The AOAC has tasked the NTAC to assess a proposed Internet2 layer-2 service offering. This document, having its origin in the NTAC Layer-2 Services Group, is the response from the NTAC on that assessment.

The potential service offerings have been specified only informally by the AOAC, but they have provided a list of desirable technical features:

1. Access to the service should be provided through **very** inexpensive 10GbE ports,
2. The service should be built on very dense and inexpensive “throwaway” switches at the PoP,
3. Support for OpenFlow should be investigated,
4. VLAN configuration should be user-controllable,
5. The service needs to have a priority queueing mechanism, even though almost all of the traffic is expected to be best-effort and
6. The service needs a less-than-best-effort scavenger option.

These features are being viewed as stronger than a wish list, but weaker than full requirements. We do note that this group has been tasked to consider the service only in the context of the Internet2 backbone. Inter-domain aspects (whether to connectors or peers) are outside the strict scope of the charge, but nevertheless must be considered.

First we will present potential usage cases, then we will provide our interpretation of the desirable technical features. We will then discuss how these might be melded into a service offering.

Usage Cases

Two candidate uses, which could legitimately be categorized as lying at opposite ends of a continuum, for this new service are: 1) the establishment of longer-term point-to- point paths (or possibly broadcast domains) which may be used for production services (for example, a connection between two RONs for backup peering) and 2) the configuration of traffic-engineered paths for high-bandwidth flows between two or more end hosts (or perhaps switches—the real differentiator between these two cases might be that the former could be categorized as “network-based” usage whereas the latter is “host-based” usage).

As a reminder, Internet2 already offers a layer-2 service—ION. The question of how ION might be folded into this new service is, strictly speaking, outside the charge to this group, but merits consideration.

In the first usage case we would expect that connections would be requested by persons actively engaged in network engineering and operations, not by end users

or applications. The configuration itself would, at least on a conceptual basis, be done manually (that is to say, not on-demand by a user application). Because we are dealing with connections that are likely to be in the production path (but not necessarily—an ongoing research project could be a candidate for this service), the expected lifetime would be on the order of weeks, if not months or years. They would be, in general, point-to-point connections (specifically Ethernet) between a pair of routers or switches at the campus or regional network edge (or DMZs involving a small number of such devices) and could have either guaranteed or best-effort bandwidth (a backup path between two RONs could be an example of the latter).

In the second usage case the expectation is that connections would be initiated by end users (such as researchers) or even under application control, not by network engineers. Since these connections are designed to facilitate the transport of high-bandwidth flows, they would typically be short-lived (on the order of minutes, hours or perhaps days). For these connections, the availability of the requested bandwidth between the endpoints is the important component. Having these connections terminate within site boundaries implies that there are significant inter-domain considerations.

Interpretation of Requirements

The list of desirable features divides naturally into pairs. First we will examine OpenFlow support and user-configurable VLANs, followed by inexpensive ports and "throwaway" switches and finally priority queuing and a scavenger service.
A commonality of OpenFlow and VLANs is that they are both mechanisms for creating virtualized paths across a network. Our interpretation of the AOAC feature list is that it is the virtualization that is important, not the specific implementation of that virtualization. Either would be a viable implementation—VLANs have the advantage of ubiquity; OpenFlow of a common API across platforms.

We will interpret "user-controllable" VLANs in the more generic manner of user-controllable paths, be they OpenFlow- or VLAN-based. Particularly for short-term paths, there is the need for an interface, preferably web-based, usable by "civilians" and an API usable by their applications to configure endpoints and bandwidth characteristics.

The desire for very inexpensive 10 GbE ports on dense, inexpensive switches implies that the layer-2 service not be implemented directly on the core network infrastructure, but as some type of overlay or orthogonal network. "Inexpensive" is from the viewpoint of the backbone—we do not address issues (particularly backhaul) relating to access to the layer-2 service, but recognize that they can be significant.
The need for some form of priority queueing or quality of service is almost mandatory. We would envision it being used primarily for paths with committed

bandwidth. We are somewhat skeptical of the demand for a layer-2 scavenger service, but see no reason why it could not be implemented as easily as committed bandwidth.

Analysis of Requirements

We will now offer analysis of the AOAC feature list and our interpretations to show a layer-2 service to be a potentially viable offering. We will maintain the pairwise treatment, but reorder the points because of conceptual (or real) dependencies. We will first examine OpenFlow- and VLAN-based implementations of a layer-2 service. Both have the capability of provisioning virtualized network paths between (or among) endpoints. Without, we believe, any loss of generality, we will refer only to point-to-point paths. We will next look at the requirements for user-controllable paths and implications for ports and switches. Finally, we will examine the QoS requirements.

Both OpenFlow-based paths and VLANs offer sufficient capabilities to implement the usage cases outlined above. From an end-user perspective (either a network engineer or a researcher) they are both implemented over Ethernet ports and can offer the same internal structure on that port. Because OpenFlow can base forwarding decisions on the matching of 12-tuples (including a VLAN tag), it *prima facie* offers broader network virtualization capabilities than does a VLAN-based service. Unlike VLANs, which have vendor-specific configuration mechanisms, OpenFlow offers a well-defined API for provisioning switch forwarding tables, removing the need for an additional layer of configuration abstraction in heterogeneous systems. VLANs are ubiquitous; OpenFlow appears to have wide and increasing vendor support. We have a working consensus that, if it can be shown through thorough testing that the stated capabilities of OpenFlow can be realized in a production environment, an implementation based on OpenFlow would be preferable to one based on VLANs. In the following discussion we will assume, where necessary, the service to be OpenFlow-based.

User-controllable paths imply some form of user interface for effecting that control. Because the service is highly distributed and may be configured by end users, a web- based tool is, if not an absolute requirement, a very strong preference, particularly given the desire that the service be easy to use. We have two such tools in use in our community: NLR's Sherpa and Internet2's OSCARS suite. To quote from the NLR website, "Sherpa provides guided, secure, interactive [web-based] dynamic circuit configuration. It allows authorized users to provision, modify, enable, and disable dedicated or non-dedicated VLANs on FrameNet [NLR's layer-2 service] in realtime, without requiring intervention from the NLR NOC". The OSCARS suite provides similar functionality for Internet2's current ION dynamic circuit service. Either could serve as a basis for the control software for this new layer-2 service; the ease of incorporating OpenFlow support might be the gating factor for implementors.

Given that a long-term path, as described above, might cross NREN boundaries, it is clear that some degree of inter-domain coordination may be required, but it is difficult to conclude that fully-functional inter-domain controllers are strictly necessary when dealing with longer-term paths, particularly because the points of demarcation will generally be backbone switches or routers. A pertinent question is whether the "users" of the control interface are "real" end users (or their campuses or RONs) or personnel at the NOC. There seems to be little need for an end user to be able to create dynamically a potentially permanent circuit; in fact, connectors might desire this capability not to exist (or the ability to create a long-term path might be restricted by specific access controls). We would not expect that involving the NOC in the management of long-term connections would be particularly onerous to either the NOC or an end user, but there are clear advantages in the use of common administrative tools.

However, the "short-term path" usage case has philosophical differences from the case involving potentially permanent paths. Having end hosts in different administrative domains necessitates the use of inter-domain controllers—the alternative, having connections coordinated by the NOC, is not a realistic expectation when dealing with dynamic connections. Giving end users and applications access to the control plane implies some degree of access control—trust relationships need to be established. AAA is not a component of OpenFlow—but it is commonly implemented in OpenFlow controllers. That AAA mechanism could be the one used in OSCARS for ION or it could be the one in Sherpa—the specification is an implementation detail, but one of critical importance.

There are on the market a number of very inexpensive switches which have moderately high densities of SFP+ ports, supporting both 1 GbE and 10 GbE. Some of them even have QSFP+ ports for 40 GbE access or uplinks. Some have OpenFlow support out of the box; others have open-source operating systems to which OpenFlow could be added. Thus, potentially appropriate candidate switches do exist. However, these switches tend not to have a great deal of redundancy, a point which would need to be understood when using the service. Any given switch would require careful evaluation before being adopted, but a requirement for SFP+ ports is very compelling because of the ability to support both GbE and 10 GbE connections in the same port.

We envision that this service will be implemented by connecting the switches directly to lambdas (10 Gb/s or perhaps even 40 Gb/s) on the layer-1 network, creating an infrastructure distinct from the IP-based production network (although it may be more cost effective to provision the service over Ethernets or MPLS tunnels—this is an implementation decision subject to oversight by Internet2 staff). Access to the service could be either over existing connections to the Internet2 network (assuming that layer-2 connections are possible between the T1600s and these switches) or over new connections directly to the layer-2 switches. We expect this to be driven, in

large part, by connector economics.

Even though the expectation is that most traffic on the service will be best-effort, the need for some form of quality of service is apparent—be it for a scavenger service or a point-to-point connection with dedicated bandwidth. In most current implementations, OpenFlow cannot be used to provision QoS capabilities. Thus, it is necessary to rely on the QoS functionality of the substrate. Assuming the service is implemented as an overlay on Ethernet, QoS would likely be controlled by using appropriate values of the Ethernet priority code point (informally referred to as 802.1p). Because we are dealing purely with layer-2, there is no obvious need for hierarchical quality of service.

However, particularly in host-to-host usage, there may be a need to support Ethernet- based QoS in a heterogeneous environment.

Discussion

One item which was not included in the charge from AOAC deals with economic aspects of the service, which are over and above the costs of connections to access the service. It is likely that, at some point, Internet2 would adopt a charging model for the service. Campus and RON networking groups would expect there to be charges for the use of the service, particularly for permanent connections with dedicated bandwidth. However, researchers, who, in general, do not include funding for network connectivity in their grants, might be able to use the service only if any charges were not passed along to them. Further analysis is outside this group's scope, but it does require consideration.

Our overall conclusion is that an Internet2 layer-2 service is feasible with few, if any, anticipated technical obstacles. The service can be built on either OpenFlow or VLANs; we recommend the former because of the additional flexibility it affords. GUI tools exist which can serve as candidates for tools to configure the service. The network can be built as a discrete network using switches which have high port density and a price point which is considered in the community to be attractive. Sufficient controls exist to implement the network with degrees of security and quality of service to accommodate both administrators and users. Access controls and accounting will be difficult to implement, particularly in inter-domain scenarios, but that is not a strong argument against deploying a layer-2 service.

Paper prepared by Michael Lambert with the NTAC Layer-2 subcommittee. Subcommittee members: Cort Buffington, Linda Winkler, Matt Davy, Matt Valenzisi, Michael Lambert and Cas D'Angelo. Also, the following Internet2 staff participated: Chris Robb, Dale Finkelson, Eric Boyd, and Matt Zekauskas.