

Alcatel, Lucent, Alcatel-Lucent and the Alcatel-Lucent logo are trademarks of Alcatel-Lucent. All other trademarks are the property of their respective owners.

The information presented is subject to change without notice.
Alcatel-Lucent assumes no responsibility for inaccuracies contained herein.

This slide must be kept when distributed externally.

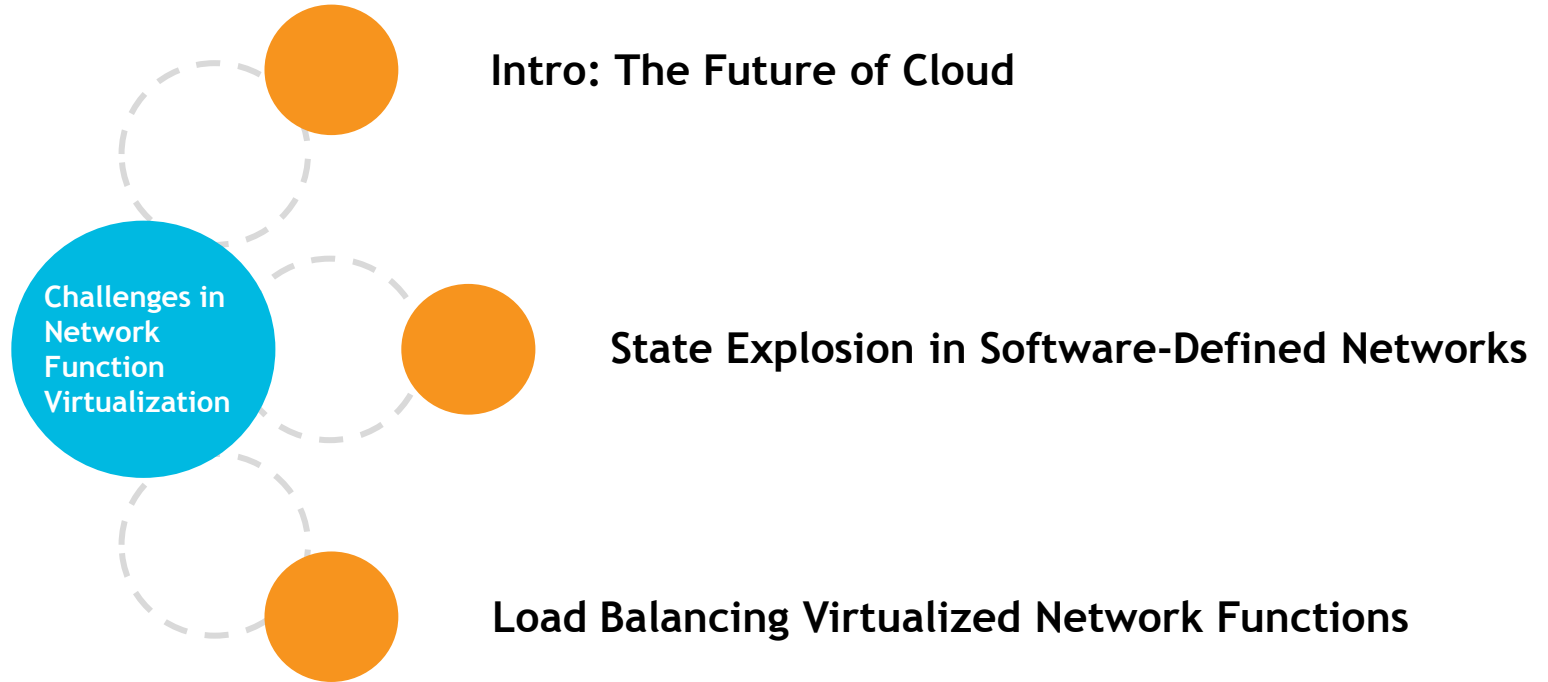


Challenges in Network Function Virtualization

Ivica Rimac (rimac@bell-labs.com)

Dec. 7, 2015

Agenda



The Future of Cloud*

*Original material: Volker Hilt

The (Brief) History of Cloud Computing



1999

Customer Login

Contact Us

Company Profile

Join Our Team

News & Events

User Name

Password

Go!

Join Our Pilot Program!

For a limited time, selected customers can take advantage of the benefits of salesforce.com -- at no cost.

I'm Interested!

Act quickly. Our Pilot Program ends February 29, 2000.

Stay Informed!

Enter your email below and we'll add you to our mailing list.

Enter



Just Sign On!

Exploit the Power of the Internet to Harness Your Sales Information!

Introducing salesforce.com, the new Internet site that allows you to easily access, manage, and/or share all your organization's sales information - immediately, efficiently and reliably - right from your computer. salesforce.com gives you instant sales force automation services online at a lower cost than ever before.

- There's no software, hardware, or networks to buy, and nothing to install or maintain - we do it all.
- Your sales information is completely safe and secure, and available to you 24/7 from anywhere using the Internet.
- Access, manage, and/or share accounts, contacts, opportunities, forecasts, and reports.
- Simply log on to the Internet; sign on to salesforce.com, and go. It's really that easy.

To gain the competitive edge, start harnessing your sales

WHAT CUSTOMERS ARE SAYING...



"Within 24-hours, salesforce.com was able to get our national sales team set up and running with a complete sales force automation solution. The product is outstanding!"

- Joel O'Neill, Sales Leader



"We test hundreds of Internet applications at our labs every year, and after testing salesforce.com, have selected it to coordinate our national and international sales efforts."

- Jeff Schueler, President



"Not only were we amazed at how rapidly our sales force was up and running, we were also astounded at the enterprise-class functionality that the application provides."

- Jeff Johnson, VP of Sales

2006



Welcome to Amazon Web Services

Amazon Web Services provides developers with direct access to Amazon's robust technology platform. Build on Amazon's suite of web services to enable and enhance your applications. We innovate for you, so that you can innovate for your customers. Browse developer innovations in our [Solutions Catalog](#) to see the possibilities!

What's New?

[Give Us Your Feedback - Developer Resources](#) (August 09 2006)

Where can we improve to help you build on Amazon Web Services? Your feedback is very important to us as we release services that you use to run your businesses. Please take 5 minutes to complete the brief survey in Newsletter #17. By completing the survey, you will be entered into a drawing for one of 250 \$5 Amazon.com gift certificates. (NO PURCHASE NECESSARY. Ends August 31, 2006. See the [official rules](#) for details.)

[Announcing Alexa Site Thumbnail](#) (July 26, 2006)

The Alexa Site Thumbnail web service provides developers with programmatic access to thumbnail images for the home pages of web sites. It offers access to Alexa's large and growing collection of images, gathered from its comprehensive web crawl. This web service enables developers to enhance web sites, search results, web directories, blog entries, and other web real estate with Alexa thumbnails images. Including web site thumbnail improves user experience by allowing end users to preview sites before clicking on the thumbnail's associated link.

[Amazon Simple Storage Service \(Amazon S3\) - Continuing Successes](#) (July 11, 2006)

La Nacion, Microsoft, and SmugMug represent the breadth of companies choosing to use the web scale storage offered by Amazon S3. Global enterprises like Microsoft are using Amazon S3 to dramatically reduce their storage costs without compromising scale or reliability. On the opposite end of the spectrum, small businesses that depend on storage, such as SmugMug, are using Amazon S3's benefits of scale and cost-efficiency that were previously only available to large companies. Amazon continues to use Amazon S3 for its own business as well, recently launching

[Your Web Services Account](#)

Sign-up Today!

Reasons to Sign-up for AWS:

- Access several Amazon Web Services for FREE.
- Receive FREE newsletters about AWS.
- Join an innovative developer community
- Learn to build new solutions and applications to make money.

[Click here to Sign-Up.](#)

Customer Spotlight

MediaSilo

Learn About Amazon Web Services

[AWS Home](#)
[Why Use AWS?](#)
[What's New in AWS?](#)
[Upcoming Events](#)
[Success Stories](#)
[Solutions Catalog](#)
[Create an Account](#)
[Contact Us](#)
[FAQs](#)

Browse Web Services

[Amazon E-Commerce Service](#)
[Amazon Historical Pricing](#)
[Amazon Mechanical Turk \(Beta\)](#)
[Amazon Simple Storage Service](#)

2015

AWS: 1.5 million servers

Google: 1 million servers

Microsoft: 1 million servers

Limitations of First Era Cloud Services



Latency



Security and Trust



Network Cost

Beyond First Era Cloud Services

The Trend Towards Providing Localized Resources



11 AWS data centers
53 Edge Cloud data centers - 20 added in past 2 years

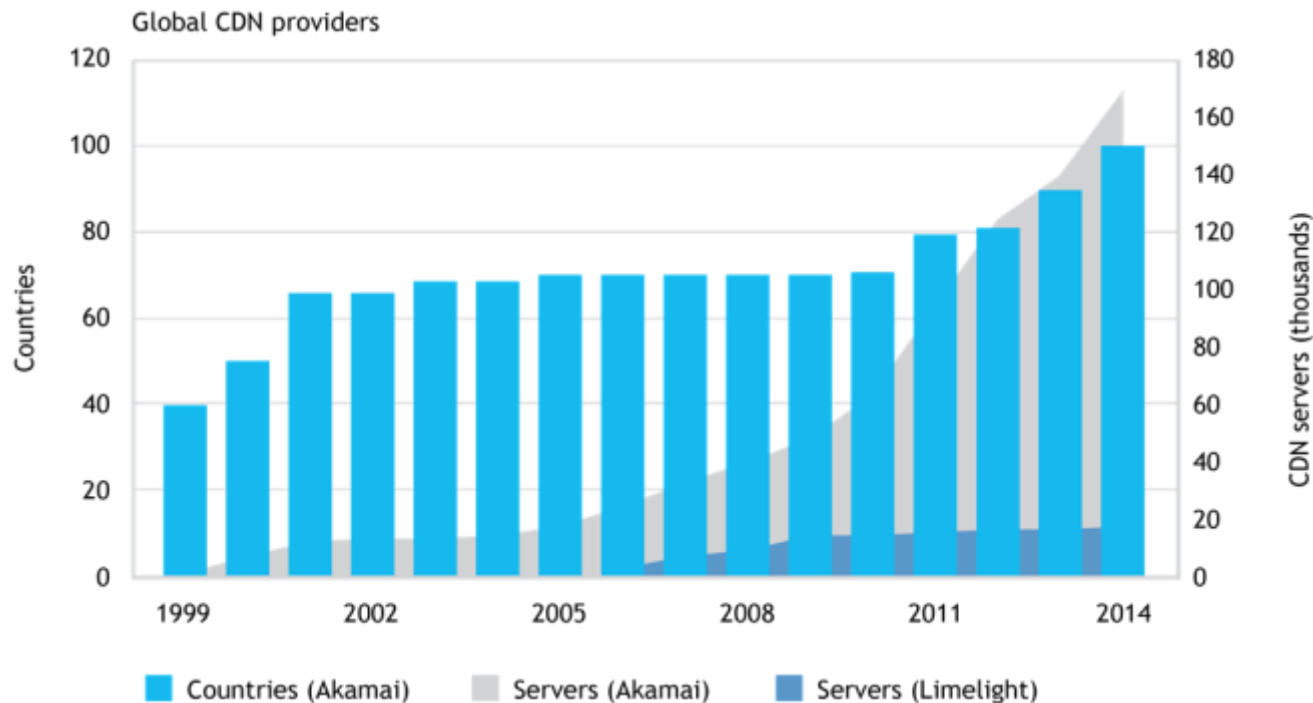


> 100 data centers

The cloud is beginning to be distributed but not yet truly local

A Look at Content Distribution Services

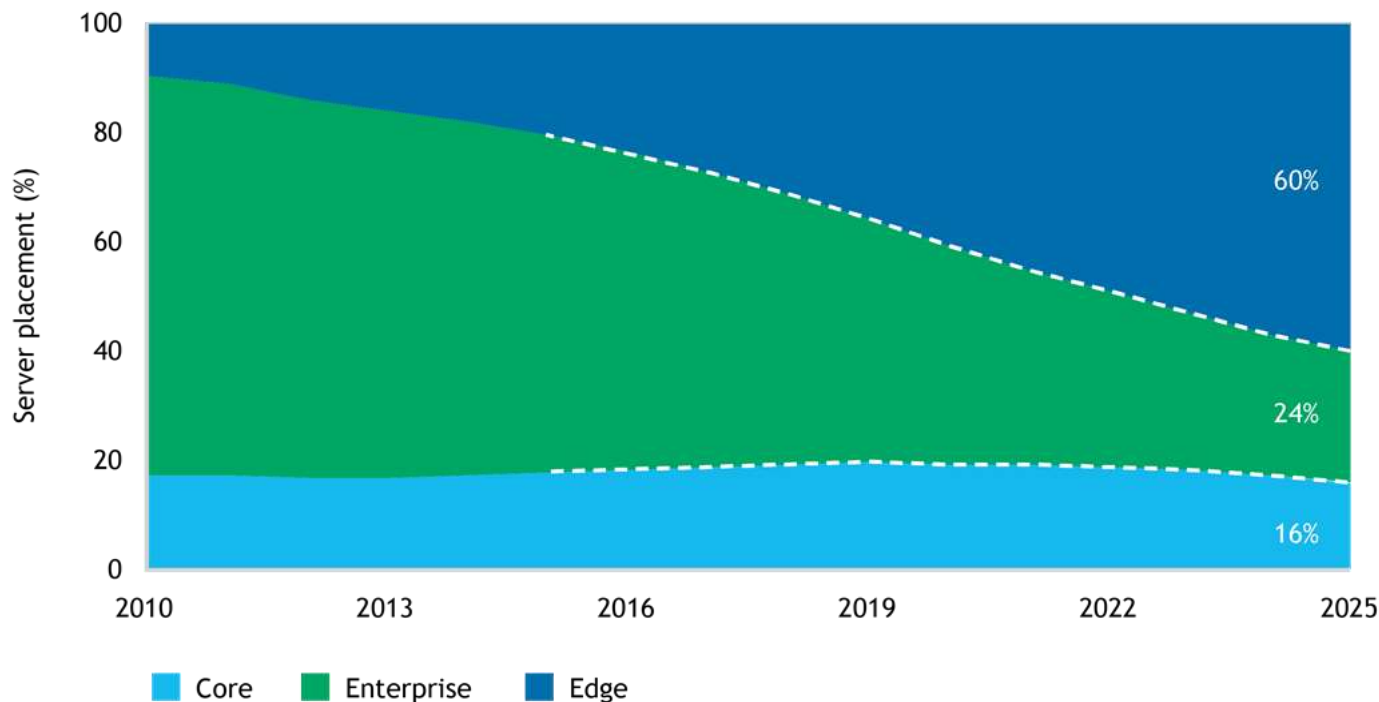
The Growth of the Network Edge



Global CDN providers are bringing CDNs as close as possible to end users

A New Paradigm: The Global-Local Cloud

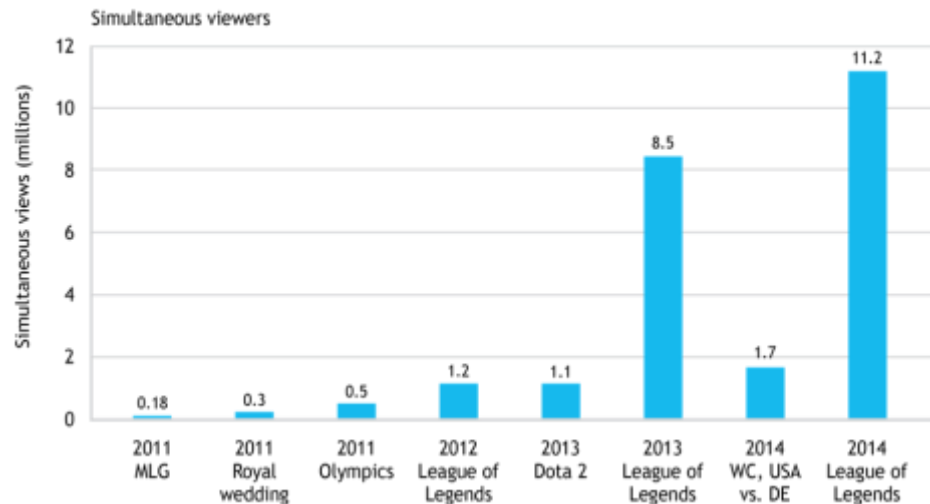
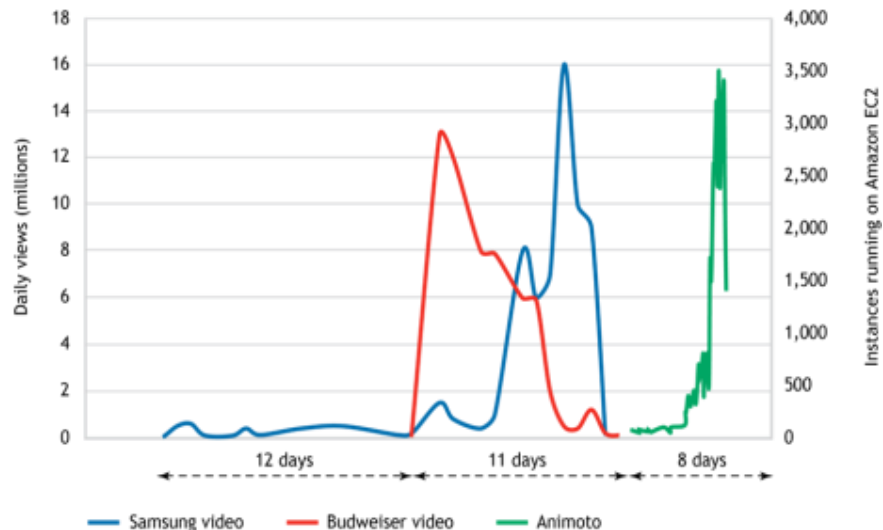
Binging the Cloud Closer its Users



Edge cloud growth will outpace growth of centralized clouds

The Dynamicity Challenge


Rapidly Changing Popularity of Applications and Content



The Future of Cloud Driven by Content and New Applications

Continued growth in video





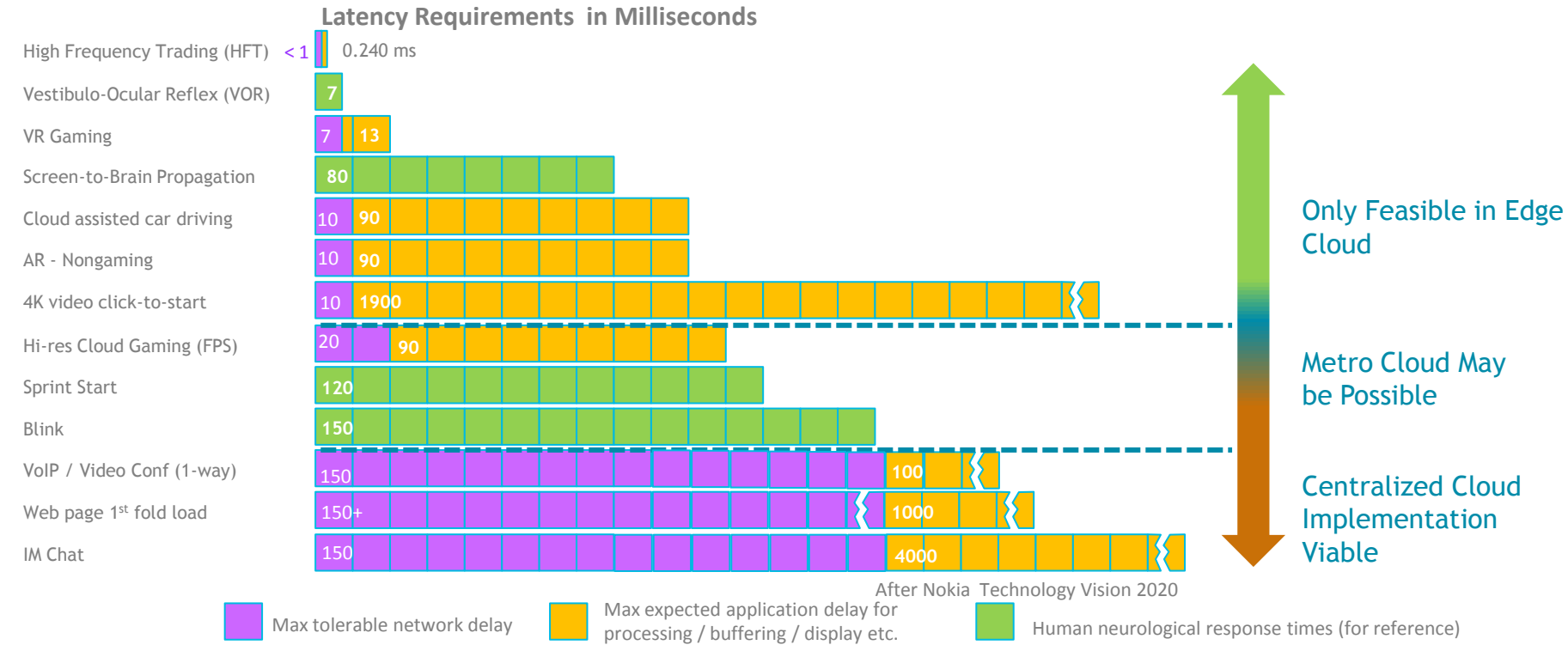
The Future of Cloud Driven by Content and New Applications

Local processing of IoT data streams

The Future of Cloud Driven by Content and New Applications

Virtualized mobile network functions

Application Latency Requirements In Comparison with Neurological Response Times



Building the High-Scale Cloud

- Creating massively scalable communication solutions over resource constrained platforms and networks
 - Hide resource limitations,
 - Adapt based on application and user needs,
 - Effective resource management and failure prevention/recovery.

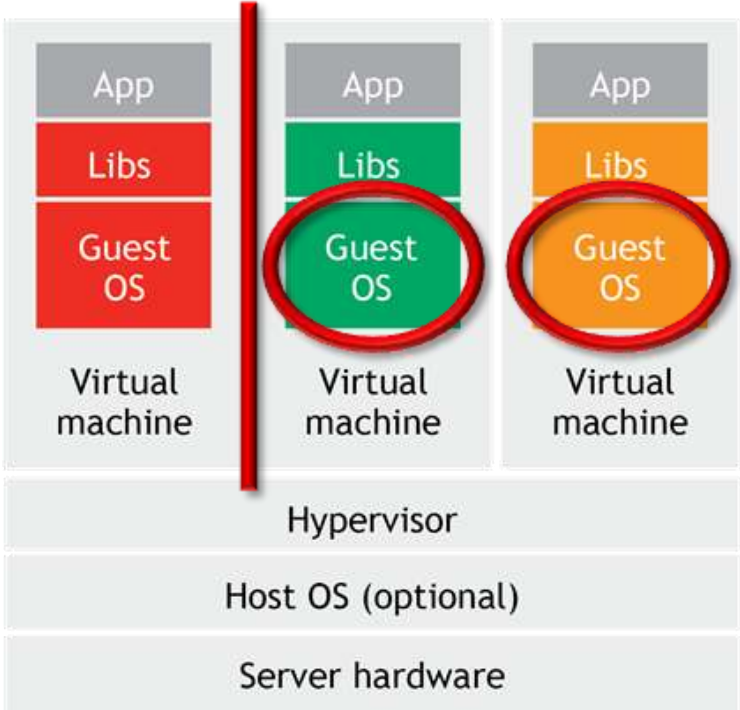


- **Challenges:** the needs of agile services require new software and networking technologies
 - How to virtualize service components?
 - How to connect components of a distributed service?
 - How to place, monitor, scale and manage services?

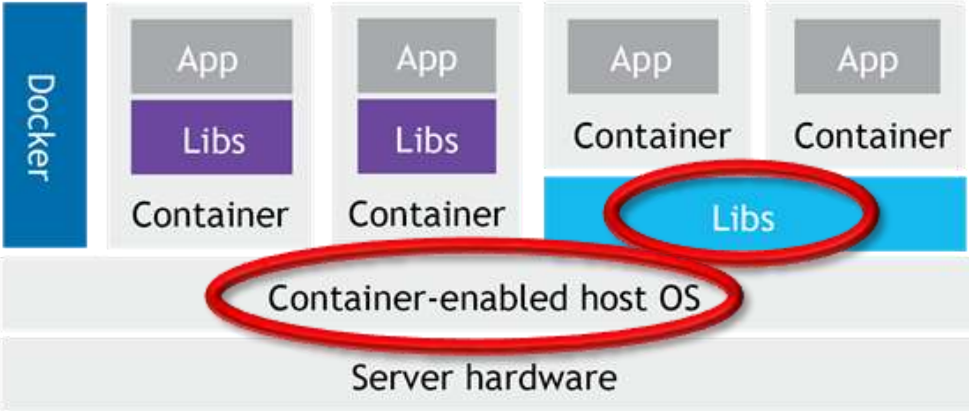
Virtualization vs. Containerization

Looking Behind the Hype

Classical virtualization



Lightweight containers



Virtualization vs. Containerization

A Comparison

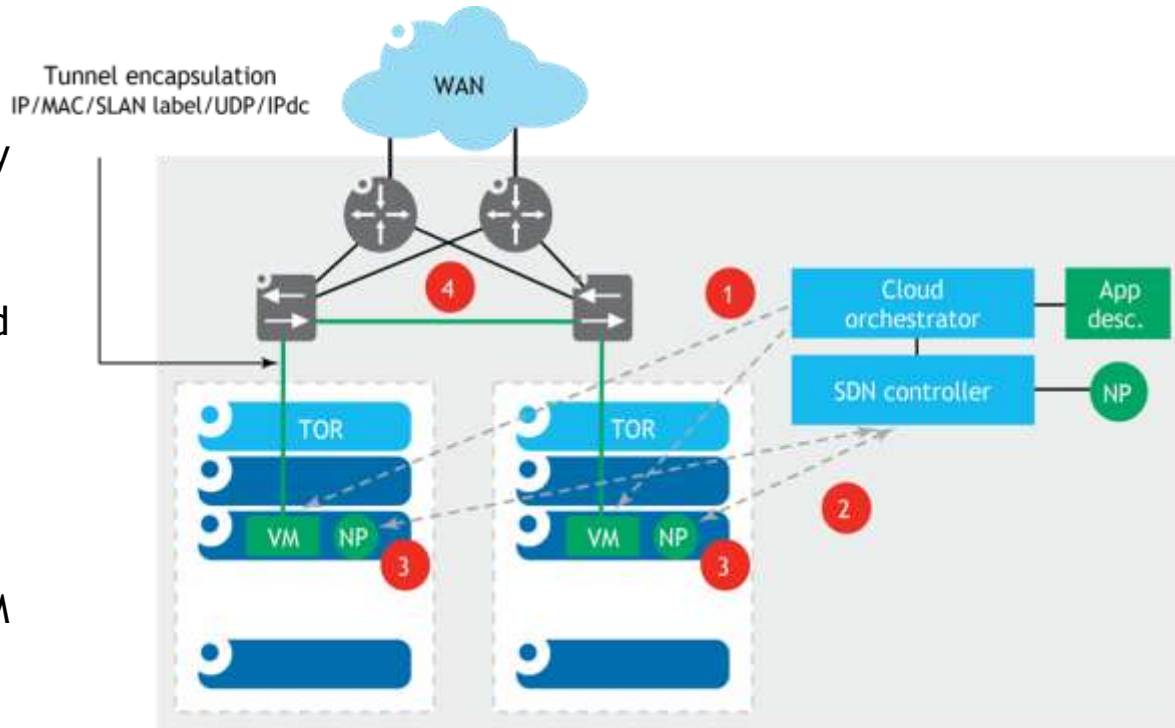
Criteria	Virtual Machines	Containers
Performance overhead (disk I/O)	> 50%	Negligible
Relative image size	>3x	1x
Boot time	10s of seconds	Few seconds
Performance overhead (typical workload)	> 10%	< 5%
Isolation of applications	Good	Fair
Management ecosystem	Mature	Evolving
Security concerns	Low-Medium	Medium-High
Impact on legacy application architectures	Low-Medium	High
OS flexibility	Excellent	Poor

- Containers and VMs are attractive for specific domains and will continue to evolve
 - Containers: micro-services designs, performance-sensitive applications and flexible deployments
 - VMs: multi-vendor environments, heightened security and legacy software

Networking the Distributed Cloud

Highly dynamic applications require the network to automatically adapt connectivity

- 1) Cloud orchestrator instantiates service VMs AND sends required network policies to SDN controller
- 2+3) Each VM retrieves required policies when it is instantiated
- 4) Virtual tunnels between the VM components are created



SDN provides this automation within and between the distributed data centers

Global-Local Cloud Business Models

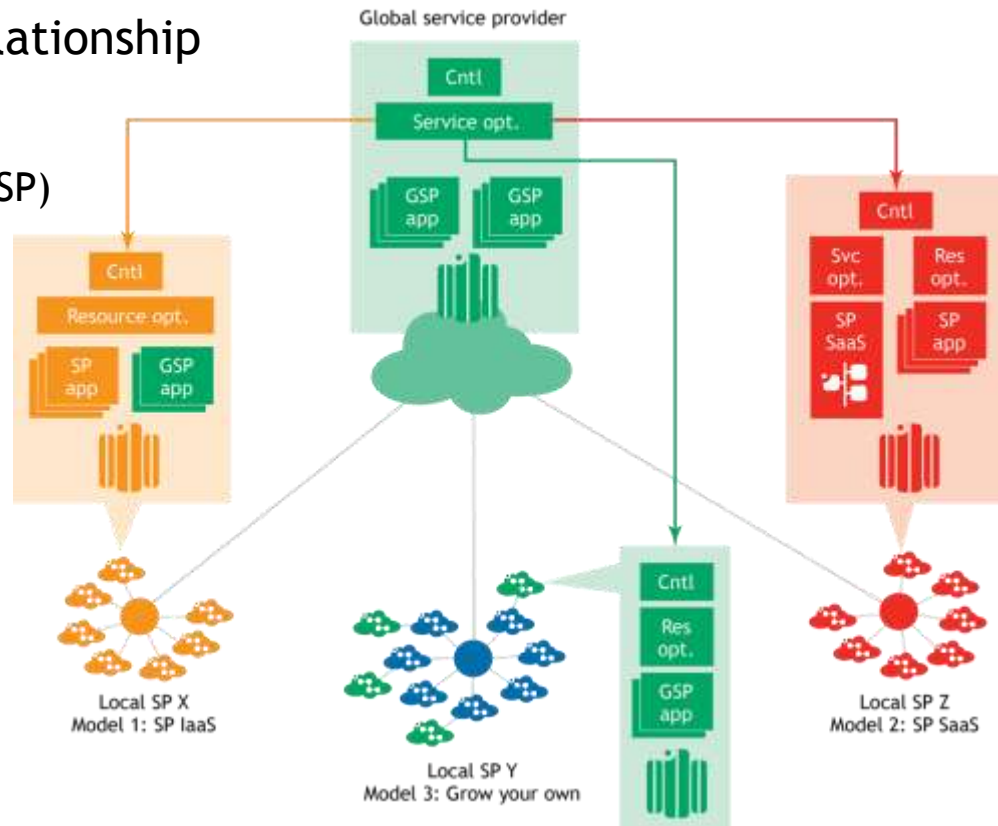
The Local/Global Service Provider Relationship

1. Local service provider (LSP) provides infrastructure for global cloud services (GSP)

- LSP offers cloud infrastructure to GSP and GSP operates services
- Limits LSP's participation in value chain and GSPs ability to fully optimize services

3. GSP provides local infrastructure

- GSPs will grow own local infrastructure
- Expensive but enables GSP to fully optimize service



Global-Local Cloud Business Models

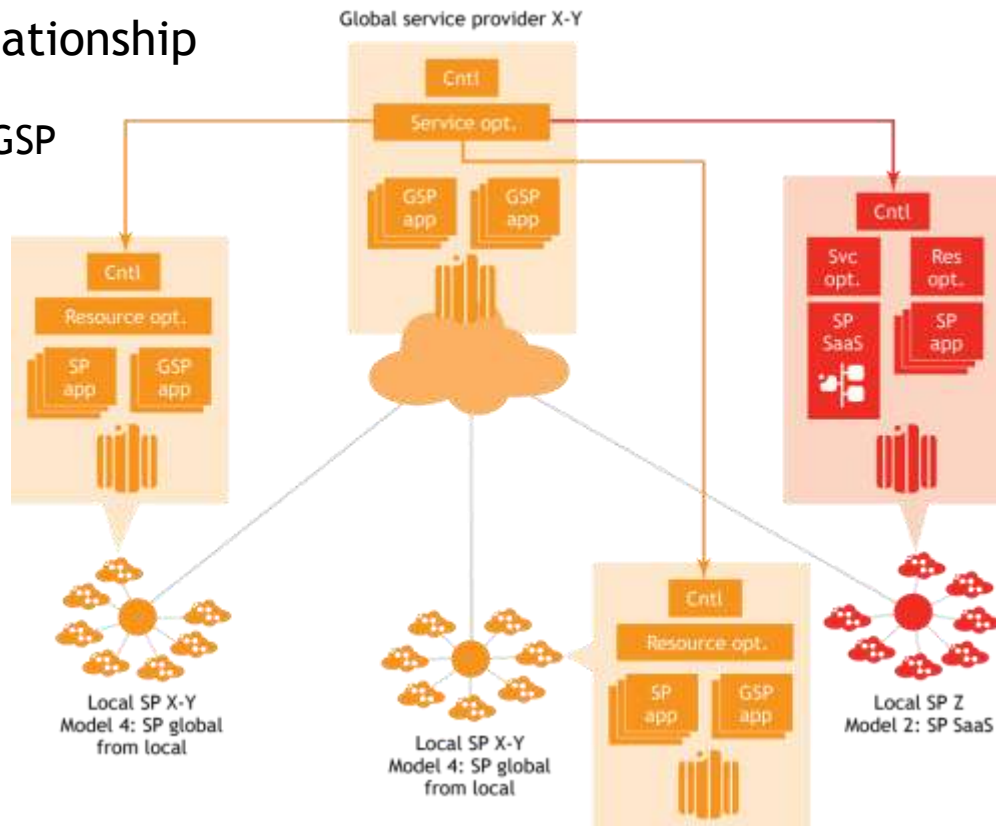
The Local/Global Service Provider Relationship

2. LSP provides hosted service functions for GSP

- Example: LSP offers CDN service to GSP
- LSP participate in value chain but limits GSPs control over the service

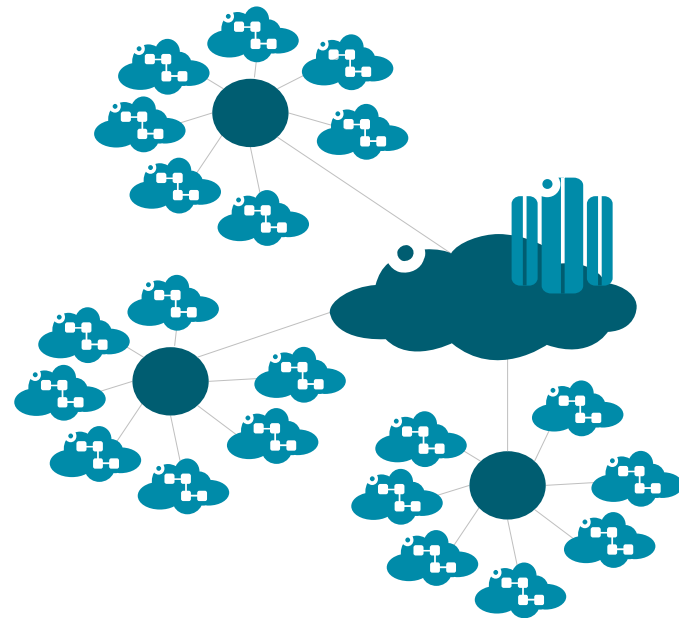
4. LSP provides global service

- LSPs create global service (e.g., through mergers or partnerships)
- Similar to model 3 but driven by LSP. LSP need to create an attractive global service.



Summary

- We are at the brink of a new era of cloud computing
 - Highly distributed
 - Constantly changing workloads
- Challenges: the needs of agile services require new software and networking technologies
 - How to virtualize service components?
 - How to connect components of a distributed service?
 - How to place, monitor, scale and manage services?
- Bell Labs is creating new approaches for NFV orchestration leveraging advanced service models and prediction algorithms



State Explosion in Software-Defined Networks*

*Original material: Adishesu Hari

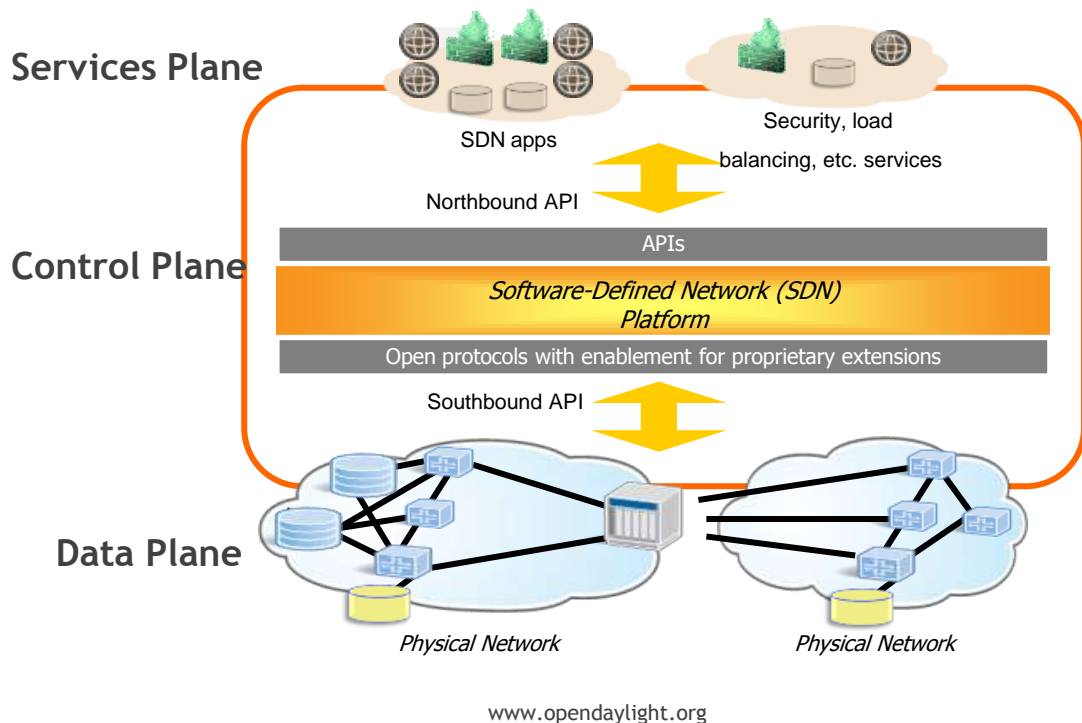
Introduction: What is SDN?

SDN:

- Centralized Control Plane
- Distributed Data Plane
 - Richer 5-tuple per flow state
 - State installed reactively or proactively

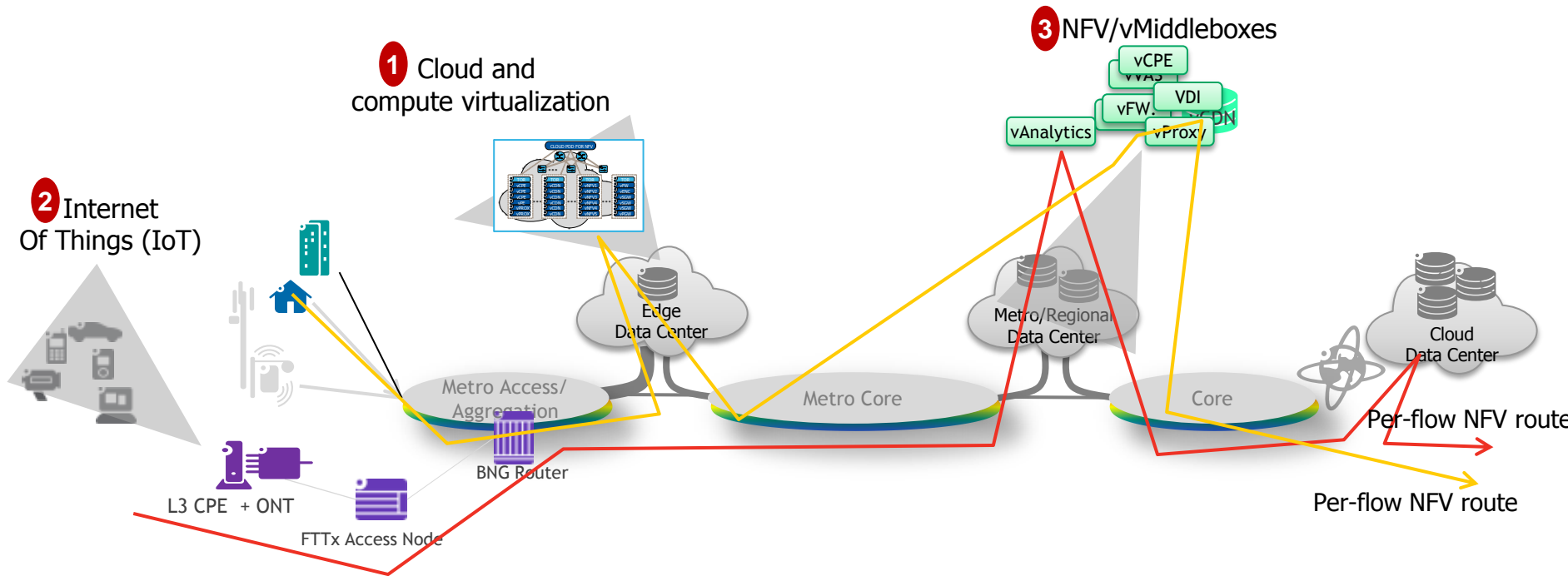
Why SDN:

- Moving intelligence to the controller enables better network management and services
- Move control plane state from switches to controller
- Provide personalized, per flow services



SDN to the Extreme: Can we reduce data plane state as well?

Problem: Data Plane State Explosion

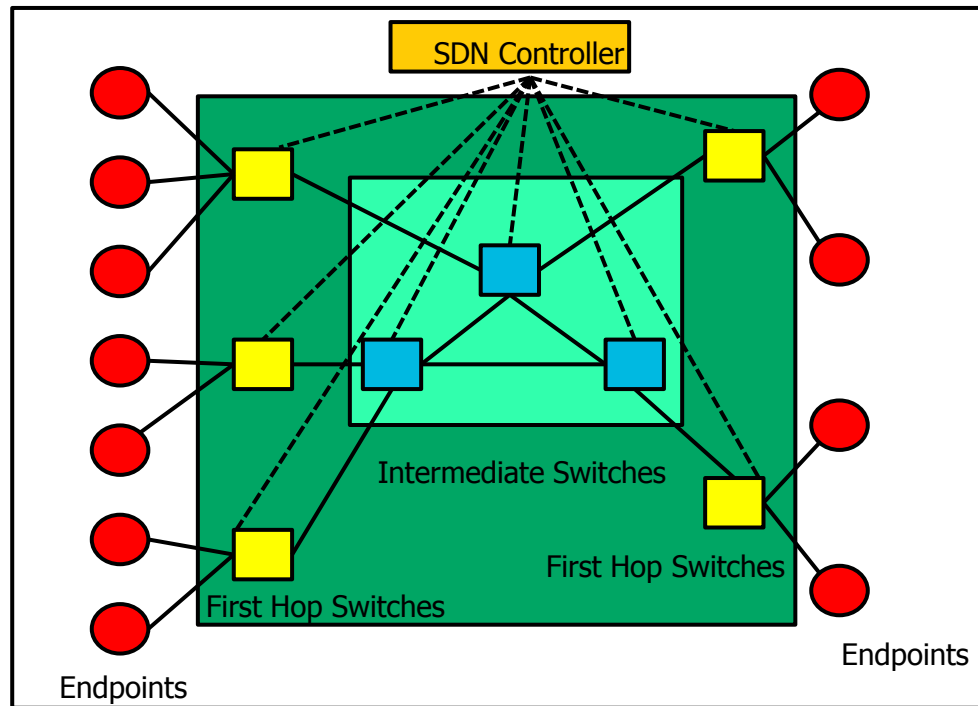


IP Prefix aggregation and tunneling cannot provide per flow service chaining needed for NFV

How to Reduce Data Plane State

Driving Principles:

- End Systems unchanged
- Forwarding State only at the edges
- Solution must work for both SW and HW solutions (no complex HW needed)



Path Switching will eliminate forwarding state in intermediate switches

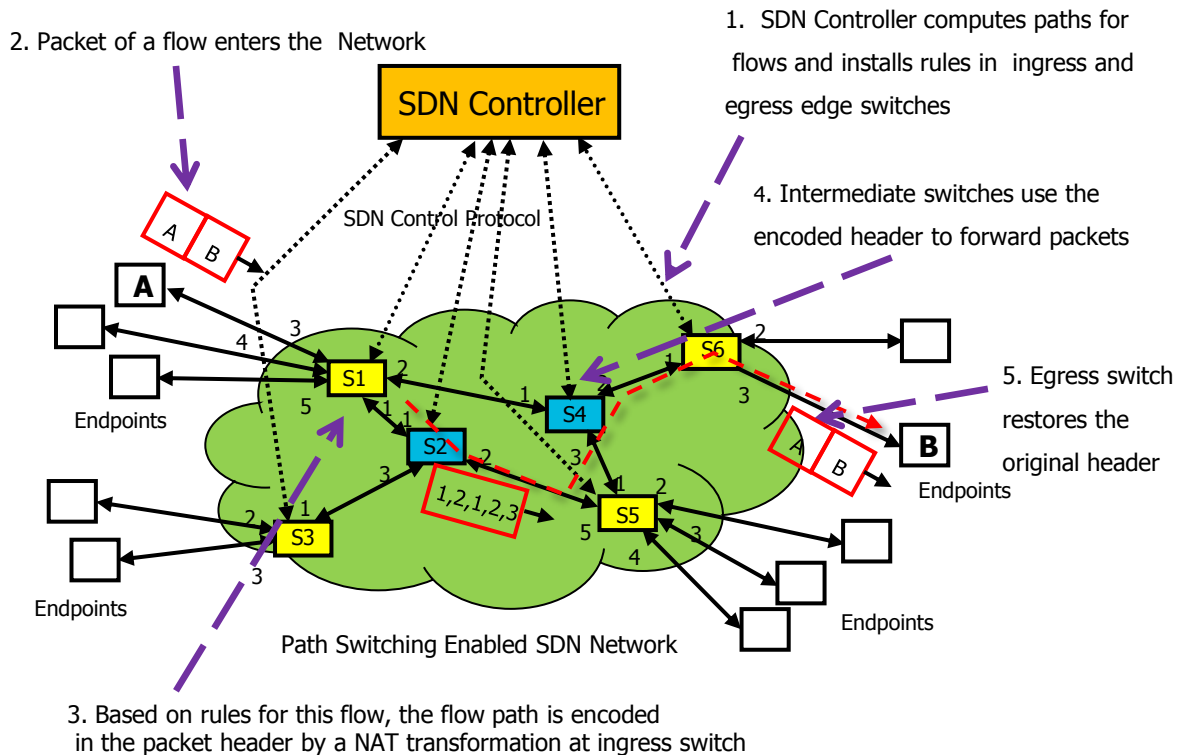
Solution: Encode Path of Packets in Packet Header

Path Switching

- is SDN friendly
- topology agnostic
- embeds network paths in the existing network address fields
- provides solution for NFV/middlebox routing

Technical Challenges:

- Efficient algorithms for encoding paths in fixed size packet headers (ISIT 2015)
- Efficient systems design and implementation with zero changes to SDN protocols and edge switches



Path Switching: Fine-grained, aggregation-free per-flow routing in the SDN data plane

Source Routing

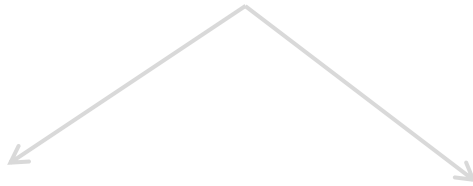
Regular Packet



Source Routed Packet



Source Routing



Loose Source routing

- Need forwarding state for all nodes
- Loose Traffic Engineering

Strict Source routing

Path Switching is not Source Routing

Transforming Source Routing to Path Switching

Regular Packet



Source Routed Packet



Fixed per-hop entries in variable sized header

Shrink path to fit within a path header

Replace src/dst fields in packet header with path header



Path Switched Packet

Variable sized per-hop entries in a fixed size header

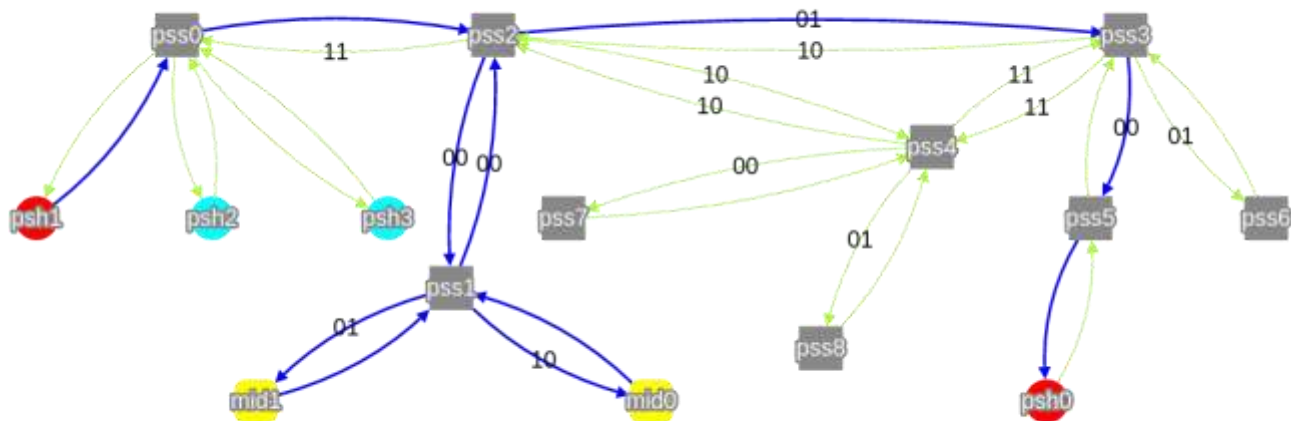
Path Switching is not Source Routing

Comparing Source Routing into Path Switching

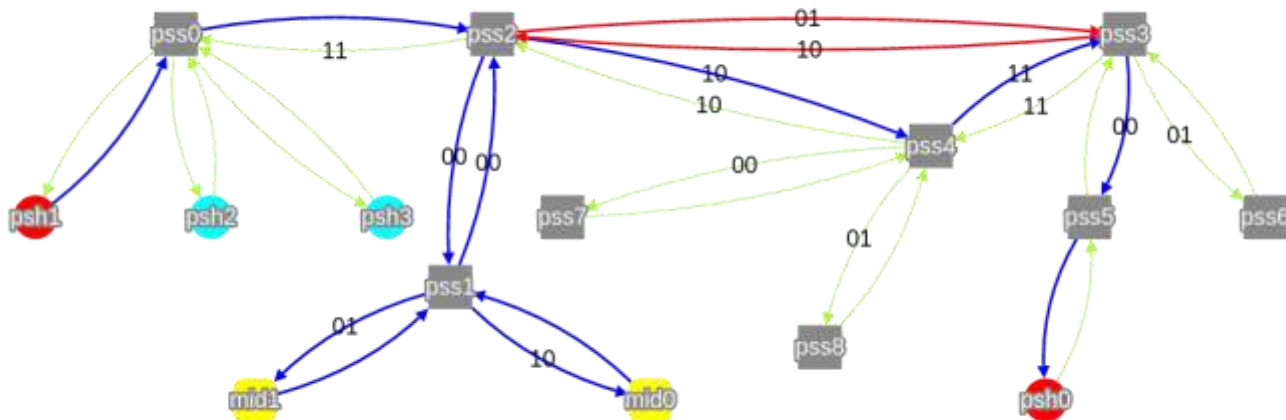
	Source Routing	Path Switching
MTU	Reduced, varies per path	Fixed, same as original
Virtualization	Separate virtualization header	Built-in
Edge Switches	Need to know source routing header	Need to know NATing
Middleboxes	Need to know source routing header + virtualization header	Only see plain packet with L2 Path Switching
Config Protocols	Need to know source routing header	Any string based control protocol

Path Switching = Source Routing without overheads

Path Switching System Demo - Fast Reroute 1



System in normal state, link pss2-pss3 active



Path Switching: Other Advantages

- **No change in MTU:**
 - All paths have same MTU. Full network utilization and dynamic path switching
- **No change to SDN control protocols or packet formats:**
 - Works with existing protocols
- **Traffic Engineering:**
 - Each flow can be routed along a different path, providing better network utilization
- **SDN compatibility:**
 - ideally suited for centralized control plane
- **Follows end2end principle**
 - State only in edge nodes, not in the core
- **Provides consistent updates and easy maintenance:**
 - Easy to update core nodes without disrupting the network
- **Integrated Network Virtualization**
 - Same technology used to create paths can also be used to restrict them
- **Green Forwarding Elements**
 - No power hungry CAMs/TCAMs
- **Single Flat Network**
 - Flattens networks of any projected size
- **Unidirectional paths**
 - Each direction can be optimized independently

Path Switching provides numerous advantages beyond per flow routing

Path Switching: System Design Issues

- How to encode variable length paths in fixed size headers?
- How to provide compatibility with existing conventional and flow switched traffic?
- How to encode the endpoints?
- How to support virtual networks?
- Micro flow routing
- Macro flow routing
- Scaling the Path Switching control plane
- Multicast
- Linux kernel Implementation with OVS
 - Basic flow routing
 - Fast reroute
 - In band Service chaining
- Testbed implementation
 - Web based graphical management and monitoring
 - Advanced services

Encode path of a packet in its header to scale per-flow state in the SDN data plane

Issue 1: Encoding Interface Labels Efficiently

- Greedy Encoding

- Encode N interfaces of a switch using $\text{ceil}(\lg_2(N))$ bits for each label
- Fixed length labels for each switch, variable length labels across switches

- Optimal Encoding (ISIT 2015)

- Variable length labels within a switch
- Use prefix free encoding for unique decoding
- Define a Path Label of a path as the sequence of interface labels for the path
- Given a set of paths P , the goal of the optimal path encoding problem is to
- find a prefix-free labeling of the interfaces for each switch so that the resulting length of the longest Path Label for the paths in P is minimized.

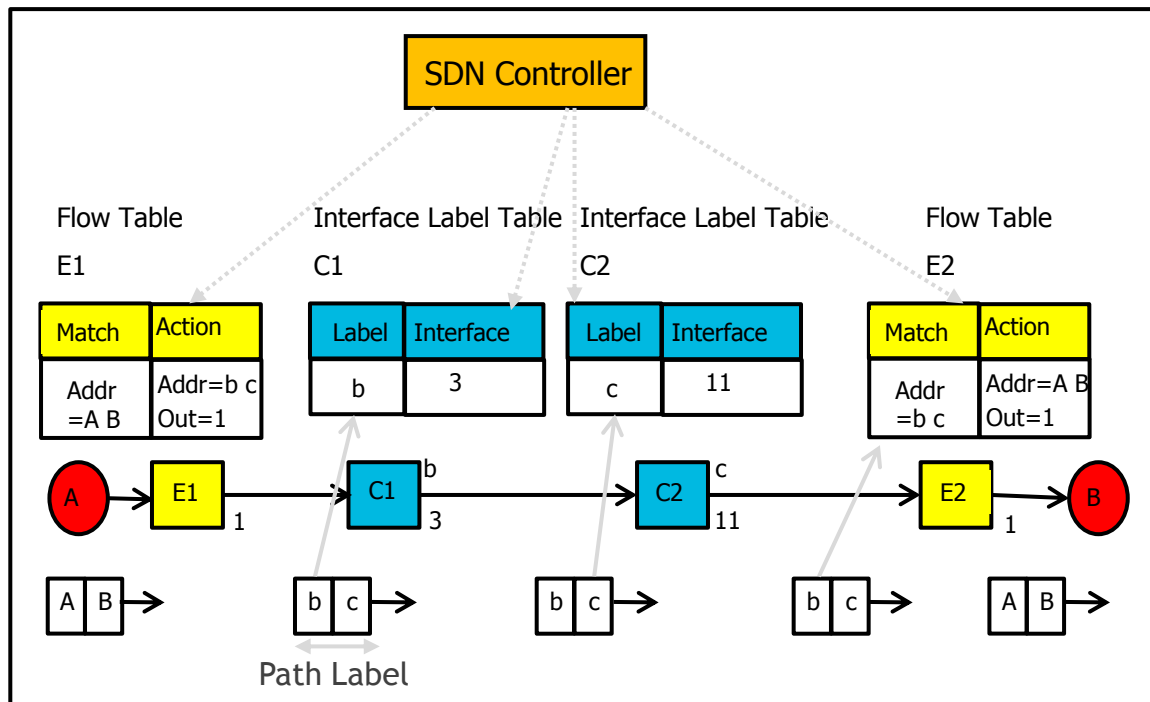
- Results

- Optimal path encoding problem is NP-hard to approximate within $8/7$ of optimal.
- Polynomial time approximation algorithm whose result is guaranteed to be within a factor 2 of optimal.
- Optimal encoding uses 30% less bits than Greedy encoding, based on RocketFuel topologies

Fit the longest paths into fixed headers using efficient encoding

Path Switching Data Structures

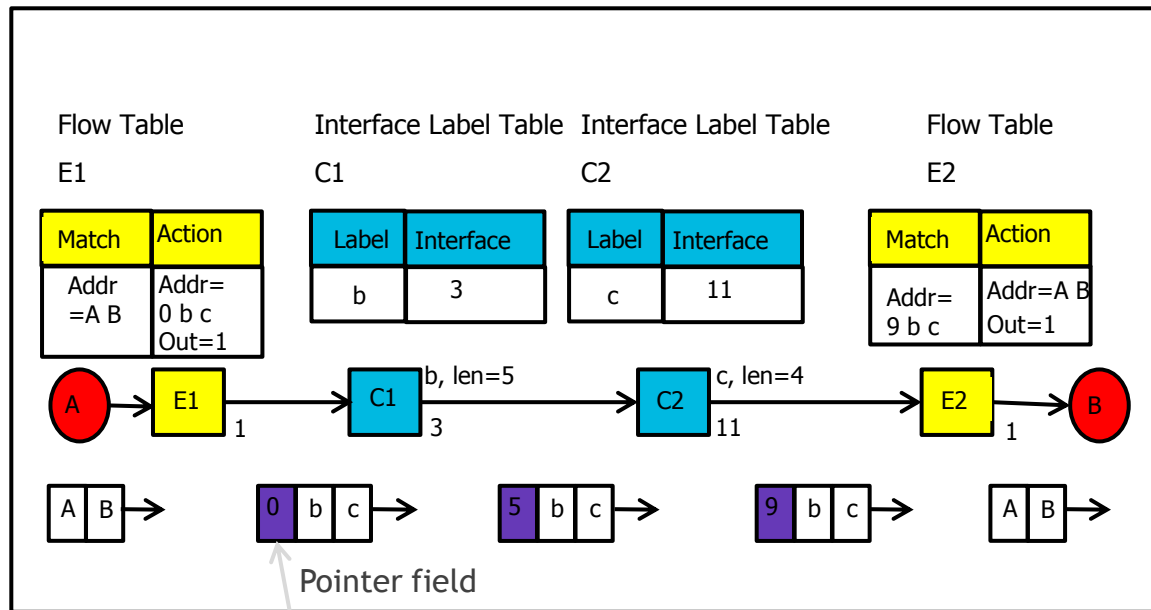
- Interface Label Table
 - Intermediate Switches
 - Maps label bit strings to interfaces
- Flow Table
 - Used in ingress and egress switches
 - Already exists in all SDN switches
 - Replaces regular addresses with Path Headers
 - Path Headers encode the sequence of labels of the path
 - One component is the Path Label
 - Sequence of interface labels



Encode path of a packet in its header to scale per-flow state in the SDN data plane

Issue 2: Parsing the Path Label

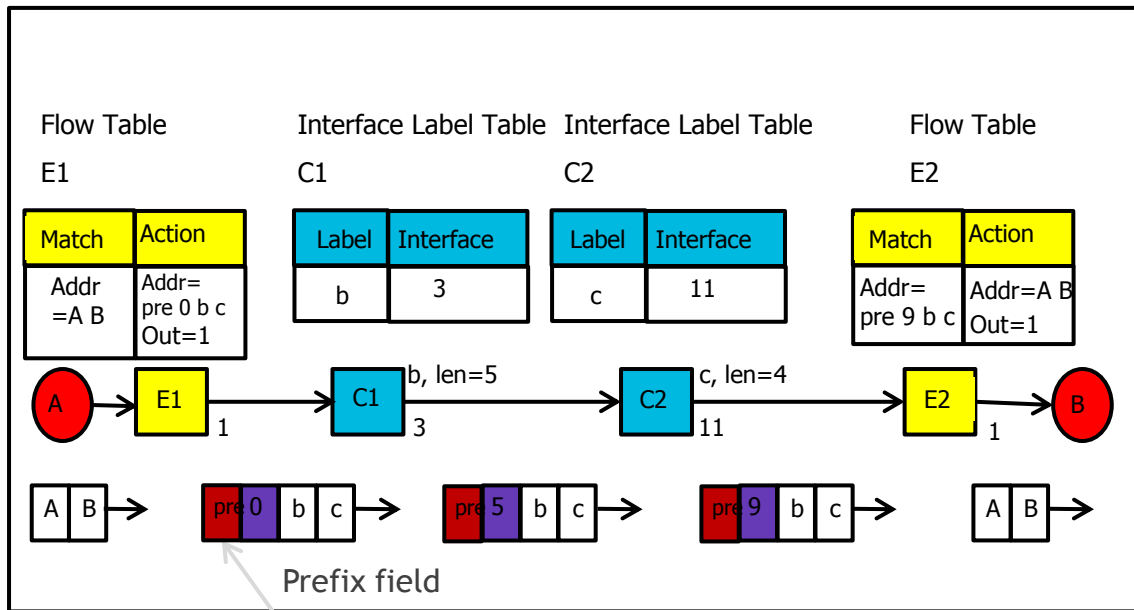
- Pointer field encodes start of current interface label
- Initialized to 0
- Updated by each Path Switch by the length of its interface label



Path Header Pointer field points to the current label

Issue 3: Coexistence with Other Forwarding Technologies

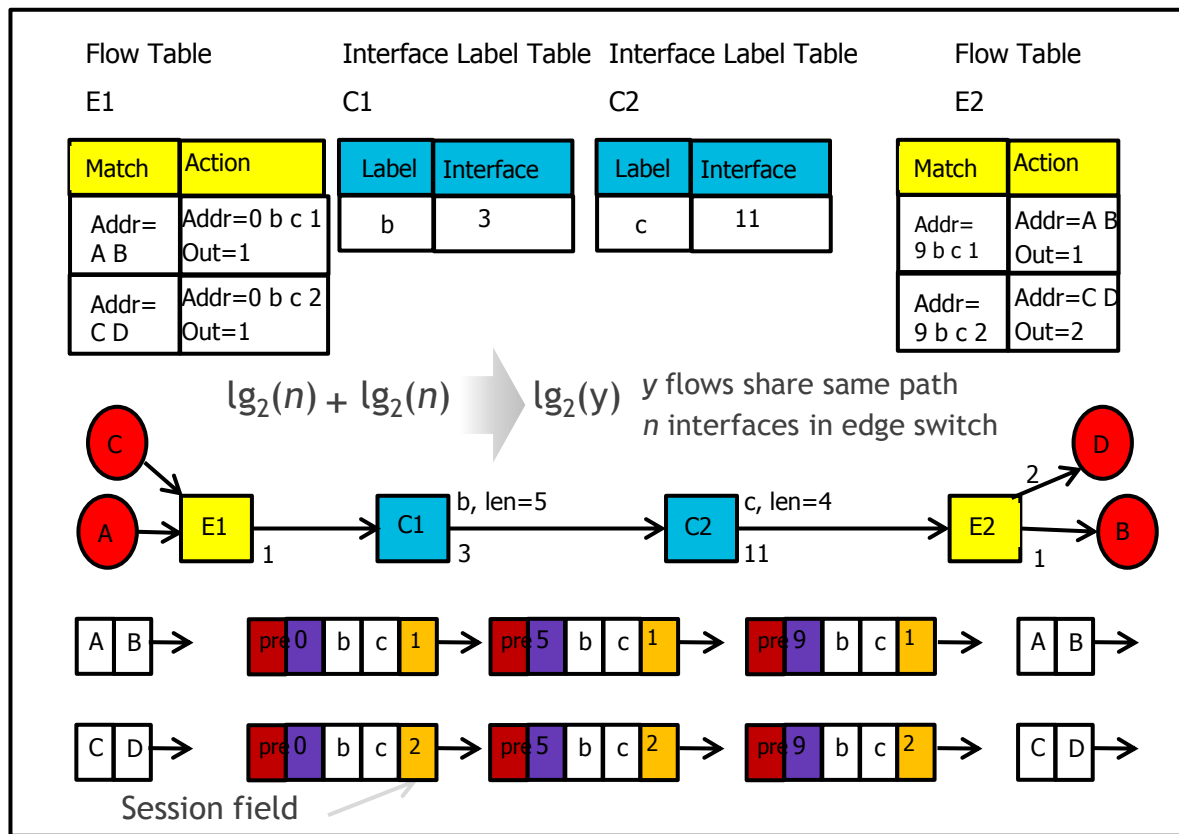
- Prefix field distinguishes Path Switched packets from other packets
- Corresponds to a private address space not present in regular packets
- Four types of packets:
 - Src/Dst regular
 - Src regular/Dst encoded
 - Src encoded/Dst regular
 - Src/Dst encoded



Path Header Prefix field allows coexistence with other packet forwarding technologies

Issue 4: Endpoint Encoding

- Session field demuxes paths between the same edge switches, but with different endpoints
- Allows endpoints to be efficiently encoded
- Allows for efficient virtual networks
- Shares space with the Path Label



Efficient endpoint encoding increases number of encoded hops and allows flexible network virtualization

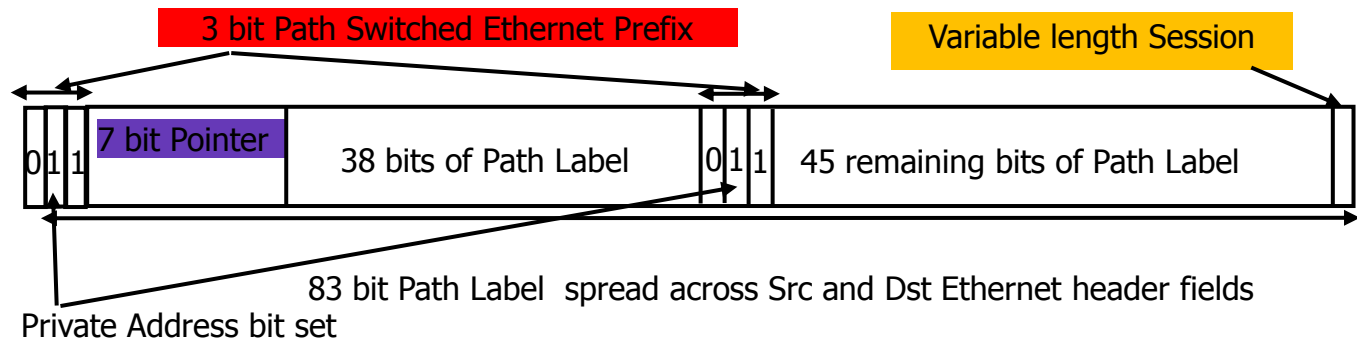
Issue 5: Which Layer to Apply Path Switching to?

- Path Switching can be applied at IP/MPLS/Ethernet layers

- **Best use case is Ethernet**

- Bigger address fields
- Allow Ethernet to scale (currently non scalable technology)
- Ethernet is ubiquitous
- Access, aggregation, core, DC

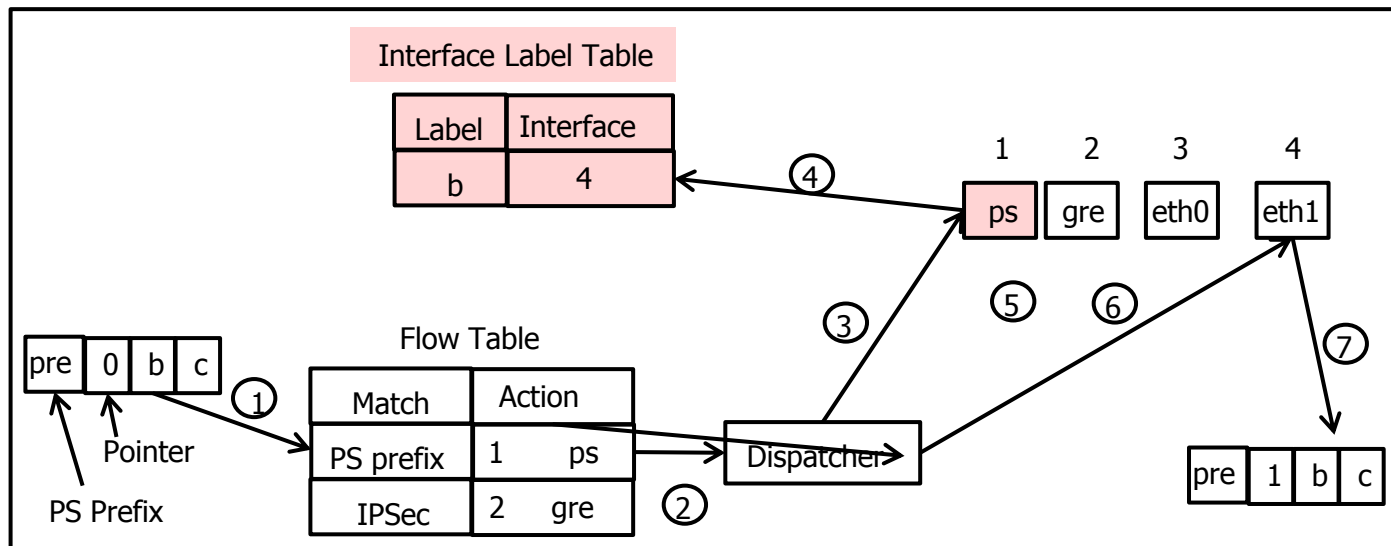
- Flat yet scalable network
- Low latency networking
 - Packet forwarding before first bit of L3 is processed
- Allows macro flow aggregation at IP level (discussed later)



Great advantages to Ethernet Path Switching

Linux Kernel Implementation

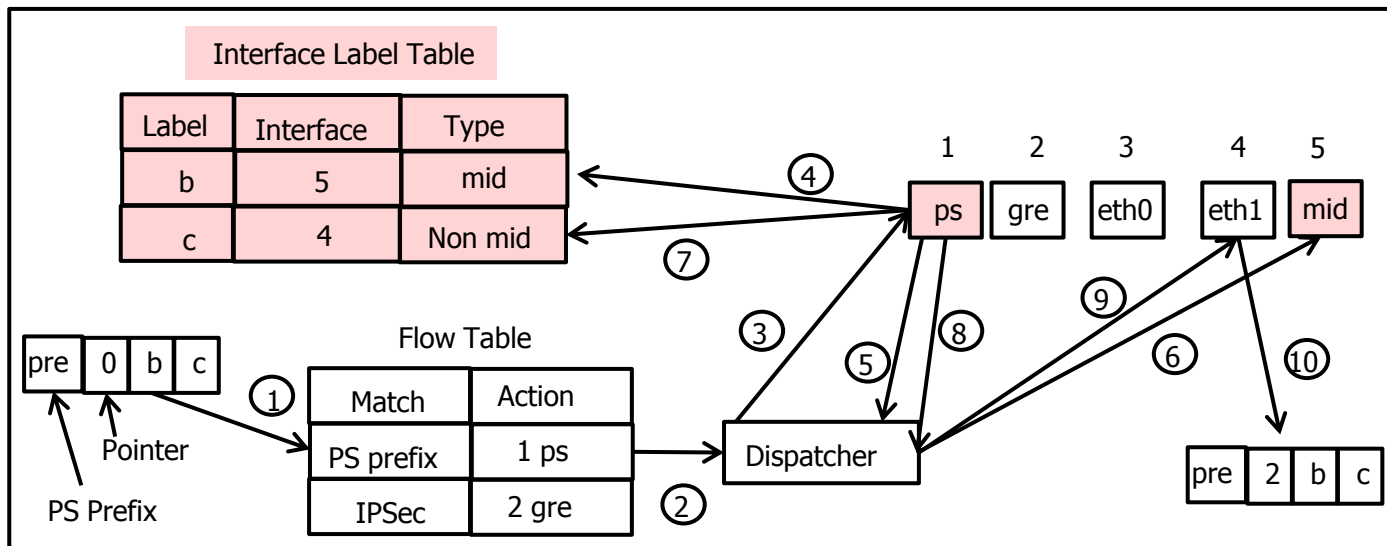
- Implemented in Open vSwitch 2.3
- No change to SDN control protocol (OF) or to config protocol (OVSDB)
- Coexists with flow switching and conventional Ethernet forwarding planes



Path Switching implemented in Linux kernel in Open vSwitch

OVS Midbox Implementation

Integrate middleboxes into OVS directly for low latency services



Middlebox framework for Path Switching added to OVS

Path Switching: Summary

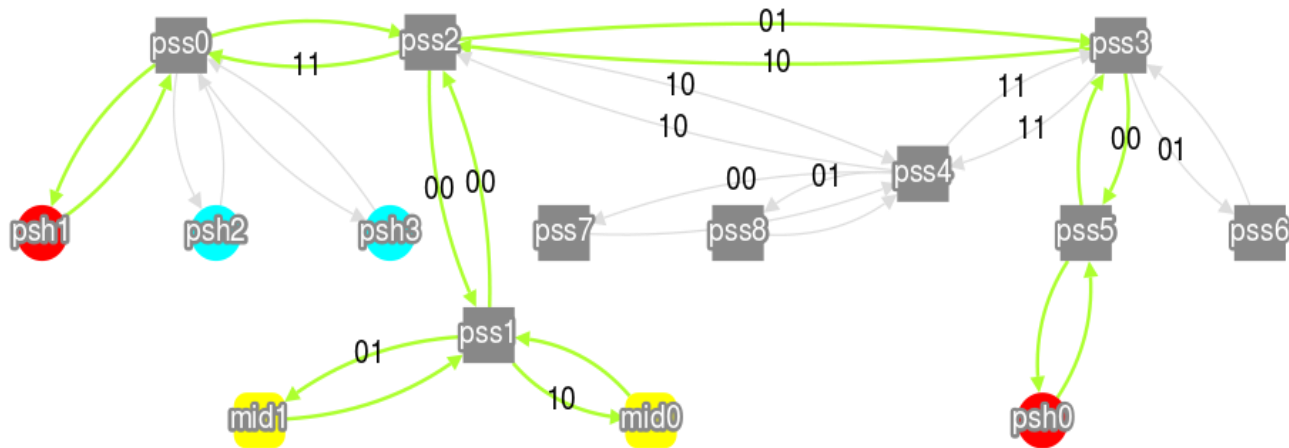
Path Switching enables fine grained per-flow paths with the following features:

- No change in MTU
- No change to SDN control protocols or packet formats
- Traffic Engineering
- SDN compatibility
- Follows end2end principle
- Consistent updates and easy maintenance
- Integrated Network Virtualization
- Green Forwarding Elements
- Single Flat Network
- Unidirectional paths

Path Switching provides numerous advantages beyond per flow routing

Web-based Path Switching Testbed

- Web based Path Switching management system
- Graphical network visualization
- Full suite of advanced services
- Network virtualization
- Asymmetric routing
- Internal and external midboxes



Advanced services managed and visualized via an HTML5 interface

Conclusions

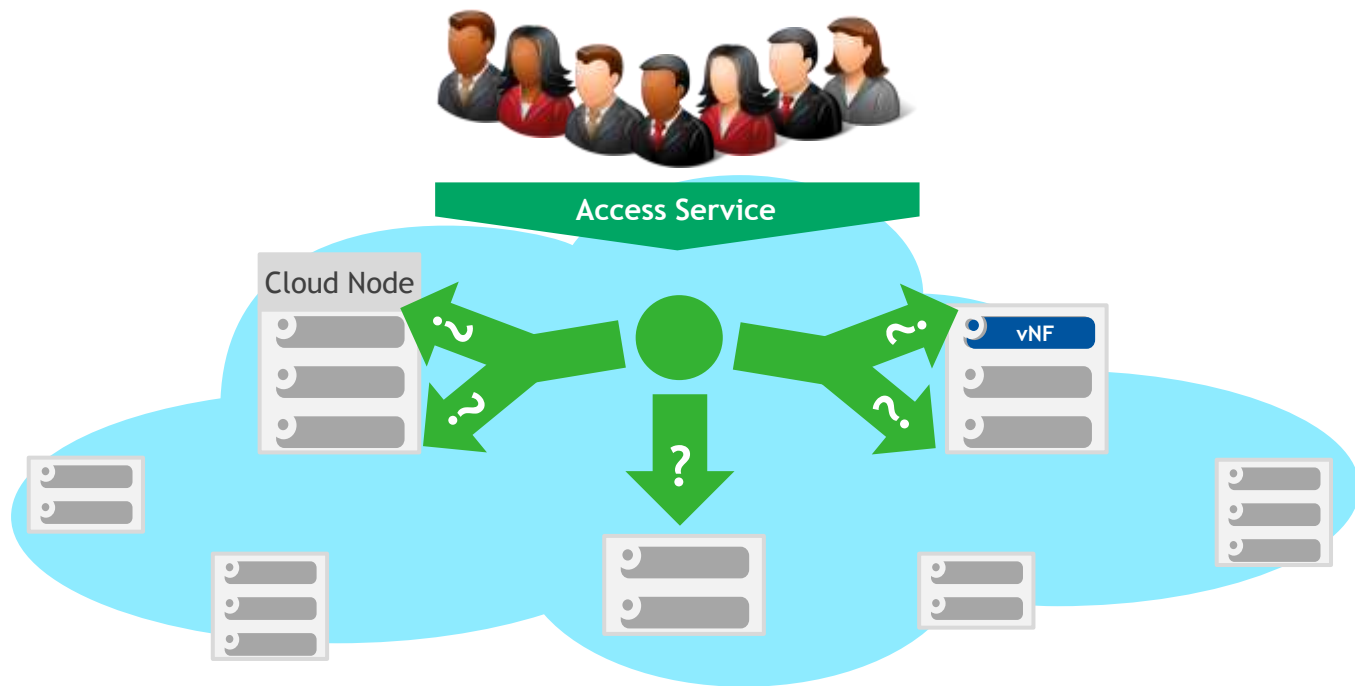
- Path Switching is a new SDN forwarding technology designed for *fine grained per flow routing for service chaining*
 - Encodes paths in preexisting network headers
- Provides many other advantages
 - Small, fixed state, power efficient forwarding elements
 - Topology agnostic and deployable over full ISP
- Efficient interface label encoding
 - Optimal Path Encoding for Software-Defined Networks. In ISIT, 2015.
- Efficient systems design and implementation in Linux
 - Adishesu Hari, T. V. Lakshman, Gordon Wilfong. *Path Switching: Reduced-State Flow Handling in SDN Using Path Information*, CoNEXT 2015.
 - <http://conferences2.sigcomm.org/co-next/2015/img/papers/conext15-final232.pdf>

SDN to the EXtreme

Load Balancing for Virtualized Network Functions*

*Original material: Ivica Rimac

Load Balancing is a Key Function



Shortcomings of Existing Solutions

Integrated L4/7 load balancer appliances:

- Expensive and bloated
- Choke point in a service chain

Conventional L4 switches:

- Break flow stickiness in dynamic cloud environment
- Don't support dynamic load distribution policies

OpenFlow switches:

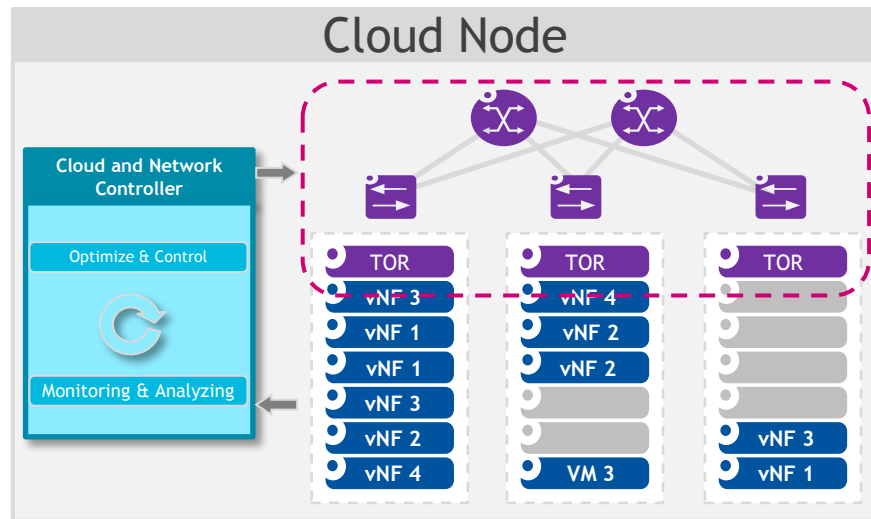
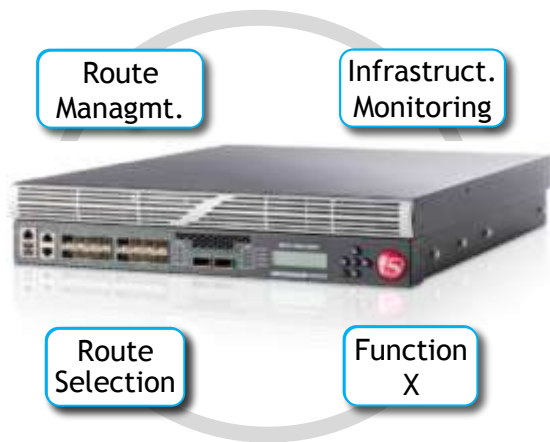
- Compromise scalability
- Inflate lookup latency and memory requirements

	Traditional Load Balancer	Cloud Requirements
Scale	42Gbps @ ~\$140k	100s of Gbps ... >10Tbps
Reliability	1+1 redundancy	N+1 redundancy
Elasticity & Dynamics	Assume static servers, static network load	VMs are added & (re-)moved, dynamic network load
Placement	Direct Server Return (DSR) supported only in same L2	VMs and LB placed across L2 boundaries
Functionality	Layer 4/7, stateful, app specific	L3, stateless, generic

“... load balancing device failures accounted for 37% of all live sites incidents [in the Windows Azure public cloud].”
Source: Patel et al. “Ananta Cloud Scale Load Balancing”, Sigcomm 2013

High cost vs. limited functionality - choose one.

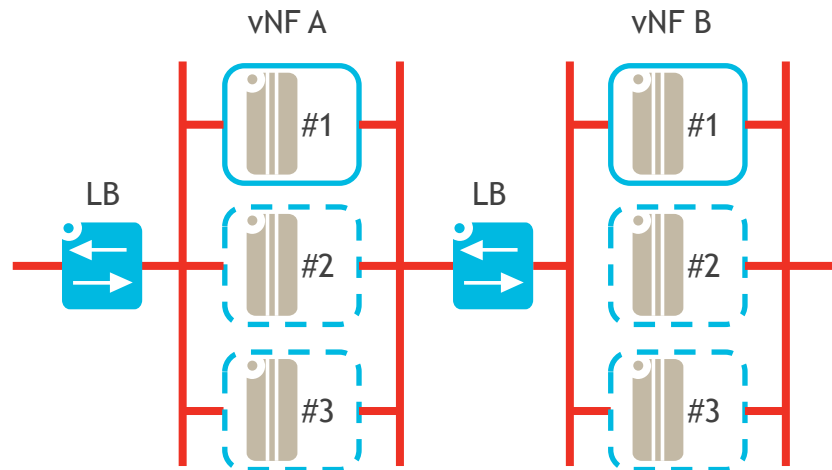
A Modern Approach to Scaling Deaggregating the Box



Agility through SDN architecture, disaggregation, and virtualized functions.

Technical Challenges

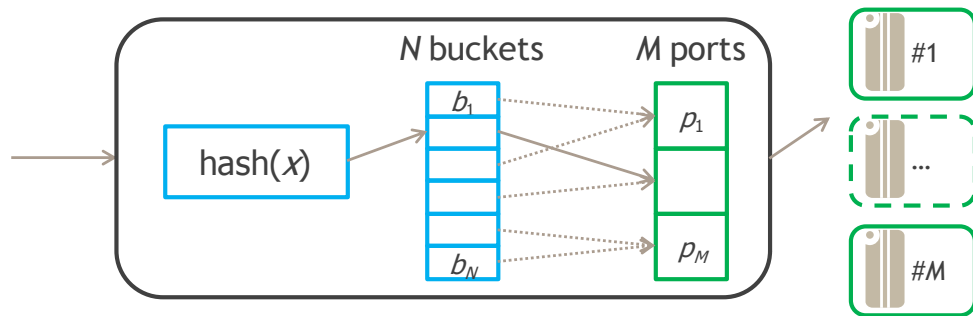
- Conventional network-layer methods are not well suited for the cloud environment characterized by high dynamicity:
 - set of VMs is dynamic (scale in/out, failure)
 - performance of any VM varies dynamically
 - capacity of any VM is unknown
- Stateful vNFs require sticky flows despite reconfiguration events



Coping with dynamicity (scale in/out, failure, load redistribution)

Addressing Challenge 1: Unequal Load Distribution

- Hash incoming packets consistently into N buckets
 - N determines the granularity of load distribution and shift;
trade-off between granularity vs. memory & lookup performance
- Partition N buckets into M disjoint sets of buckets $\mathbf{B} = \{B_1, \dots, B_M\}$
 - $B = \{b_1, \dots, b_N\}$, $\cup B_m = B$ for $m=1..M$, $B_i \cap B_j = \{\}$ for all $i \neq j$,
- Map each set B_m to one of the M ports corresponding to the M connected VMs
 - Size of set determines weight of each VM, i.e., $w_m = |B_m|$



Addressing Challenge 2: Stickiness Despite Reconfiguration Events

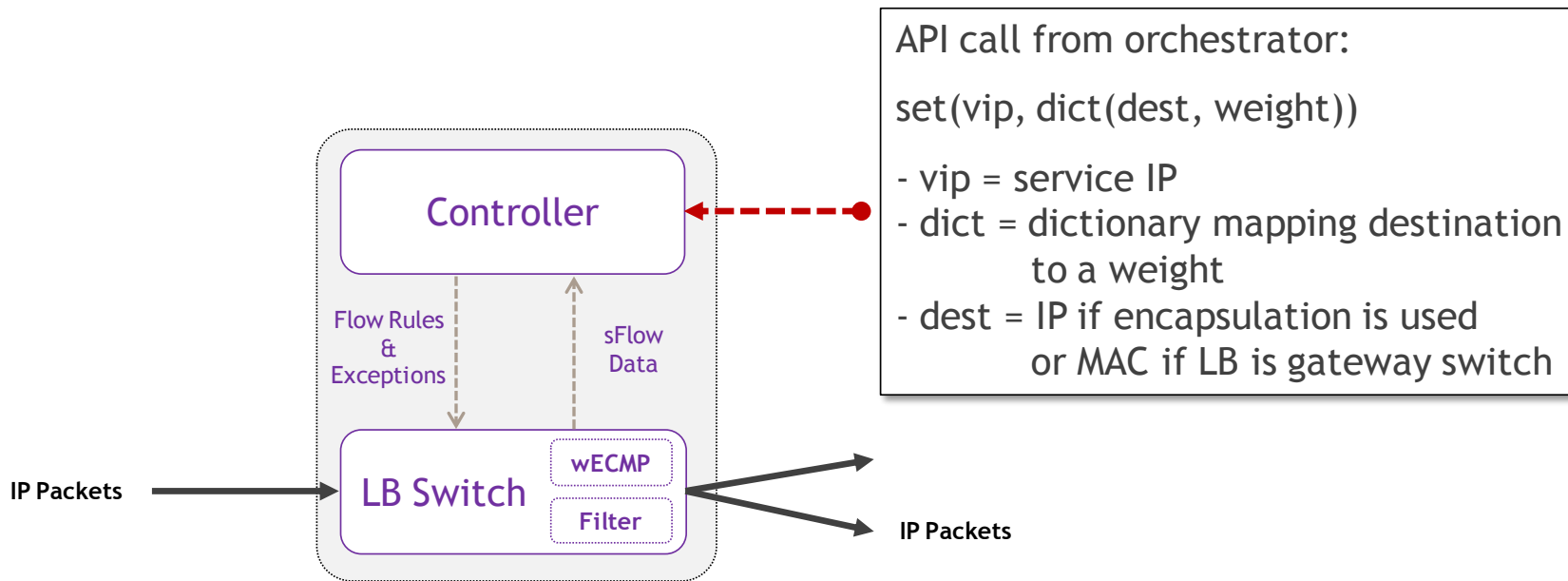
Shortcoming of conventional approaches

- ECMP rehashes on each event and breaks large fraction of active flows
- Micro-flow table for active flows does not scale

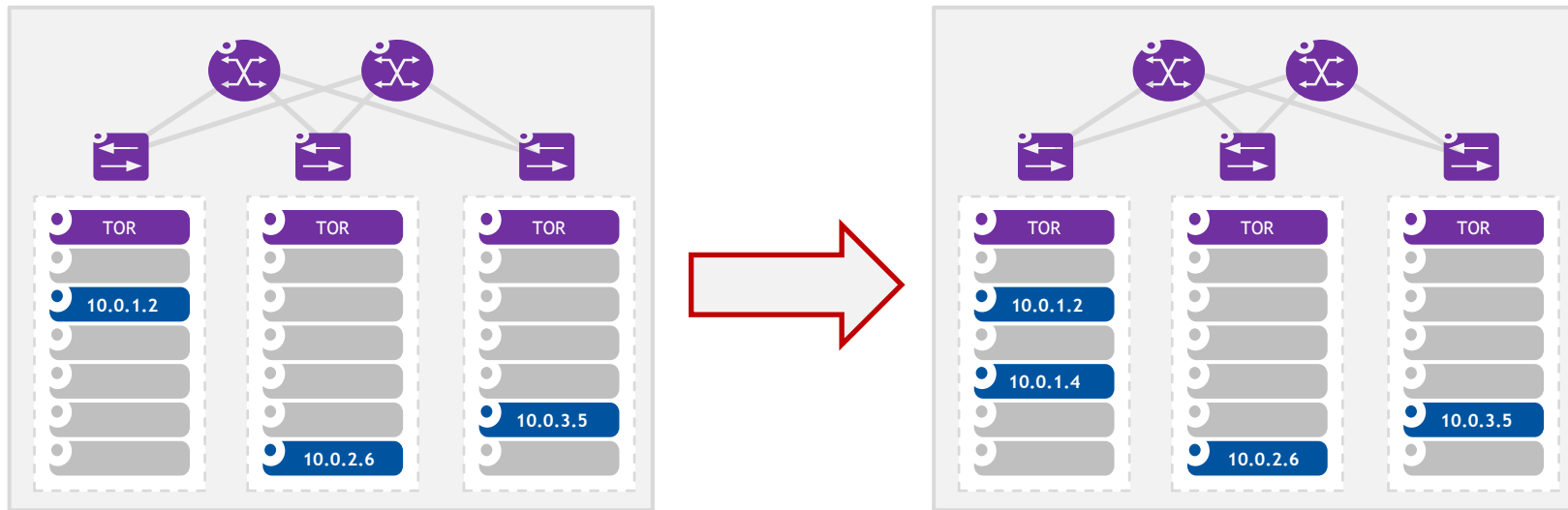
→Idea: Hashing with Exceptions

- Design of space-efficient data structures w/fast lookup
- Approximate membership lookups using dynamic BL-like structures

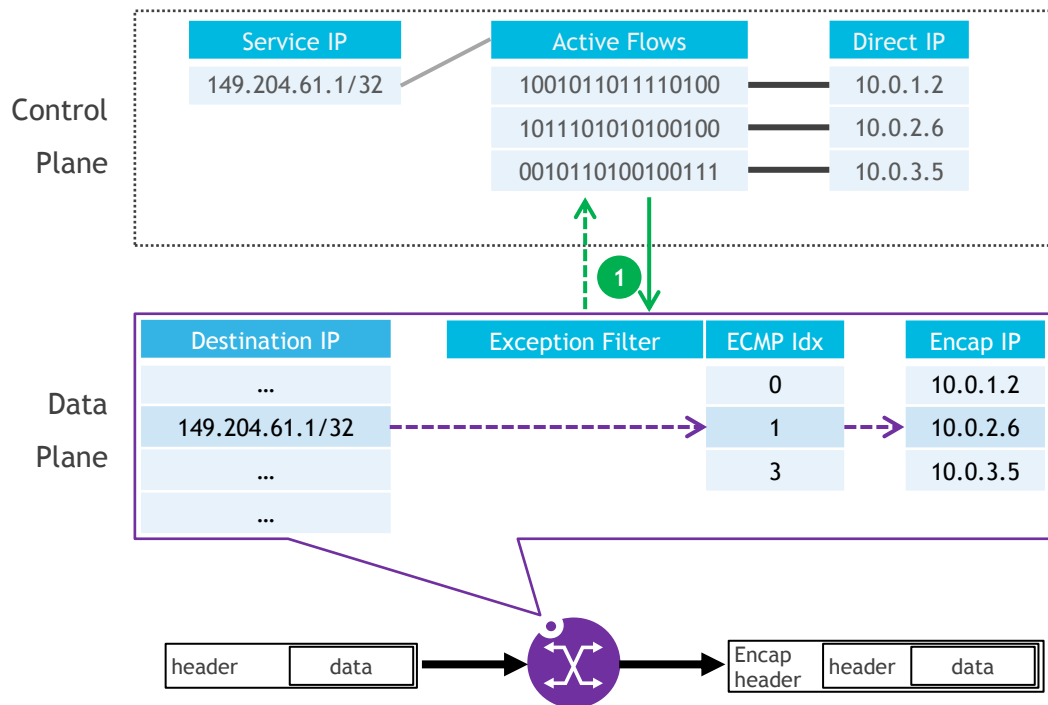
Architectural Overview



Example: Scale-Out Event

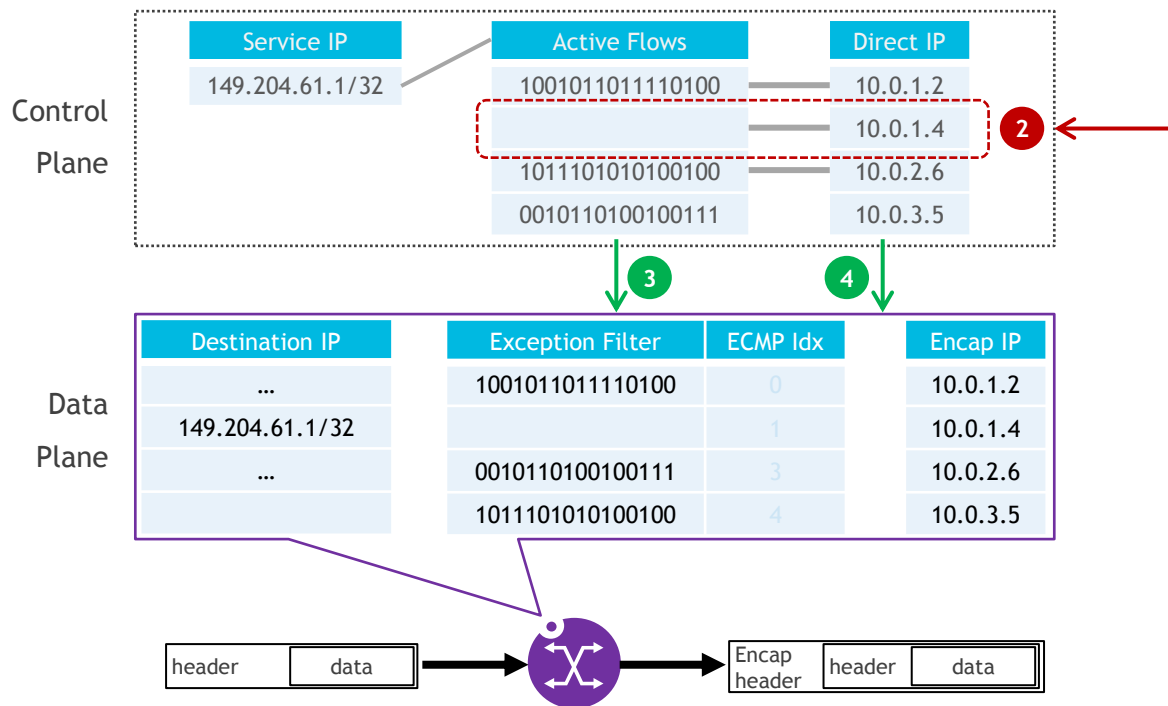


Hashing with Exceptions: Steady-State ($t=0^-$)



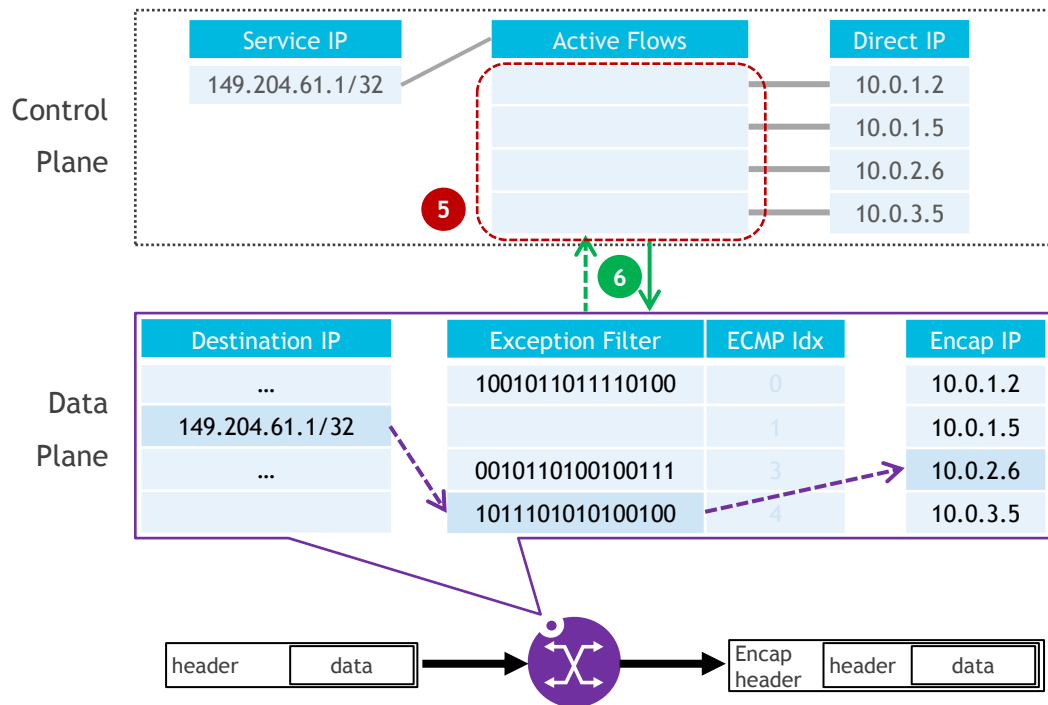
- 1 Control plane maintains info about currently active flows
 - Needs to be updated whenever a flow starts or terminates
 - E.g., by monitoring the data plane

Hashing with Exceptions: On Reconfiguration Event ($t=0$)



- 2 Control plane receives signal on scale-out event (new service instance 10.0.1.4), e.g., from orchestrator
Adds entry into its data structures
- 3 Pushes compact representation of active flows into exception table of data plane
- 4 Configures new ECMP and next-hop entries in the data plane

Hashing with Exceptions: After Reconfiguration Event ($t=0^+$)



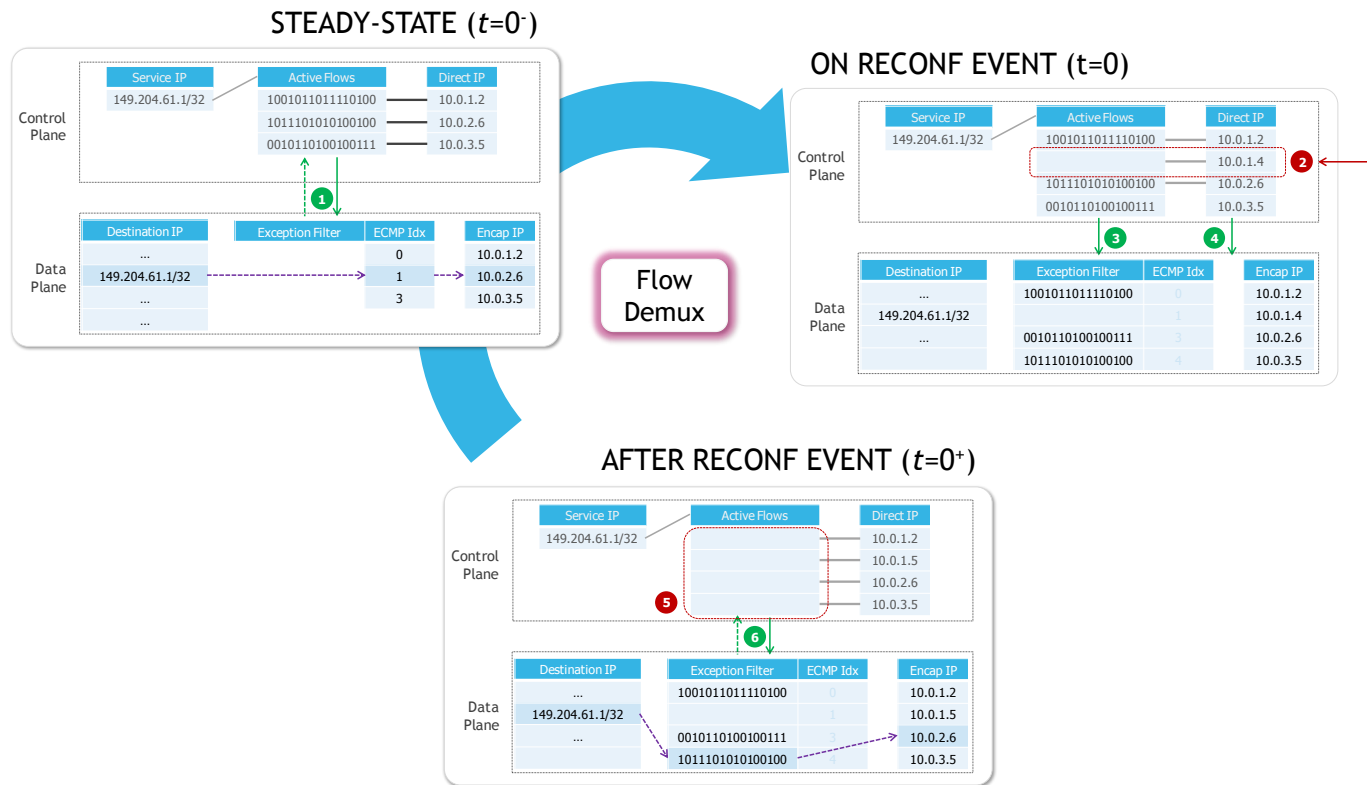
- 5 Control plane flushes active flow entries and starts new epoch

Active flow entries in the control plane need to be updated (add/del flow)

- 6 Exception filter needs to be updated whenever previously active flow terminates (del flow)

Eventually, exception filter will be empty if no other reconfiguration event occurs

Hashing with Exception: Summary



Conclusions

- Load balancing is a key function in a scale-out design
- Stateful network functions differ in requirements from stateless services

Our design principle: Blend the performance and simplicity of stateless L4 switches with adaptability of stateful L4/7 load balance appliances

Technical approach: Weighted hashing with exceptions

- Maintains flow stickyness with little incremental overhead over regular switch/router operation
- Enables runtime changes to LB policy

THE FUTURE X NETWORK

A BELL LABS PERSPECTIVE

Visit us:

www.bell-labs.com

www.alcatel-lucent.com



THE FUTURE X NETWORK

A BELL LABS PERSPECTIVE



www.bell-labs.com



linkedin.com/company/bell-laboratories



twitter.com/belllabs



facebook.com/Bell.Laboratories

