Kumori
Steering Cloud Traffic at IXPs to Improve Resiliency

Antoine Fressancourt, Cristel Pelsser, Maurice Gagnaire

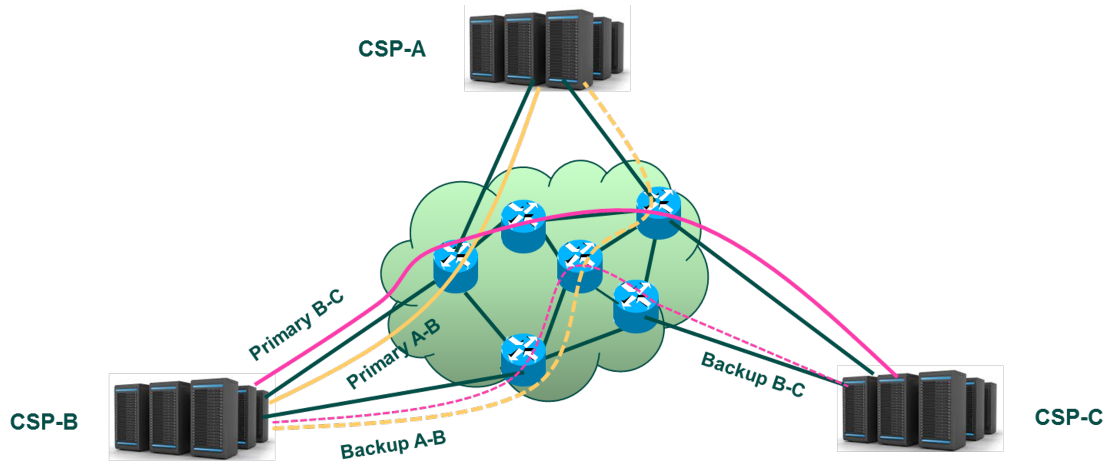
17-03-2016

Context:

Software-Defined Networking in a WAN environment

Inter-datacenter connectivity

Current state of affairs and limitations

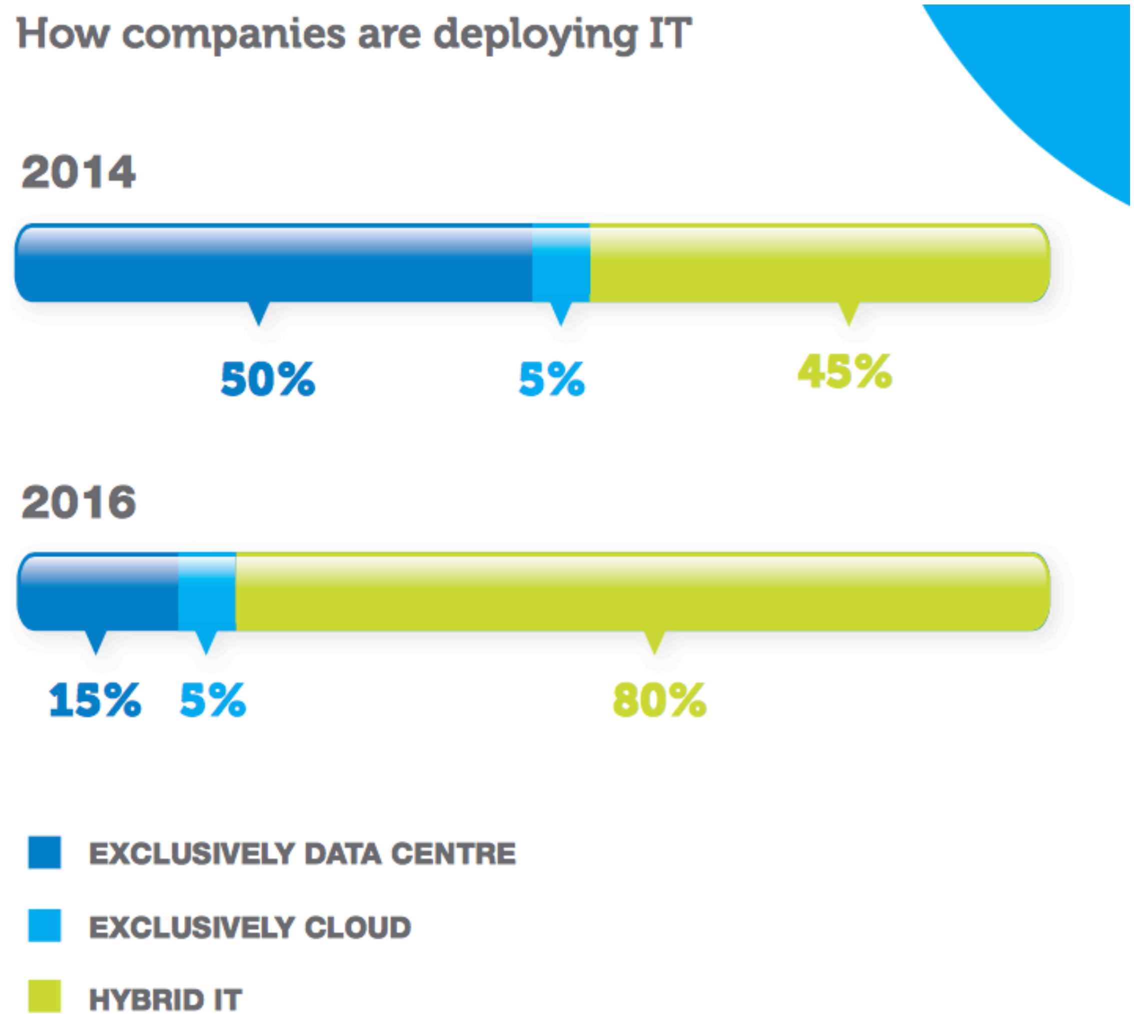


- To ensure cloud applications **resiliency**, CSP deploy applications in several datacenters ➔ Need to **connect datacenters together**
- Today: Dual private links forming a full mesh between datacenters
    - ➔ **Costly** solution
    - ➔ **Dependency** on the network connectivity provider
    - ➔ **Long** setup time

In the near future

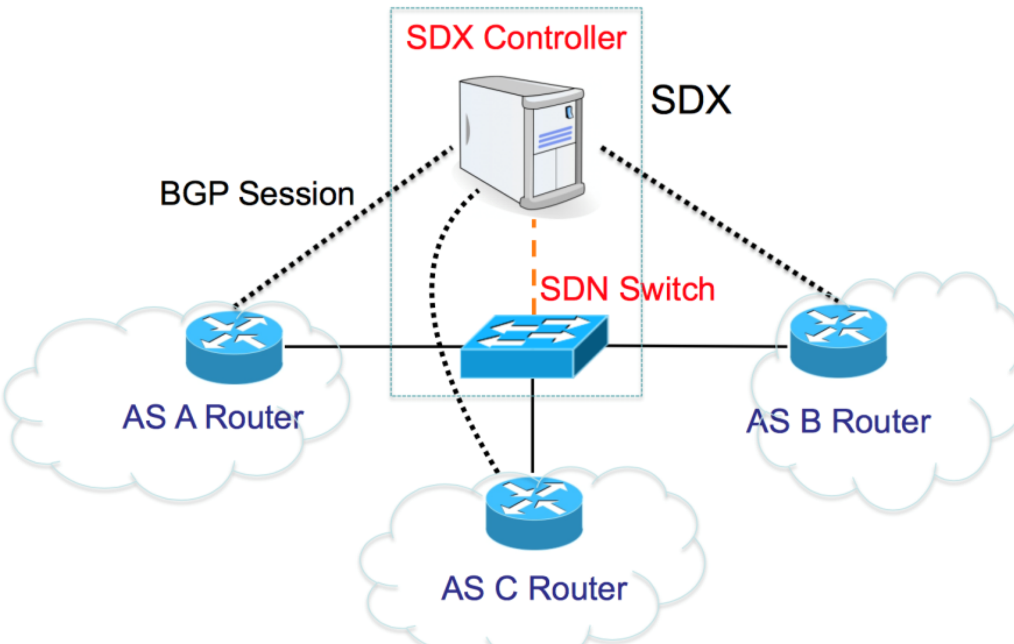
Connectivity needs related to hybrid cloud model (IDG study)

- Hybrid cloud ➔ Deployment of applications in private datacenters AND in public clouds.
- Deployment model gaining popularity (see IDG Study "The rise of Hybrid IT", 2014)



Software-Defined Internet Exchange (SDX)

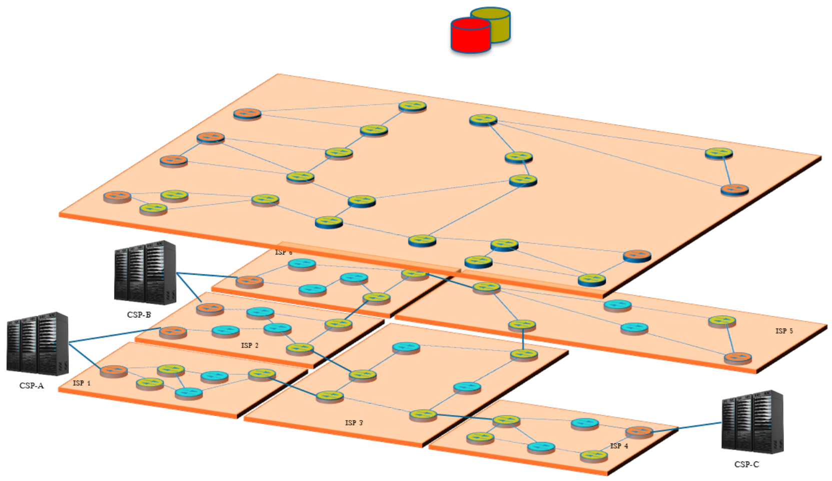
- Research work started at Princeton
- Goal: Replace BGP at Internet exchanges to enable new use cases:
  - Inbound Traffic Engineering
  - Application-specific peering
  - DDoS mitigation
  - …
- SDX controller implementation available
- Test deployments in some regional IXPs: Toulouse, Atlanta, "A large European IXP"…



Presentation of Kumori

Overview of the Kumori architecture

- Kumori is an intermediate solution between
  - Using **private MPLS circuits**
    - *Expensive*
    - *Long setup delay for each destination*
    - *SLA-based resiliency*
  - Using the **plain Internet**
    - *Cheaper*
    - *Works out of the box*
    - *Best effort*



- Using an overlay of nodes at **Internet exchange points** and **network operator facilities**
  - Resiliency through capacity to **choose alternative routes in the Internet**
  - Taking advantage of **connectivity ecosystem at IXPs**
  - Software control of the path **by the CSP**

Kumori in details

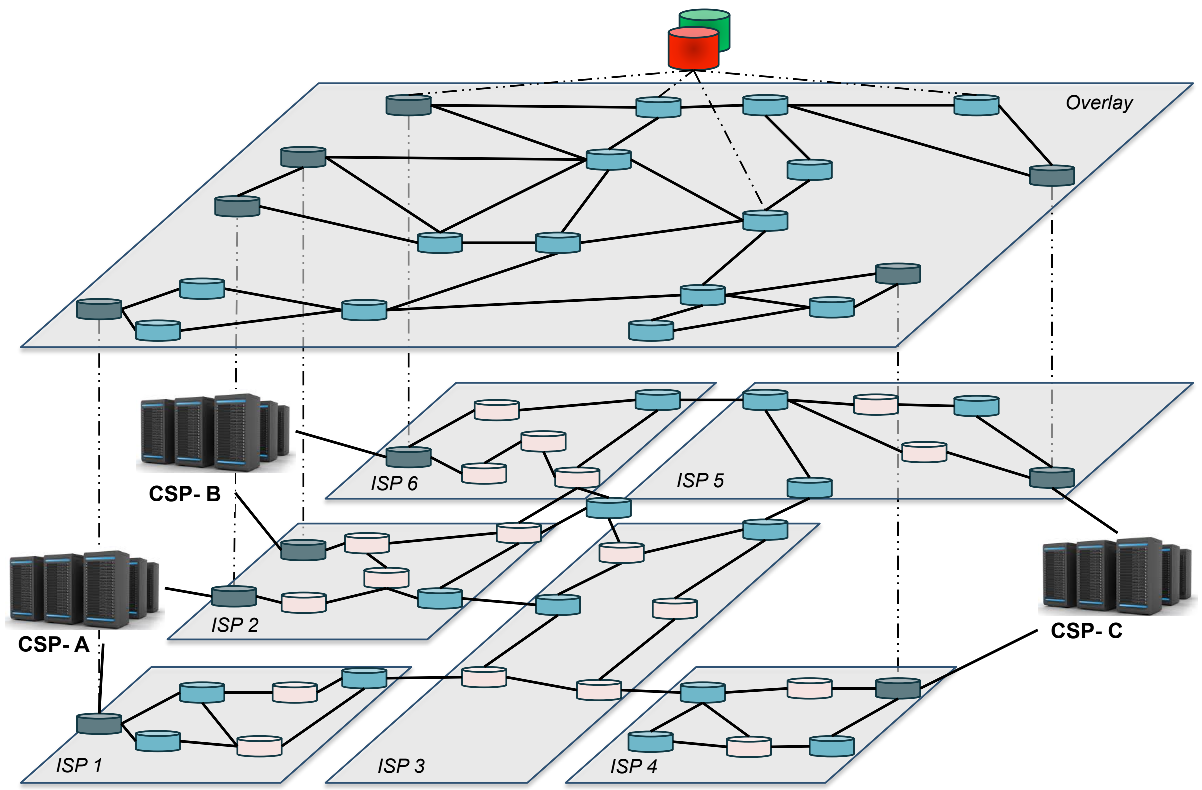
SDN-based overlay network architecture

Goal:

- Influence the way traffic is routed over the Internet between DCs
- Fast reroute over the overlay

Architecture composed of **3 elements**:

- Routing inflection points
- Egress points
- Central controller



Routing inflection points

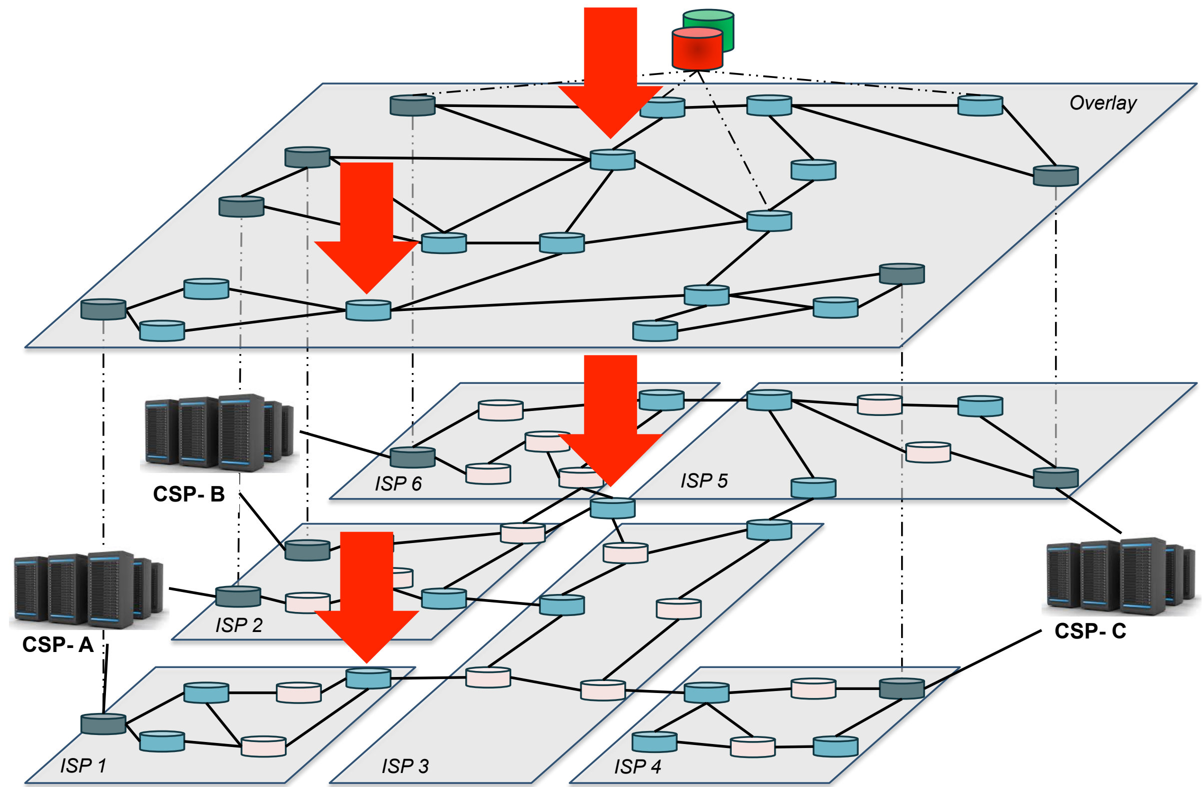
SDN-based overlay network architecture

Role:

- Apply routing policy provided by the controller
- Provide inter-DC network measurements to the controller

Location:

- Internet Exchange Points
- Inside ISP networks (cache / CDN area)



Egress points

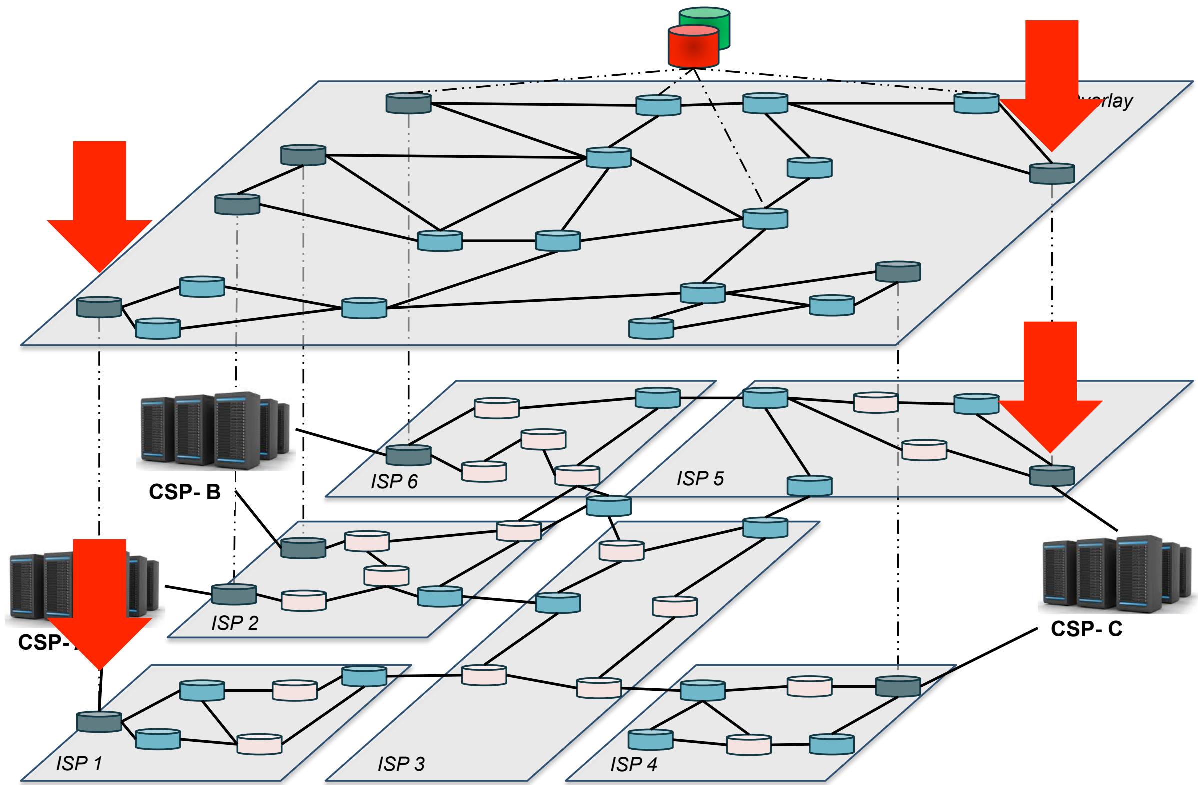
SDN-based overlay network architecture

Role:

- Contact point between intra and inter DC domains
- Gateway between Segment routing and overlay routing
- Provide intra-DC net. measurements to the controller

Location:

- Router connecting DC to a given ISP



Central controller

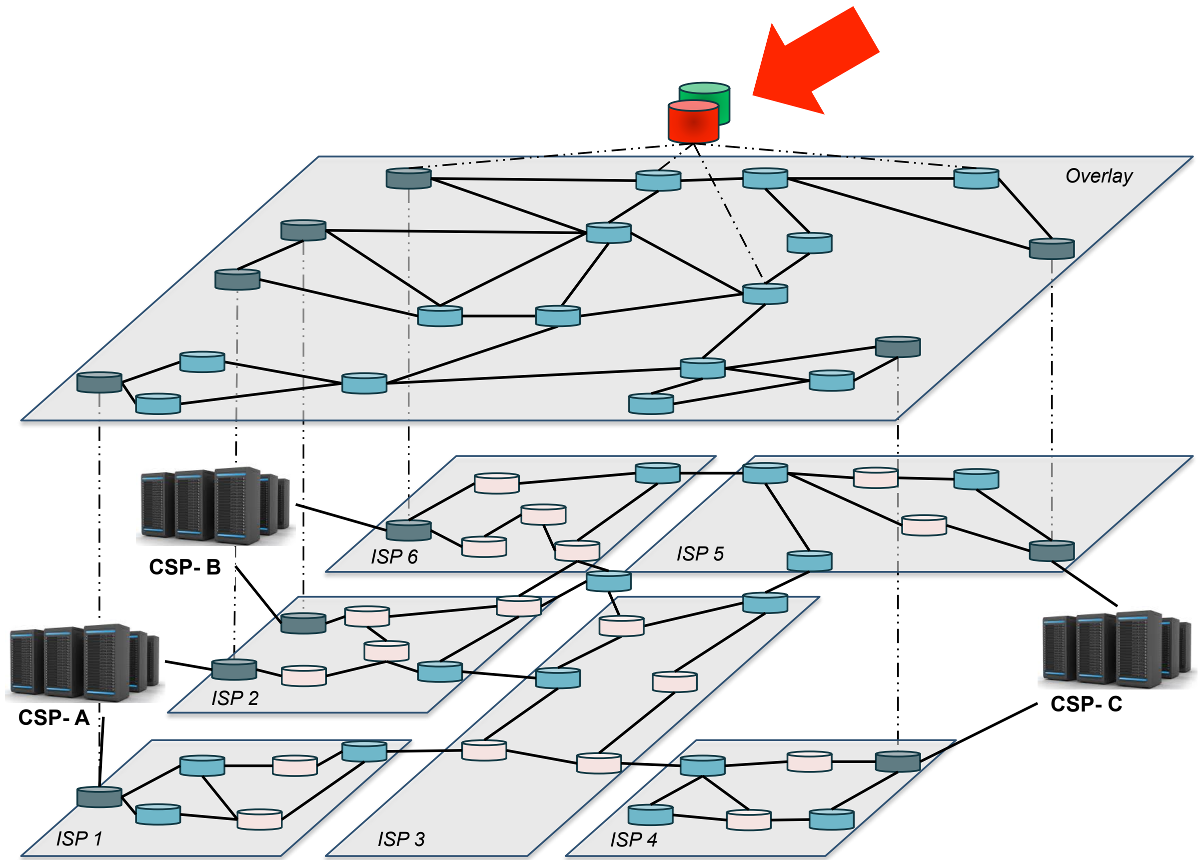
SDN-based overlay network architecture

Role:

- Provide routing rules to the various elements in the architecture
- Gather measurements to modify routing policy
- React to detected failures

Location:

- At the CSP premises



IIJ INNOVATION INSTITUTE TELECOM ParisTech worldline an atos company

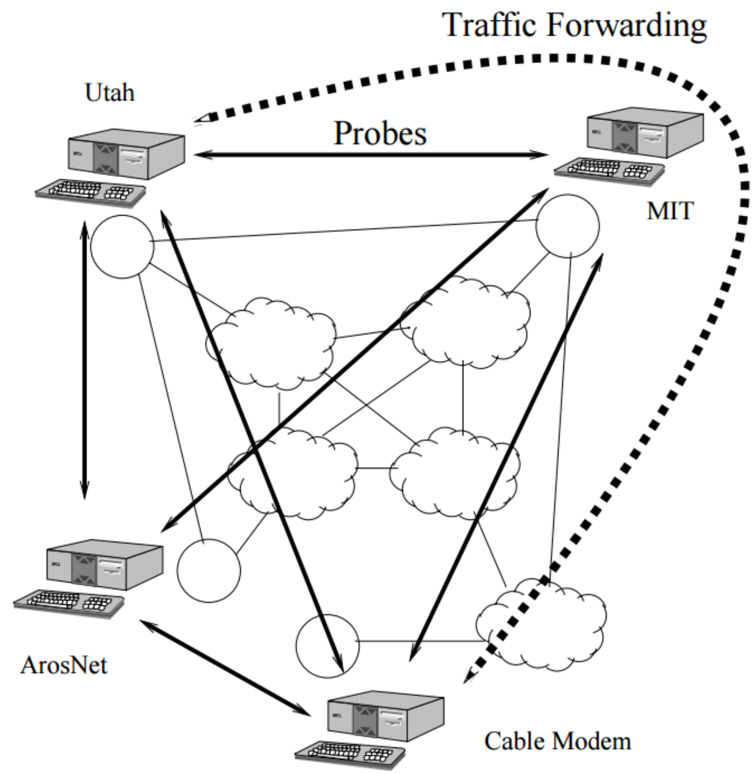
Evaluating Kumori

(And having a better view on path diversity accross the Internet)

What can we compare to?

Resilient Overlay Network (2001)

- First major project using an overlay approach to enhance link resiliency
- Goal:
  - Detecting link or node failures and routing traffic around them
  - Converge around failures quicker than BGP: 20 s Vs. ~5 min.
- Principle
  - Active probing between all node pairs
  - Link-state routing in the overlay
  - Robust
  - Does not scale beyond ~50 nodes



Evaluation methodology

iPlane dataset

- Use of **iPlane** dataset
  - Summary of **traceroutes** performed on the 15th of February 2015
  - Undirected weighted graph with:
    - **190 028** vertices
    - **916 390** edges
- Identification of nodes belonging to 12 major **CSPs**
  - Amazon, Microsoft, Google, Atos, Dimension Data, WIDE...
  - **1 604** vertices
- Identification of nodes belonging to **IXPs**
  - Combining PeeringDB and Packet Clearing House
  - **2 177** vertices
- Measurements
  - Shortest paths among CSP pairs ➔ RON applicability
  - Shortest paths between CSP and IXP nodes ➔ Our architecture's applicability

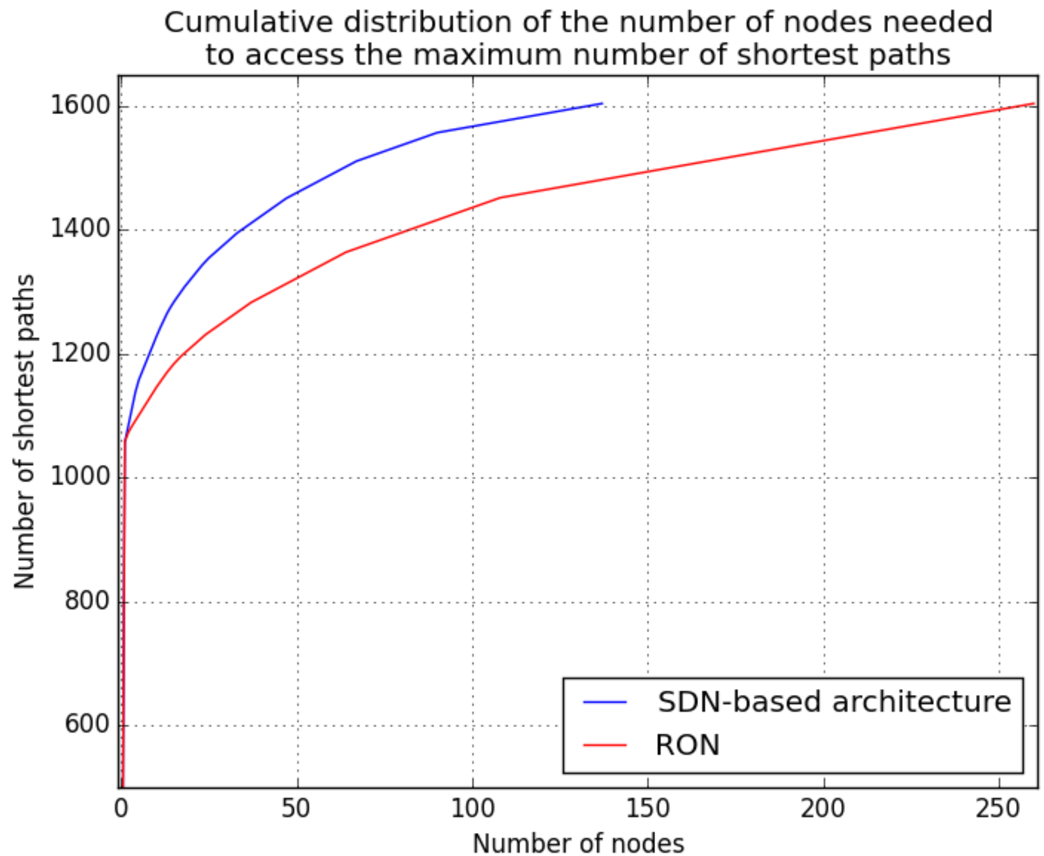
Cloud services providers together

Performance / Path length

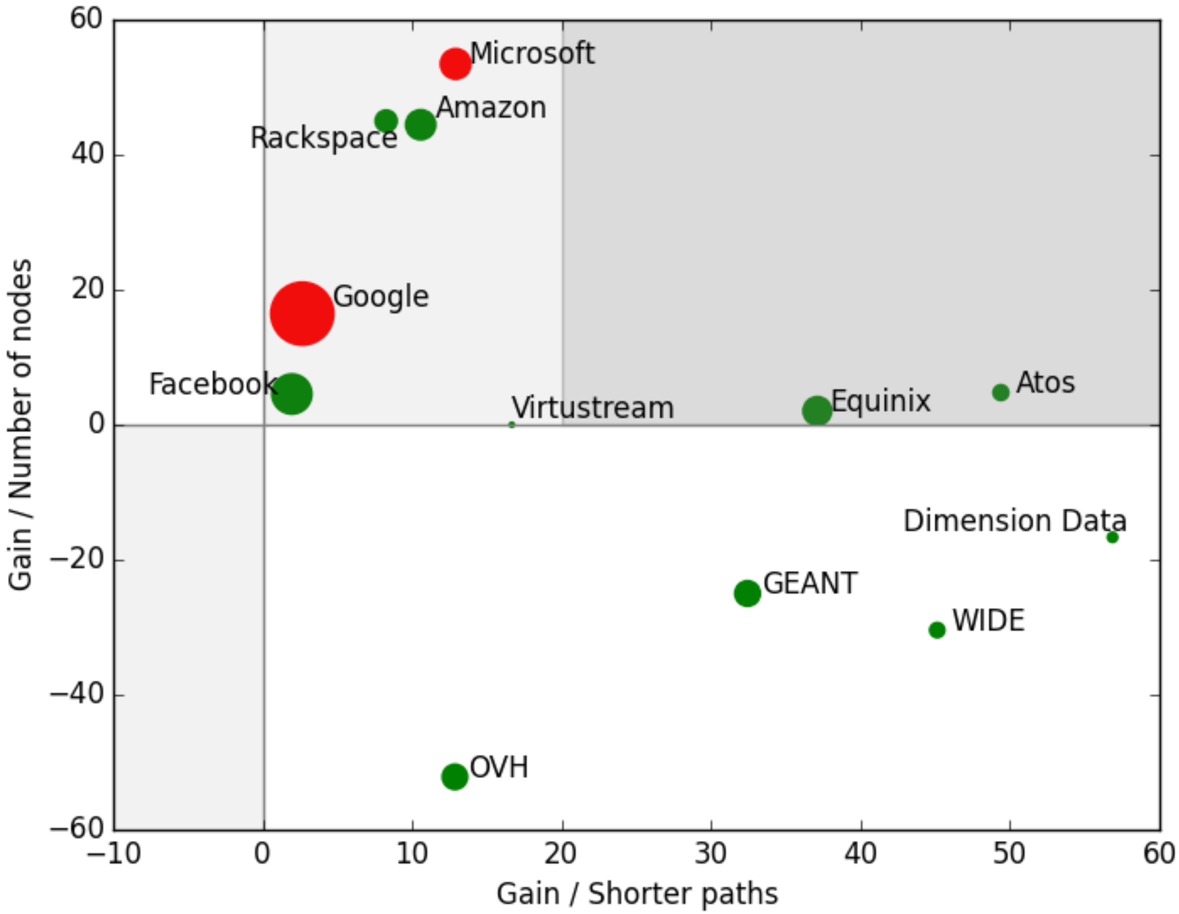
- Better or equal in **97.69 %** of the cases
- Strictly better in **5.72 %** of the cases

Performance / number of nodes needed

- 80 % of shortest paths reachable with **9.43 %** less nodes
- 99 % of shortest paths reachable with **50.41 %** less nodes



Mixed benefits depending on the CSP



2 major groups of CSPs

For major CSPs:

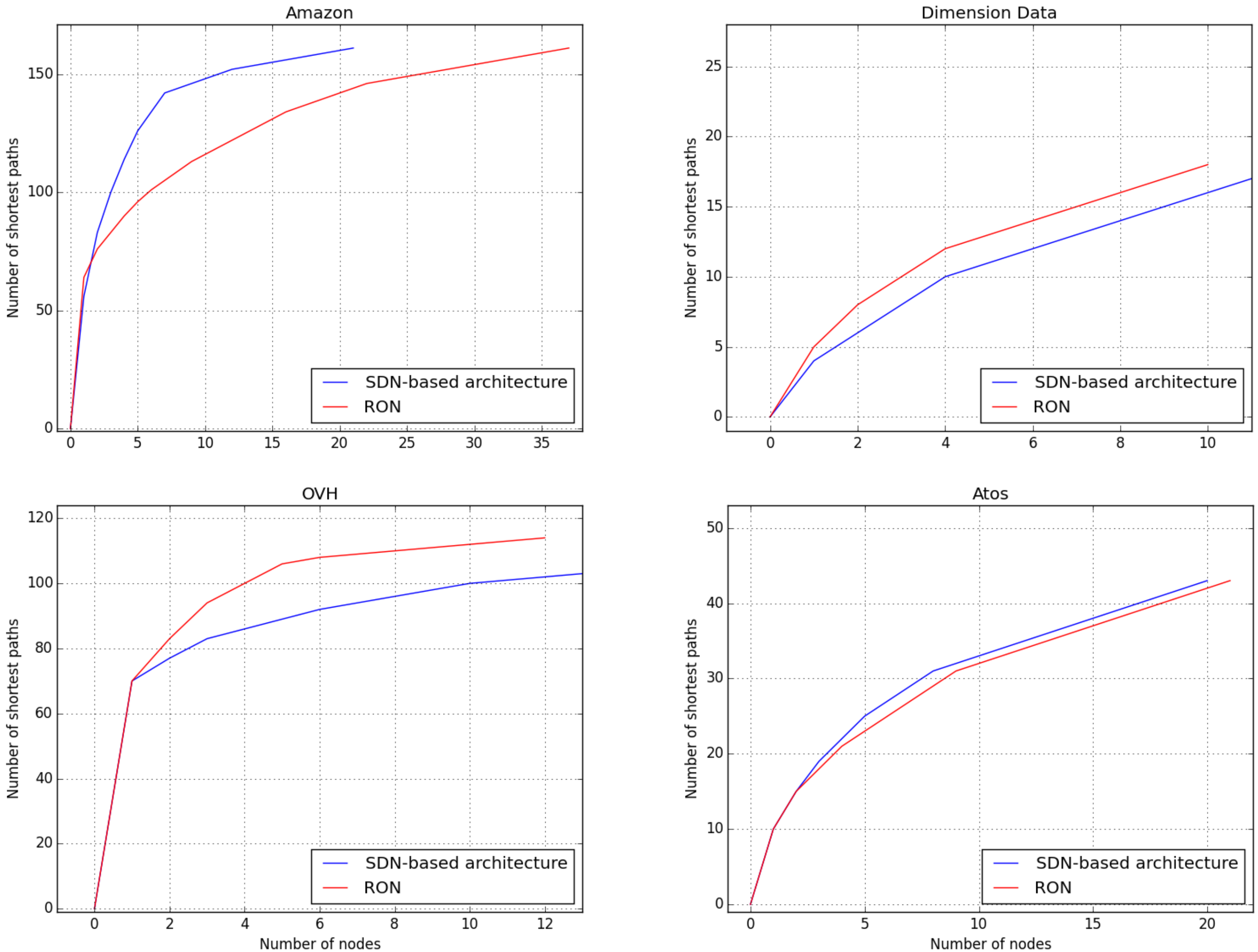
- Benefits on the **number of nodes** required to access paths of similar length

For smaller size CSPs

- Benefits on the **length of the path** accessible via the architecture

Mixed benefits depending on the CSP

Cumulative distribution of # nodes needed to access shortest paths



Conclusion and next steps

To conclude...

Summary

- Design of an architecture to enhance inter-datacenter connectivity resiliency
- Compared with RON:
  - Shortest paths for small CSPs
  - Paths accessible via less nodes for large CSPs

Future works

- Pushing the graph study
  - Investigation of multihoming benefits
  - Path diversity
- Evaluation on real testbed
  - Feasability of real time measurement and reaction
  - Hysteresis effect
  - Impact on ISPs

IIJ INNOVATION INSTITUTE

TELECOM ParisTech

worldline an atos company

Thanks

For more information please contact me:

antoine.fressancourt@worldline.com
antoine.fressancourt@telecom-paristech.fr

17-03-2016

Results summary

	General	Amazon	Atos	Dimension Data	Facebook	Google	Microsoft	OVH	Rackspace	Virtustream	WIDE	GEANT	Equinix
CSP Node pairs	1285606	12880	903	153	41328	255970	13861	6441	3570	6	780	6670	10585
Better or equal	1255950	10748	883	148	41238	255476	10982	6370	3494	6	753	5830	8851
Percentage	97,69	83,45	97,79	96,73	99,78	99,81	79,23	98,9	97,87	100	96,54	87,41	83,62
Strictly better	73511	1361	446	87	807	6783	1790	828	295	1	352	2164	3928
Percentage	5,72	10,57	49,39	56,86	1,95	2,65	12,91	12,86	8,26	16,67	45,13	32,44	37,11
CSP Nodes	1604	161	43	18	288	716	167	114	85	4	40	116	146
IXP - 80%	15	6	12	9	1	1	11	6	5	3	16	23	23
IXP - 90 %	46	9	16	11	6	21	16	13	7	3	20	34	35
IXP - 95%	74	13	18	12	13	39	20	19	8	3	22	40	42
IXP - 99%	121	20	20	12	21	66	27	23	11	3	23	44	48
CSP - 80%	38	15	13	7	1	1	26	3	6	3	9	17	23
CSP - 90%	104	22	17	9	4	20	43	5	12	3	13	23	36
CSP - 95%	180	29	19	10	10	51	51	7	16	3	15	29	43
CSP - 99%	244	36	21	10	22	79	58	11	20	3	16	33	49
Benefits 80 %	9,43	25	4,76	-16,67	0	0	25,86	-13,04	5	0	-30,43	-13,64	0
Benefits 99%	50,41	44,44	4,76	-16,67	4,55	16,46	53,45	-52,17	45	0	-30,43	-25	2,04
CSP-IXP matches	6	2	0	0	0	2	2	0	0	0	12	2	14