Role of Open Exchanges in the evolution of global research and education networking

"Open Networks for Open Science"

19 May 2011

At the kick off meeting of the GLIF Open Lightpath Exchanges ad hoc Policy Group it was agreed amongst the attendees that we needed to draft a problem statement of the importance and necessity for GOLEs. At the meeting several attendees asked what is the problem that we are trying to solve? Why do we need a policy or governance working group on GOLEs?

This document tries to briefly address these issues and layout the questions that needs to be addressed in terms of GLIF Open Lightpath Exchanges.

Some History

Prior to the development of the Internet, telecommunication networks such as the telephone system were always hierarchical. Municipal telephone networks connected to regional networks which in turn connected to national networks. This was done in order to aggregate traffic to efficiently utilize scarce and expensive long distance or trans-oceanic telecom infrastructure. However, while the utility of voice communication became indispensible to society the phone network rapidly became a powerful monopoly blocking innovative use and novel services. I

The Internet changed all that. Since the "Internet" was *specifically* created to provide and enhance "Inter-network" communication, it required a new architectural paradigm to allow for the exchange of traffic between networks. And in order to provide robustness, the management and control was purposely decentralized and made autonomous in both its control as well as its policy – a dramatic new approach to telecommunications .

As the Internet developed, routing protocols such as Border Gateway Protocol (BGP) were developed to enable and to automate the ability of networks to interconnect as they deemed necessary amongst themselves. There was no "design" that forced or necessitated a particular network interconnection scheme in the internet. Thus, a rich and very flexible environment was created that allowed each emerging region or industrial sector to construct the network infrastructure and inter-networking relationships that served its constituents' purposes. ***A hierarchical network was no longer***

intermediating who could establish interconnections, or how they had to do it, or why or why not.
The resulting amorphous collection of networks was able to find a natural balance between cost effective hierarchical aggregated networks infrastructure was expensive, and flat direct peerings with dis-intermediated connections where groups of like-minded networks could establish common self-defined services and policies.

In practice, realizing this powerful, open interconnection environment was expensive – each network had to acquire circuit-based connections to every other network it wished to peer directly with. Indeed, the expense of doing so inhibited such interconnections for much of the early internet.

The cost of interconnecting Internet Service Providers for the purposes of peering drove the creation of "exchanges"; a few centralized locations were created where networks could establish a point of presence. Each network at an exchange could easily and cheaply establish many direct peering connections to other networks present. This proved to be a cost effective solution that provided bi-lateral peering in a rich "town square" paradigm. Examples of these early exchange points were the Metropolitan Area Exchanges – "MAE-East" in the Washington DC region, and "MAE-West" on the US west coast. Such exchanges continue to thrive today in much more sophisticated services offered by facilities such as AMS-IX, NYIX, "telco hotels", and corporations such as Equinix.

This open exchange model provided substantially greater control to each network to form alliances and to develop innovative new distinguishing and differentiated services. Exchange points enable a) bilateral peering based upon only the policies and priorities of the two peering networks, b) they allow interconnections using many different types of technologies and protocols, c) they provide a cost-effective means to realize the broad flat peering fan-out of the internet that provides both robustness and enhances innovation.

Internet exchange points enable or enhance emerging new service architectures such as content distribution networks and "clouds" services (e.g. Amazon, Google, Skype, etc), allowing application service providers to peer directly with ISPs including National R&E networks (NRENs) avoiding the high costs of transit and policy constraints imposed by the traditional carriers. They significantly improve the "user experience" by reducing hop counts and by providing more direct performance oriented paths between content providers and consumers.

Current R&E network environment

Although the Internet has radically changed network topology from the old hierarchical telephone system, the global R&E network community remains highly hierarchical. Virtually every R&E network serves a province/state or nation, which in turn connects to a national (NREN) or pan national backbone. The national or pan-national backbones often then make arrangements for complex interconnection agreements between each other and often discourage provincial/state networks from interconnecting directly .

It should be noted that although national or provincial/state R&E networks are almost exclusively monopolies, this is generally done on a voluntary basis and the arrangement does not imply a

government imposed monopoly or a breakdown in open markets. The driver for these voluntary monopolies have been economy of scale for the R&E community, as well as serving state or national science. However, these networks and their organizations did not evolve as the cost structure they were created to defend against has come down. Indeed, the price/performance ratio at all layers of the telecommunications technologies and protocols has dropped substantially over the last decade for most of the R&E constituency. Yet our network organizations in general remain tied to service models established in the last century. These same [policy/financial] strains have now started to develop between and within R&E networks that still subscribe to a hierarchical business aggregation model.[1]

These strains have evolved over the past several years where many NREN and provincial/state networks have established direct connections with each other through a number of mechanisms including "cross border" fiber connections, wave based services, and even Ethernet VLANS or traditional SONET/SDH circuits. And we can expect the trend toward flatter collaborative interconnections to accelerate as virtualized IT infrastructure and services mature.

These "bypass" connections were established for any number of reasons:

(a) To bypass policy limitations of national or pan-national hierarchical networks;
(b) To reduce costs by bypassing more expensive conventional networks;
(c) To seamlessly connect distributed campuses scattered across multiple networks;
(d) To directly connect across traditional boundaries sites belonging to a pan national science project;
(e) To reduce latency, hop count and path complexity by routing traffic directly between adjoining networks, rather than going through intermediate transit networks.
(f) To quickly deploy innovative new products or services such as lightpath services which would take much longer to implement on traditional hierarchical national or pan-national networks due to their larger more general purpose community and more conservative risk aversion; and
(g) Greater range of innovative, experimental activities as like-minded networks and scientists can experiment with new protocols or services without waiting for universal consensus on a commonly agreed solution of all participating networks in the traditional hierarchical structure of R&E networks.

Cross border fiber connections have been widely established between NRENs in Europe and provincial/state networks in North America. A growing portion of traffic between adjoining European NRENs and regional networks in North America now pass directly between these networks rather than through the hierarchical pan-regional or pan-national networks. As wave and/or layer2 based services proliferate, the same issues exposed by cross-border fiber will emerge as these virtual services allow

[1] Indeed, at one early point in the R&E network development, client policy regarding university membership fees for campuses of a single university system were evaluated according to how many sports teams the campuses had: One team = one fee, two teams = two fees. This type of policy consideration is perhaps done in a more realistic fashion today, but nevertheless demonstrates how policy can inhibit collaboration in sometimes completely baffling ways.

network interactions that do not conform to strict hierarchical models. In this paper, where we refer to cross border fiber, we generally intend it to mean any type of connection/lightpath capability that can flatten or otherwise bypass traditional topologically based policy.

While cross border fiber helped reduce costs and spur on innovation it is not a complete solution. One of the challenges with cross border fiber is developing end to end solutions where traffic must transit multiple intermediate regions and the associated regional provider(s). The need to transit intermediate networks has been a problem that has bedeviled R&E as well as commercial networks for some time. The problem is even more challenging when one needs to forward new service capabilities (e.g. lightpaths) not necessarily supported by intermediate networks. Differing policies, transport technologies, costs and availability of capacity make this a very challenging problem to solve. It is fraught not only with technical challenges but business issues as well. A major focus of current research in lightpath networking is related to the problem of inter-domain lightpath provisioning and related network management.

Transit networks also pose a political and policy problem for governments at multiple levels. A new generation of research intensive economies is developing in China, Brazil, India, South Africa, etc. who will not be satisfied with data transit policies and services unilaterally decided by Europe or America.

Open Exchanges can defuse this situation. Similar to internet exchanges for IP traffic, Open Exchanges allow all networks to establish a presence at a common Exchange point and establish bilateral peerings for the interchange of traffic. Interconnecting at Open Exchange points means that all networks are seen and treated as equals. Neither party is subject to transit policies of an intermediate network. As more and more science and research involves global collaboration we need open architectures that enable this type of science without any single network or country being able to impose policy or traffic restrictions on another country or collaborator.

This fundamentally removes the issue of non-aligned intermediate transit policy. However, one subtle requirement must be made explicit for Exchange Points to function as policy neutral interconnect facilities: These facilities must be engineered to be *non-blocking*. Non-blocking means that there are no technical or performance limitations imposed by the Exchange point itself that could create contention for the Exchange point resources. If there is contention for Exchange Point resources among the clients, a policy is required governing which clients will get the resource. This places some networks at an advantage and others at a disadvantage. This undermines the purpose and effectiveness of the global Exchange Point architecture. The Exchange Point must therefore provide a carefully defined service specification that offers a level playing field for all clients – whether the Exchange point offers physical fiber cross connects, circuit switching, or packet based VPN services. A client's ability to establish bilateral inter-network connections is solely a function of the clients' available resources at the edge of the Exchange facility.

Implicit in this level and non-blocking requirement of Exchange Points is that it applies to *access* to the Exchange point as well. Any network or service provider should be able to gain access to the Exchange Point for the purposes of making use of the Exchange point services. Carriers or regional/national

service providers must all have equal and open access to the facility. Without such open access, the constraints of allocation or usage policy will limit flexibility and innovation.[2]

Unfortunately for emerging R&E services, the richness of Exchange points is not yet realized at the scope or reach that establishes "self-determination" and "freedom of association" as a fundamental tenet of future network Exchange architectures. Indeed, few R&E networks can afford direct access to all major international Exchange points. Therefore, transit services between Exchange points are needed to allow client networks at one Exchange point to reach client networks attached to another Exchange point.

Such inter-exchange transit services have the potential to reincarnate the very problem Exchange points are supposed to solve: constraining transit policy. Ideally, we would like to develop an architecture and service model that provides inter-exchange transit services that extend and expand the non-intermediary policy of an Open Exchange. However, such an extension of Exchange Point openness to a "distributed Exchange Point" concept requires a fully provisioned bi-section bandwidth allowing non-blocking service across geographically distributed points – which could be enormously expensive for large Exchange Points separated by long distances.

This poses important scaling issues and by implication creates a more complex Exchange Point service offering – compelling the community to look at the reality of "distributed" Exchange Points with a more analytical and pragmatic realism, i.e. stitching a set of individual Open Exchange Points together to create an open fabric or globally distributed Exchange point with similar properties will be difficult to achieve both technically and financially. Open Exchanges do not create unlimited capacity or a single global network or resource governance, and so fully qualified *distributed* Exchange Points will require substantial inter-connection bandwidth or careful crafting of the Distributed Exchange service definition in order to maintain a non-preferential service profile. It should be noted here that the cost of operating a single geographically co-located switching node is substantially less (orders of magnitude less) than the cost of long fat pipes for connecting to other distant Open Exchange Points.

As arrangements must be made to provide capacity between Exchange points, the owners of that inter-exchange capacity can make that capacity available to third parties. This is a different sort of problem than transiting a network where usage policies and path determination challenges can prohibit or inhibit such transit services. This is an open topic for both research and experience to inform. Although much work needs to be done it is reasonable to expect that independent yet "open" policies can be developed for the usage and allocation of the transport resources between open Exchanges.

As well as a richness of many Exchange points, over time [commercial] Internet exchange points have established a number of tools such as route registries and route servers to help networks interconnecting at an exchange point identify the appropriate peering network with the shortest hop routes to the final destination. To date such tools are in their infancy for lightpaths.

---

[2] Note that *open* does not imply *free*; the Open Exchange point may charge a fee for connecting and use of services.

The new science and demand for big pipes

As discussed at the Internet 2 meeting, we are starting to see big science applications whose data volumes would overwhelm a traditional R&E IP network(s). In conventional shared IP service environments, these big science application are constrained by non-deterministic and bursty traffic loading present in the shared network. Further, the large science user can cause degraded performance for the conventional users by creating many large flows that steal a greater portion of the capacity from the larger user community. These science communities are looking to establishing their own networks by interconnecting the proposed Open Exchange points thus affording a particular science community the ability to connect at a conveniently "nearby" Open Exchange and thereby be able to access and share purpose built capacity connecting community sites. LHCONE is an early example of this within the R&E community, but many more are expected with deployment of global scale eVBLI, climate modeling, etc. By identifying the Open Exchange Points as the regional foci of access for all communities, and then on that basis all communities contribute capacity to the inter-exchange fabric, then such capacity becomes more easily accessible, more efficiently utilized, and can be more effectively managed. However, as stated above, the sustainable and equitable models for cost sharing and delegation of inter-exchange transport capacity is an open issue that will require substantial discussion among the community.

Initial models for such community networks were initially aligned with the hierarchical network architecture, often reflecting also (for the LHC community) the assumed hierarchical data distribution pattern. Practical experience shows that data movement and replication of data sets is only partial hierarchical and there is a much greater demand for access to any data set anywhere resulting in a more mesh type access architecture. As discussed earlier, a full mesh of interconnections between sites around the world hosting data sets is not practical. And a full bisection bandwidth provisioning of inter-exchange capacity among all Open Exchanges is similarly prohibitively expensive. However, where experience dictates, coordinated and *collaborative* build out and dynamic allocation of inter-exchange capacity will address the particularly high-demand inter-exchange routes and reflect the reality that data distribution patterns do not fall completely into either hierarchical nor full-mesh models, and that these flows will morph over time as projects and teams come online. Interconnecting at Open Exchanges allow all sources and sinks of these massive data sets to exchange data without network imposed policy constraints, and will allow the most efficient build out and utilization of international and inter-continental capacity across the entire R&E community.

This is not a new phenomenon. Internet content distribution networks (CDN) and cloud providers have discovered this same solution independently. They also have to connect massive data sets and computational resources to users. Large CDN providers like Akamai, Google, Limelight, etc. have deployed extensive private optical networks and distributed servers to transfer these large data sets through the global set of Internet Exchange points where they interconnect to last mile ISPs. LHCONE is simply a variant of a CDN network dedicated to distributing LHC data.

Network Innovation

One of the important policy objectives for the advancement of digital society by governments is the development of new network protocols and architectures. Funding councils in Europe, America and elsewhere around the world have spent millions in research in network security, reliability, and greater flexibility. One of the biggest challenges facing researchers and network operators is how to smoothly transfer these new protocols and architectures into a production environment without a major forklift to the existing production network or disruption of existing services. The deployment and adoption of IPv6 is a classic example of the significant barriers to move to a new protocol. The adoption of standard video compression and distribution protocols is another example. Deploying new protocols, transport technologies, and services on a global scale is difficult largely due to the inability of those groups of users and/or networks that require these new services to interoperate successfully with their like-minded peers across non-conforming networks. In some cases, virtual networks suffice to hide the intervening transit networks, but with in some scenarios, such virtualized tunnels are difficult to realize or hide innate capabilities of the underlying technology. Open Exchanges provide for a much greater flexibility as they allow like-minded networks to peer with each other directly using any protocol or architecture they so wish. Again we have seen these developments occur in the commercial Internet world where CDN networks such as Akamai and Google have deployed their own versions of TCP/IP to optimize the transfer of their massive data sets through Internet exchanges around the world.

The problem statement of Open Exchanges

If one accepts this basic premise of the growing importance of Open Exchanges for the advancement of science and research, we need to work on global coordination of deployment and interconnection of Open Exchanges. To date there is only a handful of Open Exchanges located around the world, many of which could easily be single points of failure. There is very little diversity and a limited number of interconnections between them. While these initial few facilities have been useful, the future strategic potential of Exchange Points generally will not be achievable without a coordinated effort both technically as well as financially. We need to develop funding and pricing models to support a richer set of Open Exchanges and a diverse set of interconnections. We need to further develop emerging Open Exchange tools and services. We need to develop a consensus framework for emerging Open Exchanges, and reference implementations from which to gather best common practice. This framework must clearly state the defining characteristics of Open Exchanges to maintain the architectural principles of openness and non-intermediation. Open Exchange policies need to assure that any party can bring in a link and access the Exchange and that any client can be connected to any other client at the Exchange and that there are no constraints on the traffic which subsequently traverses the Exchange. We need to develop the value added services to support the science and education in the use of Open Exchanges and establishing end to end connections seamlessly and transparently. New distributed data architectures, grids, science cloud services, science as a service, and even Green IT architectures will be greatly enabled by Open Exchanges – but much needs to be done to develop the necessary tools and middleware architectures.

We also need to recognize that the direct interconnection of universities and networks to Open Exchanges may undermine existing network business models that provide cost effective network solutions to smaller and more remote institutions. These institutions cannot be forgotten and funding models and solutions must be developed to address their needs; We must be highly conscious not to create a new digital divide. We need global coordination of such efforts to avoid creating solutions that do not interoperate or result in Open Exchange islands. Science communities, universities and researchers beyond those usual suspects at the bleeding edge of high performance networks (LHC, eVLBI, etc.) must be engaged on the advantage of the Open Exchange based service architectures.

This setting of the stage of the problem statement is only the first step. Much more work needs to be done. This paper attempts to clearly articulate the fundamental architectural value of Open Exchanges within advanced networks in order to build the consensus that Open Exchanges are indeed one of the most promising new “levelers” for network innovation and research in coming years.