SDN Use Cases - Part 3 - Inter-domain SDN

Vasileios Kotronis
Outline

➡ Part I (~45 min)
  ▪ A brief recap of BGP and how it works
  ▪ Communication between SDN domain controllers
  ▪ Partial SDN deployment with BGP compatibility
  ▪ Outsourcing and centralizing inter-domain routing
  ▪ Control Exchange Points and end-to-end QoS

➡ Part II (~45 min)
  ▪ SDN at Internet Exchange Points (IXPs) → SDX

➡ General directions for inter-domain SDN (~2-3 min)
Routing Hierarchies in the Internet*

- The Internet = a network of networks/domains
- How do we route packets within such an environment?

- **Level 1: Routing within a domain**
  - Use an Interior Gateway Protocol (IGP) for intra-domain routing
  - Based on Distance Vector or Link State → RIP, OSPF, IS-IS, ...
  - Choice by administration of the routing domain
  - Example: SWITCH forms such a routing domain

- **Level 2: Routing between domains**
  - Use an Exterior Gateway Protocol (EGP) for inter-domain routing
  - Today’s standard is a path vector protocol, supporting policies → Border Gateway Protocol (BGP), Version 4 (BGP4)
  - Routing domains = Autonomous Systems (ASes)

*Original slide from Dr. Xenofontas Dimitropoulos for the CN 2014 course, ETH Zurich*
A basic view of the Internet*

Autonomous Systems (AS):

- Managed by autonomous entities
  - ISPs, Governments, Content Providers, …
- Have a unique AS Number (ASN)
  - Around ~46000 ASes out there!

Interior Gateway Protocols (IGP), OSPF, RIP, ...
Exterior Gateway Protocols (EGP), BGP4

Border Router
Routing information

*Original slide from Dr. Xenofontas Dimitropoulos for the CN 2014 course, ETH Zurich
Border Gateway Protocol (BGP)*

- Internet: Arbitrarily interconnected set of ASes
  - Not restricted to the Tier model or tree structures
  - Even denser than you might think (peering agreements at IXPs, etc.)

➤ BGP “is the glue that holds the Internet together“
  - Communicates **prefix reachability** information to ASes
  - Information collected by ASes is used to configure forwarding tables of border routers

➤ **Path vector** protocol
  - Exchange of routes to destinations in the form of AS path vectors (Dest_IP_Prefix, AS1→AS2→AS3→…)
  - No explicit distance metric exchanged!
  - ASes can detect routing loops by AS path analysis on route ads

➤ **Extensive support for defining routing policies**
  - Customers/Providers/Peers, TE, security, cost reduction, …

*Original slide from Dr. Xenofontas Dimitropoulos for the CN 2014 course, ETH Zurich*
Why do we need a new “SDN-BGP”? The Internet works fine as is, right?

- Current architecture does NOT support innovation
  ▪ Strong coupling between architecture (protocols etc.) and infrastructure (network devices)
  ▪ Architectural rigidity → barrier to innovation, not sufficient modularity

- Management, trouble-shooting and security are hard
  ▪ Manual configuration of several knobs
  ▪ “Masters of complexity” paradigm

- Large convergence times (~10s of seconds up to minutes)

- BGP scalability issues due to routing table size and churn
  ▪ Controller has plenty of CPU power and capacity, routers do not

- BGP routing inconsistencies can cause anomalies
  ▪ Black-holes, loops, routing disputes

- Difficulties with policy enforcement (outdated BGP knobs)
SDNi: SDN Controller Interconnection

- Main motivating factor: SDN Partitioning
  - Scalability (devices/controller)
  - Manageability (separate responsibilities)
  - Privacy (each domain on its own)
  - Deployment (SDN islands within legacy networks)

- Oriented to horizontal partitioning
  - In contrast with vertical e.g., like FlowVisor/OpenVirtex does

- Advocates interconnection between controllers
  - Each SDN domain: controlled by one SDN Controller/NOS
  - SDNi is an interface mechanism between SDN domains
  - Relates to control plane cooperation
What can be exchanged?

- Network topology (or slices thereof)
- Network events (e.g., ‘link-down’)
- User-defined request information
- User app QoS requirements
- Infrastructure status (e.g., energy consumption)
An Inter-AS Routing Component for SDN

- Component that runs on top of an SDN controller (NOX)
  - Extend “Switch” and “Messenger” modules of NOX
- Exchange of inter-AS prefix reachability information
- Routing based only on destination IP prefix as vanilla BGP
- Maintain DEST_IP_PREFIX, AS_PATH, NEXT_HOP, added NEXT_DPID and NEXT_DPID_PORT
- Loops are handled via AS_PATH checking (as in BGP)
- Essentially replicates BGP primitives over an inter-controller, SDN setup

Example: Steps for inter-domain routing decision

1. Packet flow arrives at OF Switch
2. First packet header is sent to controller
3. The packet arriving at Controller generates a Packet-in Event at Switch component
4. When prefix does not belong to domain, Inter-SDN is called
5. Inter-SDN search for a route on RIB, that return the Next-Hop and path, if found
6. When a route is found, next_hop information is returned to Switch component
7. Controller install the rules on OF Switch
8. Packet flow is sent to destination, if there is route

How can an SDN domain communicate with BGP?

Outsourcing the Routing Control Logic: Better Internet Routing Based on SDN Principles

Routing management and optimization is complex

Diverse objectives
- BGP policies
- Over-the-top service guarantees
- SLAs with client networks
- Peering agreements
- Transit cost reduction
- Green TE
- Scalability
- Security
- Etc…

Complex research problems
- Ideal routing
- Compile into low-level configuration

Complex multi-objective optimization
Also: we are stuck with BGP

- Has kept the Internet working for decades
- But it is (almost) the same as decades ago
- Well-known technical drawbacks
  - Poor security, adoption of RPKI very slow → several prefix hijacking incidents
  - Slow convergence times → 30% of the packet loss is due to BGP
  - Policy disputes
  - No support for end-to-end circuits
  - No support for DoS attack mitigation

- It is very difficult to evolve → ossification
Routing and Shifts in the ISP industry

- Profits in pure transit drop ($/Mbps)
- Traffic increases, but so does management complexity
  - Increased load from Content Providers, CDNs
  - Do ISPs have incentives to upgrade their carrier networks for free?
  - Who should pay for the network and its management complexity?
- Bit pipe ISP model under heavy revision
- Pressure for reduced operational costs (OPEX)

→ Focus on higher-margin services
  - IPTV, VoIP, cloud-hosting (remember the “Cloud” lecture)

→ Exploration of different financial paradigms
The case for Outsourcing

- Well-known practice to reduce-streamline OPEX
  - Benefits from economy of scale
  - Ecosystem of managed networking services, e.g., IBM outsources network management to AT&T

- Outsourcing makes sense for Internet routing:
  - Internet routing and optimization is hard
  - Gets harder as the service requirements grow
  - Large effort – Small payoff
  - Complexity hinders sophisticated routing

- Idea: Routing Logic Outsourcing
Outsourcing the Routing Logic

- Focus on profitable services on top of routing
- Buy expertise from specialized contractor
- Form interactive business relationship
SDN: simpler outsourcing of per-domain routing

Step 1: Outsourcing the per-domain routing control plane
Thinking bigger: cumulative outsourcing

New multi-domain services

TE Optimizations  End-to-End QoS  Programmable Transit

Multi-Domain Routing OS Platform

Routing as a Service Contractor

Cumulative Information Exchange (SDN API)

Client A  Packet Forwarding  Client B  Packet Forwarding

Packet Forwarding

BGP

BGP

Step 2: Cumulative routing outsourcing enables new multi-domain services
Additional benefit: legacy-compatible evolution

- Ecosystem of outsourcing service contractors - clients
- New routing-signaling protocols within the clusters
- New protocols for contractor interoperability
- Legacy Compatibility (BGP)

Step 3: Legacy-compatible evolution: multiple contractors, multiple clients
Recap: Routing Outsourcing

Benefits
- Legacy-compatible inter-domain control plane evolution
- Inter-domain optimizations
- Multi-domain TE
- Economy of Scale

Challenges
- Resiliency/scalability of multi-domain routing control platform
- Evaluation of viability of routing outsourcing business model
- Incentive-based optimizations
Our research

- Develop hybrid BGP-SDN emulation framework
  - Use real router software
  - Emulate multiple Autonomous Systems (AS)
  - Emulate SDN clusters within a BGP environment

- Evaluate effect of SDN centralization on BGP convergence time as a use-case*
  - Controller speaks BGP outside of cluster (east-west)
  - Controller speaks OpenFlow within cluster (southbound)
  - Transparency to the legacy BGP world
  - Hybrid path-vector / link state inter-domain routing

→ Code available for download (open-source):
  https://bitbucket.org/gaadrian/siren

Sample Setup

Legacy BGP World

SDN AS Cluster

BGP Router ("legacy AS")

SDN Switch ("SDN AS")

Inter-AS Link

SDN Control Channel

BGP Monitor

Cluster BGP Speaker

Multi-AS Controller

8.0.1.0/29

8.0.6.0/29
Sample Results: Route fail-over on 16-node full mesh
Control Exchange Points: Providing QoS-enabled End-to-End Services via SDN-based Inter-domain Routing Orchestration

Vasileios Kotronis, Xenofontas Dimitropoulos, Rowan Klöti, Bernhard Ager, Panagiotis Georgopoulos and Stefan Schmid

Proceedings of the Research Track of the 3rd Open Networking Summit (ONS), Santa Clara, CA, USA, March 2014
Motivation

- Internet works, but it is not as reliable and performant as we would like
- Besides classic apps, like skype video calls, we expect:
  - Telemusic
  - Telesurgery
  - Remote Control of Critical Infrastructure (e.g., energy plants)

→ Question: Can today’s Internet support those services properly?
Possible for a single provider

- An ISP can allocate resources within their domain
- ISP has full overview of link utilization
- ISP controls the embedding of its traffic matrix
- Common practice = dedicated lines with guaranteed:
  - Max latency
  - Min bandwidth
  - Max jitter
  - ...
What if the endpoints have different providers?
Inter-domain routing limits us

- No end-to-end guarantees for:
  - Availability
  - Latency
  - Bandwidth
  - ...

- Current inter-domain routing does not allow this
  - BGP focuses on *reachability*, not *QoS guarantees*
  - We can’t replace BGP (easily)
We propose Control Exchange Points (CXPs)

“V. Valancius et al. “MINT: a Market for Internet Transit”, In CoNEXT ’08
ISPs announce pathlets
User requests end-to-end path
Controller stitches pathlets together
Controller monitors guarantees
Controller detects guarantee violations
Controller chooses alternative route
Best location for CXP data plane anchors?

- Good path diversity
- Maximal coverage of potential users
- Well-connected deployments
- High bandwidth and availability
- Provider neutrality
IXPs have the desired properties!

- Internet Exchange Points are public peering points
- They can have hundreds of providers participating
- They exchange up to Tbps of traffic
- They are independent of individual members
- Ideal locations for having impact on inter-domain
  - See SDX lecture later

- But what about path diversity? Coverage?
Rethinking path selection with CXPs for diversity

**Scenario 1 → valley-free routing by allowing:**
- arbitrary p2p hops between the uphill and the downhill path
- one CXP-mediated path traversed
- “mountain with wide peak”
- pure valley-free allows a narrow peak

**Scenario 2 → unlimited number of p2p links:**
- any number of CXP-mediated paths are traversed
- “multiple mountains with wide peaks”

**Other Scenarios:**
- freedom/incentives to extend current policy scheme(s)
- exploit Internet’s path diversity, under policy compliance
Recap of the CXP concept

- **CXP data plane anchor**
- **Control & monitoring**
- **Access pathlets**
- **Transit pathlets**
- **Logical e2e path**

Diagram:
- CXP controller
- Transit ISP
- Access ISP
- Source
- Destination
General directions for inter-domain SDN

- SDN testbed federation will be the first incubator
  - Common APIs (GENI*, NSI**)
  - Open mindset from administrators
  - Will learn about such testbeds in following lecture

- Need to better quantify benefits for ISP transition
  - Downtime minimization
  - Smooth migration schemes for the core
  - CAPEX and OPEX gains?

- SDX approach quite promising
  - Potential vehicle for other research ideas (CXPs)

- PCE-based path computation/installation and SDN
  - Path Computation Elements → No need to reinvent the wheel!

- We still have a long way towards standardization
  (SDNi, IETF drafts) → open design space!

* GENI: http://www.geni.net/
Any Questions