Indiana University School of Infromatics and Computing

HATS Researach Group

SDN: Large ISP Case Study Writeup

The Large ISP: An Introduction

By large ISP we mean an ISP with a physically large network, serving many external customers and providing a broad range of network and hosting services. We have in mind ISPs with national up to global reach, including the so-called 'Tier 1' ISPs.

Before considering what a Next Generation ISP might look like, we need a reasonably simple model of a current ISP where we have:

- a number of geographically separated "Points-of-Presence" (PoPs) connected by the ISP's "Core Network". The Core Network will use a variety of routing protocols-notably OSPF or IS-IS, and iBGP-which may be overlaid over MPLS and/or Layer 2 networks.
- at any given PoP, there may be:
    - "border" connections to other ISPs: peers and (except at Tier 1) transit providers-either directly or via Internet Exchange Points (IXPs). These connections all use eBGP.
    - a local Data Centre, where there may be some quantity of:
        - ISP internal services, including: internal management and monitoring systems (OSS), DNS, email servers, etc.
        - customer equipment or equipment provided for customer use.
        - local Content Distribution Network (CDN) equipment: content cacheing. (Connections to external CDNs are, essentially, peering connections.)
    - customer connections: either simple (Default Route) connections or Transit Customer connections (using eBGP). Transit customers may be other ISPs or multi-homed end-customers.

Within the PoP the Site Network will connect things locally and to the Core Network. That Site Network will use a variety of routing protocols and lower layer networks.

Not explicit in this model of an ISP are:

- network management and monitoring: the equipment for these will be distributed across the PoPs and connected by some internal network within a PoP, and across the Core Network between PoPs and to one or more Network Operation Centres. There may be some entirely separate way of reaching some PoPs, for disaster recovery.
- network services: services such as VPNs will overlay the connections and networks shown. General Internet Access is also provided over the connections and networks shown.
- the infrastructure for the PoPs: the buildings, their security, the reliable supply of electricity and cooling, etc.
- the network infrastructure between PoPs: from the fibre upwards.

The Data Centre component of the PoP will vary in size and complexity. For this component, this use case overlaps the Data Centre and the Heathrow use cases. What distinguishes this use case are:

- geography: the ISP's PoPs may be widely geographically spread, so the network between those PoPs may have significant latency and be less reliable than the network within a PoP.
- interconnection with other networks: which is a quite different from connections within a network.

SDN/NGN and Intra-ISP Networks

The working model of an SDN described above has been successfully applied to Data Centres. In a Data Centre there may be a very large number of switches/routers and an even larger number of devices in a network with disparate requirements for connecting some devices together, and for ensuring some devices remain separate. Devices in the network may include "middle boxes", such as firewalls, load-balancers, traffic-shapers and so on. SDN brings software and processing power to bear on all this complexity.

While a large Data Centre may be complicated by scale and diversity of connections, it also has properties which fit with the SDN approach:

- it is straightforward to separate the control network from the data plane. It may be possible to physically separate the control network, using separate switches and links between the ME, CE and FE. If not actually separate, the control network may be implemented as separate VLAN(s).
- the control network can be given as much bandwidth as it needs, and is physically relatively small, so throughput and latency to and between CEs should not be an issue.
- the control network is in a controlled and benign environment, so can be expected to be reliable.

which all contribute to being able to maintain the required Shared Network State. Note that we do not, here, concern ourselves with exactly how the SDN maintains this state.

Turning to the application of SDN to the ISP, clearly within a PoP the Data Centre part can follow the model above. For the ISP's Core Network, perhaps the SDN can be extended across all PoPs, as shown in Figure 6.

Here it is supposed that the Control Network that allows the ME and CE to coordinate the SDN is extended as an overlay or virtual network over the ISP core network. The issues here are that, unlike within the data centre:

- the control network is strongly dependent on the data plane.
- the bandwidth available to the control network may be limited, and the network is physically large, so throughput and latency to and between CEs may be an issue.
- the control network is in an uncontrolled environment, so cannot be relied upon.

On the other hand, it is only necessary for the extended shared network state to cover the core network, which will generally be relatively simple.

SDN for this application requires mechanisms which replace the routeing protocols which currently make forwarding decisions based on the dynamic state of the network and distribute that state around the network. The Shared Network State abstraction allows for new and better ways to manage and make those forwarding decisions, but depends on being able to keep track of changes in the network in a timely fashion and being able to maintain stability-both issues which current routeing protocols struggle with.

At present one can only speculate whether some form of SDN, along these lines or otherwise, will replace today's routed core networks. However, if so, we believe that the SDN will comprise a network of ME, CE and FE devices-what the ME and the CE will do, exactly, remains an open question.

Transitional or Hybrid ISP Networks

Assuming that ISP networks move to some form of SDN over time, there will need to be a means to incrementally replace existing routed networks. So, whatever form the SDN takes, it will need to interoperate with current routed networks. The essence of this is the separation of routing from forwarding-so that SDN parts of an ISP network speak routeing protocols in order to exchange routeing information with existing parts of the network, but make forwarding decisions and manage the data plane in their own way.

This introduces a routeing layer to the model of an SDN:

Here the Routeing Elements (RE) speak routeing protocols to each other and to routers in other parts of the network. The RE are under the management of the ME, in the same way as the CE. The RE exchange information about the network state with the CE. Note that the RE are entirely separate from the Data Plane, except to the extent that their connections to each other and, especially, to existing routers may be implemented by the data plane.

This, essentially, completes the separation of routing from forwarding that exists in current integrated routers. It also separates not only the forwarding, but also the forwarding decision making from the distribution of routeing information. The business of routeing is divided in three parts:

1. the routeing protocol, which specifies the information which is distributed across the network, and how that is achieved. This dictates what the router stores in its Routing-Information-Base (RIB).
2. the configuration of routeing policies, such that the routers which comprise a network collectively deliver what the network operator wants.
3. deciding how to forward packets given the contents of the RIB and the routeing policy-that is to say, deciding what to store in the Forwarding Information Base (FIB).

Separating routeing from routers makes it possible to then decompose the business of routeing, so that:

- the configuration of policy can be centralised, so that the operator can configure their network, not a collection of routers.
- more software and greater computing resources can be applied to making forwarding decisions: to improve traffic engineering, pre-calculate fail-over paths, improve network utilisation, and so on.

This transitional organisation applies equally to intra-ISP and inter-ISP routeing. For inter-ISP routeing the RE would speak eBGP to peers and transit providers. The RE would speak iBGP to existing routers and/or route reflectors. The opening up of the business of routeing opens up the possibility of replacing iBGP with something less prone to spending tens of seconds or more "hunting" for new paths when things change.

There are a number of incentives for ISPs to move in this direction:

- improvements in network management and configuration-reducing cost and providing greater control and flexibility.
- with a centralised view of the network and central management of its configuration, it should be possible to model network behaviour, and check configuration changes before they are applied to the network.
- opening the market for new suppliers of separate forwarding devices.
- innovation in network management, routeing and control software.

SDN/NGN and Inter-ISP Networks

Thus far we have considered only the ISP's own networks, within and between its PoPs. Now we consider how inter-ISP-inter-AS, peering and transit connections-might change in an SDN/NGN world.

Currently, the essence of Inter-ISP networking is BGP. BGP is a simple protocol, carrying a relatively small amount of information about a relatively large number of routes. What makes BGP complicated is firstly the scale of the task, distributing routes for and across the entire Internet, and secondly all the policy bells and whistles intended to allow the ISP to manage their connections to the rest of the Internet.

The wonder of BGP is not so much that it is far from perfect, but that it works at all. Amongst the issues with BGP are:

1. the speed at which the BGP mesh can respond to changes is deliberately damped in order to maintain stability. This can lead to some routeing losses measured in (small numbers of) minutes.
2. for eBGP, particularly, there is a strong, implicit link between the BGP session and the routes advertised in that session-the control and the data planes are bound together. This is because most eBGP sessions run over a point to point connection between a router in one AS and its correspondent in another. The routes advertised in the session will naturally have a next-hop which uses the same point to point connection. For resilience two ASes may establish two separate interconnections, with two separate eBGP sessions. The failure of one connection causes a ripple at the BGP level, where it would be preferable to manage the recovery at a lower, faster level.
3. the strictly limited support for traffic engineering, particularly beyond the network's borders. This is partly to do with limitations in the protocol, but a lot to do with the independence of every AS and the implicit "best-efforts only" nature of the wider Internet.
4. BGP carries only information about reachability. It carries no information about capacity.
5. the absence of any means to verify that a route arriving via BGP is kosher-saving the presence of BGPSEC.
6. the absence of any means to detect "route leaks", BGPSEC notwithstanding.

None of these issues are directly related to the SDN/NGN separation of control plane from data plane. Mostly these are deep issues with the structure of the Internet. And the Internet is of a size that any change will take time. So, it is hard to imagine that there will be a swift resolution.

However, the separation of routeing from routers offers the best opportunity yet to extend or replace BGP, starting, perhaps with iBGP.

Operators and Stakeholders

In general terms we expect an NGN ISP to have a network comprising a Data Plane, made up of a number of Forwarding Elements, under the control of a Control Plane made up of a number of Control Elements (which directly control Forwarding Elements) and a number of Management Elements (which manage the network) and (possibly) a number of Routeing Elements (connecting to existing routed networks). It seems likely that the Forwarding Elements will be distinct devices, while the functions of the Control Plane may be combined and implemented as integrated or separate software systems and applications, spread across some number of actual devices.

Without attempting, at this early stage in the development of SDN/NGN ISPs, to define how each of the elements will, eventually, work, we can identify a general structure comprising:

- control devices: spread across the ISP's PoPs

- some network connecting the control devices-the "control network"
- forwarding devices
- some network connecting forwarding devices to their controllers-the "command network"

It seems likely that each forwarding device will be under the control of a small number of controllers (for resilience) and that those controllers will be local (within the PoP).

A "pure" forwarding device might be defined as one which does exactly as it is told by a single (or replicated) controller. A device which integrates different sets of commands from different controllers may be deemed to be a hybrid (low level) controller and forwarding device. A hybrid device would be (in our terms) connected to the "control network".

The NGN ISP may be expected to have a core network and various other networks interconnected by that core network-as now-implemented by some hierarchy of control and forwarding layers.

So, secure the control of the NGN ISP it is necessary to:

- secure the various elements: so that each one cannot be subverted or prevented from doing their intended job.
- authenticate connections in the control and command networks: so that each element only talks and listens to the elements it should talk and listen to.
- protect the connections in the control and command networks: so that data is not lost, modified, added or delayed. the data exchanged is not strictly secret, but it would do no harm to:
- encrypt connections in the control and command networks.

To ensure that the control of the NGN ISP is reliable requires redundancy of elements and network.

Implementing the control and command networks as physically separate networks has obvious advantages. Implementing them as separate virtual networks, with some priority to ensure the availability of bandwidth, is the next best thing. Where the control network extends between ISP PoPs, a virtual network layer is the best option.

Assuming that control of the NGN ISP is insulated from outside interference, we may worry about whether inside interference can be detected and deal with-for example misconfiguration arising from human error or from malice. Some consistency checking in the control plane is required to check that what the network is being told to do is valid and correct. By valid we mean that it is consistent with the network topology and capabilities. By correct we mean that it is consistent with what the operator wants the network to do. So, it is valid to tell the network to do something it can do, but it is only correct to do so if the network then does something the operator wants.

Checking for validity and correctness requires a specification of the network topology and policies which can be checked against. One of the advantages of an NGN ISP network is that the control plane may be driven by just such high level specifications. Checking lower level configuration against the higher level specification could detect errors in the generation of that lower level configuration. Closing the loop and checking the actual behaviour of the network against the high level specifications may detect errors at any point in the process of telling the network what to do, and could detect interference which all other measures has failed to detect.

There are a number of parties involved:

- the network operator, including:
    - the operator's own NOC.
    - subcontractors.

- the network's customers, who may have:
    - virtual private networks.
    - virtual private data centres.

  over which they may wish to have (at least virtual) control.
- the network's peers and transit providers: where eBGP (or some future replacement) provides an arm's length connection between the networks, but which carries information which the ISP control plane must be able to depend on for its external routes. With BGPSEC there are other parties involved, providing the data-bases which contain keys used to authenticate BGP messages and which attest to an AS's right to originate a set of prefixes.

For the large ISP its subcontractors may be a particular concern. When equipment is replaced or newly installed, it must be attached to the control network, started up and brought under the control of the ISP's NOC. If a third, fourth, fifth... party is doing the work on site in some remote PoP then the ISP's NOC needs some reliable way of remotely installing the keys required for the new device to authenticate itself to others, and of configuring it so that it will not disrupt the control network the moment it is connected. In a remote PoP the ISP may wish to consider the possibility of extra equipment being placed in their network, and of extra software and configuration being added to devices as they are installed.

For virtual private networks and data centres, the question is how deep into the control network the customer needs to be admitted. One approach may be to allow the customer access to their own high level specification of their virtual infrastructure, and nothing more. The ISP's management systems can then translate and check that specification, before allowing it to be reflected in the actual network. More complicated is the creation of vertical or horizontal slices of the ISP's infrastructure, and allowing the customer to reach in and manage those-clearly this requires some means to ensure that the customer can affect only their slice, and that what they do with that slice does not exceed capacity or other limits on the service provided.