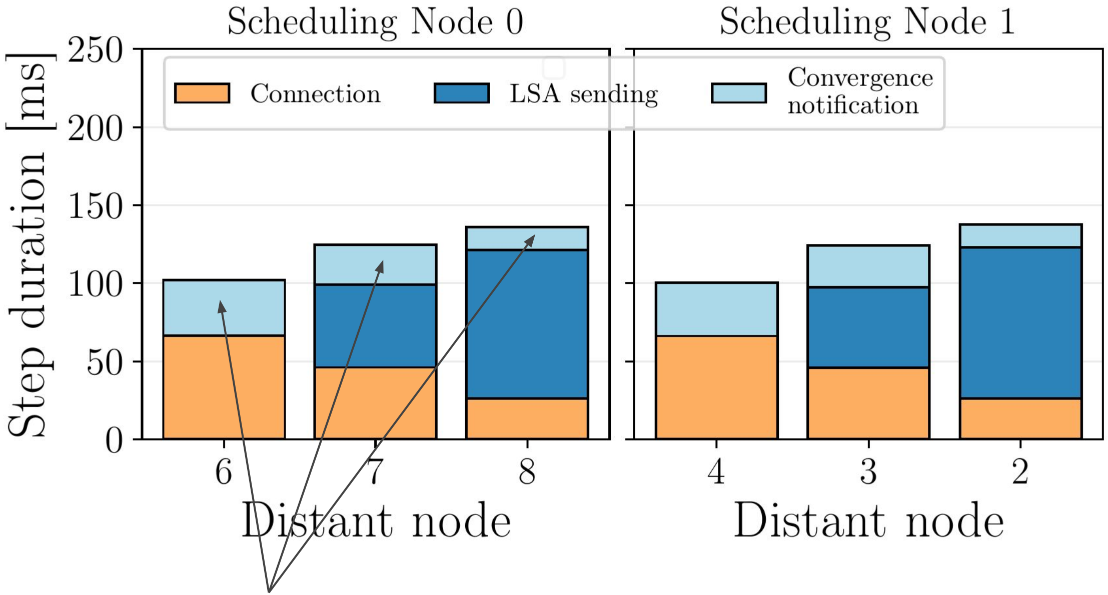


Related to path length

oFIQUIC contributions in a ring

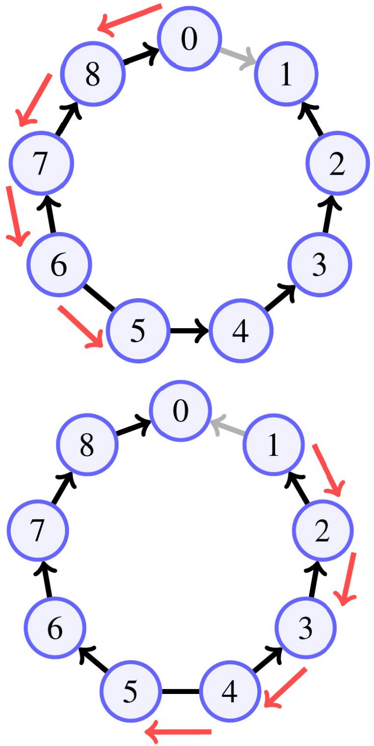
oFIQUIC performances

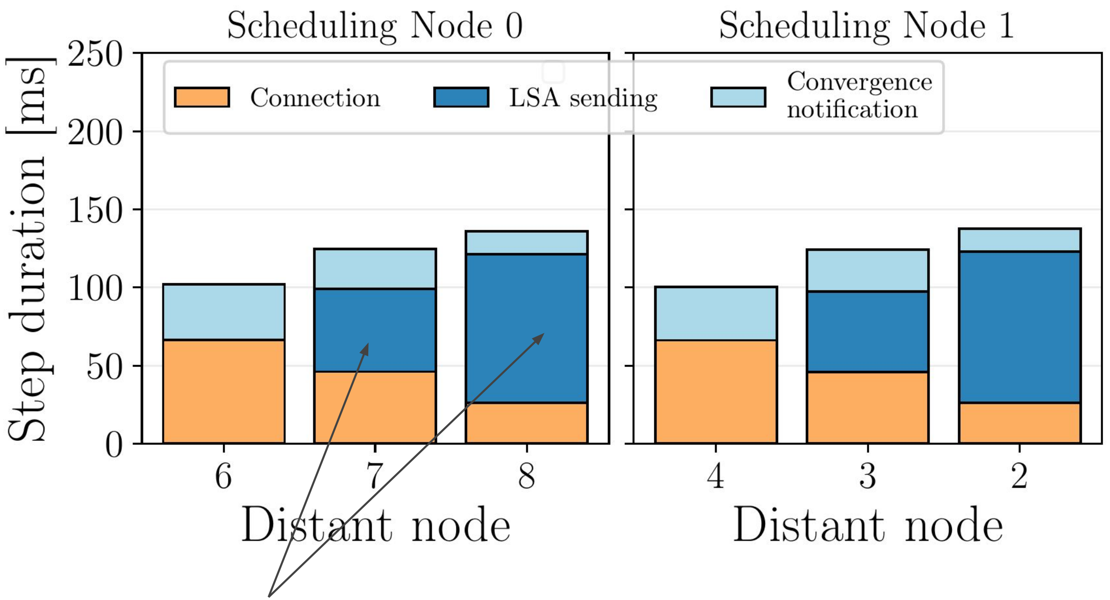




Related to path length

oFIQUIC contributions in a ring
oFIQUIC performances





Related to previous step