

Layering the Inter-Domain Layer

Vytautas Valancius and Nick Feamster
College of Computing, Georgia Tech
vvalanc2@uiuc.edu, feamster@cc.gatech.edu

1. Introduction

Data transport services are flourishing at the intra-domain level. Despite that, things look much worse when we try to cross a domain boundary. Operators are endlessly tuning Border Gateway Protocol (BGP) to achieve various traffic engineering goals [4], yet, unfortunately there is no easy way to allocate resources from point A to point B between two arbitrary domains on the Internet. Both work in the research community [9] [10] and the success of a virtual network operators (VNOs) such as Vanco [7] demonstrate the need for this type of function. Additionally, although capacity is ubiquitous and often over-provisioned, it is not exposed to market forces and it is not easy to sell. This paper proposes scalable protocol mechanisms for provisioning interdomain paths for buying and selling.

Advanced interdomain scale services obviously pose some extreme challenges. The most pressing issues include:

- *Scalability.* A service for provisioning interdomain services must scale to a large number of networks and end systems.
- *Privacy and policy enforcement.* Selling and provisioning end-to-end services requires information about available resources, but ISPs are typically reluctant to share all information about their resources.
- *Security and accountability.* Interdomain services also require transit domains to be accountable for offered services.

No proposal for a new set of protocols for interdomain service delivery can move forward without addressing these issues.

In this position paper we provide two insights. First, we take a fresh look at the interdomain control layer on the Internet. Observing that interdomain control should perform three distinct functions, we describe a possible control plane separation for such functionality. Second, we propose a new layer for abstracted interdomain topology information exchange. We explain the possibility of such layer and propose a research agenda.

2. Inter-domain Control Plane Layers

To date, the sole role of interdomain protocol was subsumed by BGP. We think it is possible and necessary to separate certain interdomain control plane features to several distinct layers. That said, advanced interdomain services shall not substitute best-effort per-hop based forwarding and conventional protocols such as BGP. On the contrary, IP forwarding can successfully coexist with new services

as it can be observed in the current intra-domain networks using both MPLS-VPN and MPLS-TE technologies[6] and standard IGP protocols.

We propose an interdomain service that comprises three distinct sub-services or layers, each providing different functionality:

- **Service setup and monitoring layer.** This layer is responsible for setting up the service and monitoring the performance. This shall be supported by every node on the service path. Fortunately existing routers already support such functionality for bandwidth reservation in the form of Resource reSerVation Protocol with Inter-Domain Traffic Engineering extensions (RSVP-TE) [1].
- **Accounting and inventory layer** This layer is responsible for keeping track of available resources and participating in joint interdomain path computation. This layer resides on dedicated platforms, separate from forwarding nodes. We can see framework for such layer provided by Path Computation Element (PCE) architecture [2].
- **Inter-domain topology layer.** This layer provides abstract interdomain topology information. It does not require any changes to the current routing platforms. The nodes at this layer can provide interdomain paths to one or more client source domains. In fact, it is still an open research question whether this layer should be distributed or centralized.

Figure 1 shows the proposed layering of the interdomain protocols. The figure shows three layers and protocols at each layer. Each of the services/layers addresses different challenges. Service setup and monitoring protocols allows for authenticated service establishment with strict accountability. Inventory layer tracks the resource usage inside domain and provides privacy and policy enforcement. Interdomain topology layer, besides providing interdomain information shall address the scalability problems.

We envision the layers interacting in the following way:

1. The *Interdomain Topology Layer* computes abstract AS level interdomain route.
2. The *Accounting and Inventory Layer* communicates with nodes on ASes along the route coordinate to compute the exact interdomain path.
3. *Service Setup Layer* sets up the service by provisioning the paths.

As was briefly mentioned above, two of these three layers are already emerging. RSVP-TE protocol with interdomain

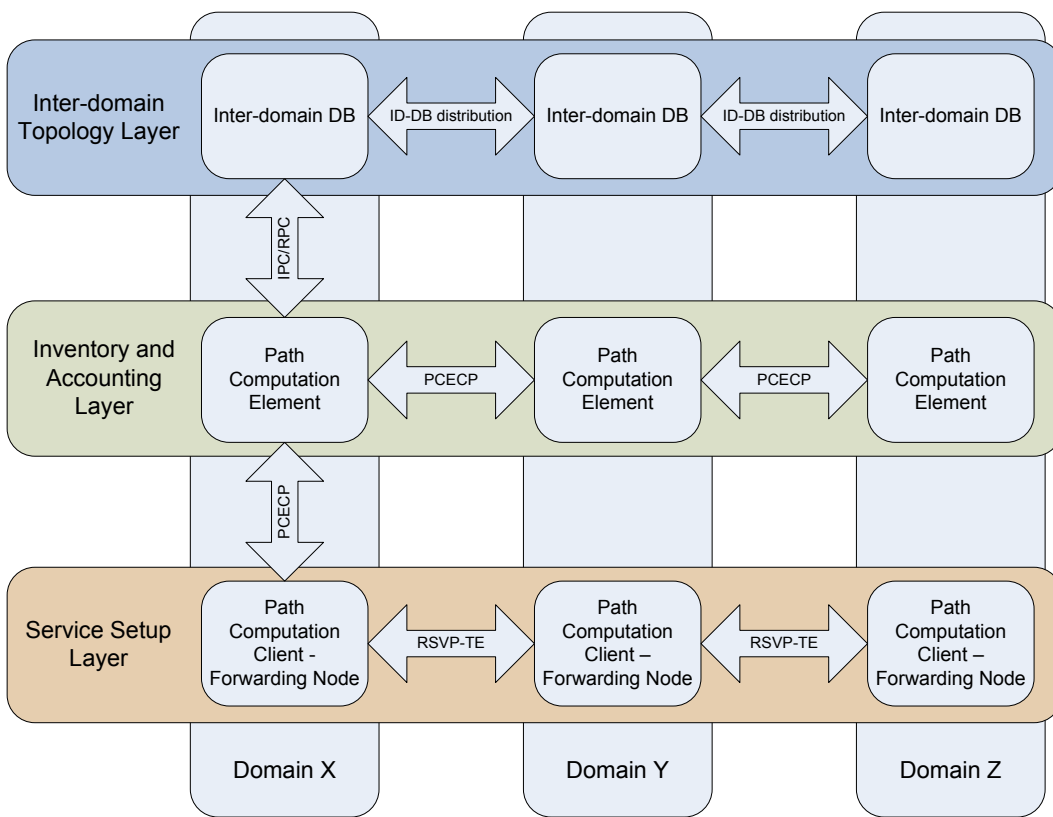


Figure 1: The Control Plane Layers.

extensions is a perfect example of service setup layer[5][1]. Inventory layer is emerging in the form of PCE architecture[2][8]. There is currently no ongoing work to develop an *interdomain topology layer*.

3. The Interdomain Topology Layer

Currently the interdomain reachability information is provided by BGP. Some initiatives [10] try to extend BGP to support greater visibility. We argue that BGP is not up to the task to provide topological resource information between domains. We propose several research directions on control plane level dedicated to interdomain resource information dissemination. We must stress again that the new layer does not compete with BGP. BGP in its present form (or more preferably in a way defined in [3]) can continue serving best-effort per-hop forwarding decisions.

Words ‘topology’ and ‘interdomain’ inevitably raise scalability concerns. The following aspects of the Interdomain Topology Layer will improve scalability.

1. The interdomain topology layer needs to provide only abstract information—treating ASes as black boxes and advertising only ingress and egress tuples with resource constraints.
2. The new layer does not need support routing policies similarly to BGP. Most policy decisions, such as allow or disallow certain traffic to pass a domain, are made at the inventory layer using a special protocol set (ie.

protocols provided by PCE architecture).

3. The interdomain topology database provided by the new layer need not be up to date. This is the most fundamental assumption that allows us great flexibility in design. We believe that this assumption is valid because any discrepancy in the database is compensated by the inventory layer.
4. The interdomain database provided by the new layer need not be completely accurate. Providers can advertise the available bandwidth, minimum reservable bandwidth, link protection levels and other information as they see fit. It does not need to reflect the actual link capabilities. If link is under-provisioned, the inventory layer, that coordinates calculation of exact path will notice the problem and proceed with a different path. If some transit domains advertise configuration that is not reliable, source domains might mark them as untrustworthy.
5. The new layer does not need to react fast to the data-path failures. The data-path problems could be detected at the setup layer, possibly using protocols such as RSVP.
6. The new layer does not need to exchange information about prefix ownership, since this information is already available through BGP.

Having all these assumptions, requirement list is short:

1. The protocol must scale.

2. The protocol must expose service-level information.
3. The resource information must not reveal private or proprietary details about a domain's internal topology.

4. Research Directions

The observations outlined above offer many directions for future research. We need to approach the design of a new layer from both theoretical and practical perspectives. The resulting design must be both scalable and business friendly. Specifically, we must consider the following points:

- What kind of information should be exposed to inter-domain layer?
- How often should this information be updated?
- What kind of architecture the protocol should use?
 - How would distributed link-state protocol work in the interdomain setting? (Link-state would become domain-state.)
 - What aspects of the protocol should be distributed versus centralized? Should the protocol make use of other existing protocols (e.g., DNS)?

5. Conclusion

This position paper has proposed a fresh look at the layering of interdomain protocols and a new interdomain topology information layer. We observe that there is a need for a layered interdomain information dissemination service that is distinct from BGP in the following ways: (1) the new service needs to deliver more visibility than BGP does, and (2) the new service does not need to have convergence properties similar to BGP, because other layers can take care of inaccurate or outdated information. We believe these propositions offer many design options and research directions for a new protocol that will enable advanced interdomain services and applications.

REFERENCES

- [1] BONAVENTURE, O., FILSFILS, C., AND FRANCOIS, P. Achieving sub-50 milliseconds recovery upon bgp peering link failures. In *CoNEXT'05: Proceedings of the 2005 ACM conference on Emerging network experiment and technology* (New York, NY, USA, 2005), ACM Press, pp. 31–42.
- [2] FARREL, A., VASSEUR, J.-P., AND ASH, J. *A Path Computation Element (PCE)-Based Architecture, RFC4655*. IETF, Network Working Group, 2006.
- [3] FEAMSTER, N., BALAKRISHNAN, H., REXFORD, J., SHAIKH, A., AND VAN DER MERWE, J. The case for separating routing from routers. In *FDNA '04: Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture* (New York, NY, USA, 2004), ACM Press, pp. 5–12.
- [4] FEAMSTER, N., BORKENHAGEN, J., AND REXFORD, J. Guidelines for interdomain traffic engineering. *SIGCOMM Comput. Commun. Rev.* 33, 5 (2003), 19–30.
- [5] KARSTEN, M. Collected experience from implementing rsvp. *IEEE/ACM Trans. Netw.* 14, 4 (2006), 767–778.
- [6] OSBORNE, E., AND SIMHA, A. *Traffic Engineering with MPLS*. Cisco Press, 2002.
- [7] PLC., V. Vanco corporate website. <http://www.vanco.com>.
- [8] VASSEUR, J., AND LE-ROUX, J. *Path Computation Element (PCE) communication Protocol (PCEP), draft-ietf-pce-pcep-07.txt*. IETF, Network Working Group, 2007.

- [9] XIAO, L., LUI, K.-S., WANG, J., AND NAHRSTEDT, K. Qos extension to bgp. In *ICNP (2002)*, pp. 100–109.
- [10] XU, W., AND REXFORD, J. Miro: multi-path interdomain routing. In *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications* (New York, NY, USA, 2006), ACM Press, pp. 171–182.