VEIL-Click: Creating Advanced Large-scale Ethernet Networks

Sourabh Jain, Gowri Chitloor Parthasarathy, Zhi-Li Zhang
{sourj,gowri,zhzhang}@cs.umn.edu
University of Minnesota-Twin Cities

***Abstract*—*VEIL* (Virtual Ethernet Identifier Layer) is a novel concept for below IP networking. It addresses the scalability, efficiency and reliability challenges facing the traditional Ethernet, while retaining its simplicity. Furthermore, it provides a uniform (below IP) *convergence* layer to support a large, dynamic *and heterogeneous* (layer-2) network that is capable of connecting hundreds of thousands or more *diverse* physical devices. The key idea in the design of VEIL is to introduce a *topology-aware, structured* virtual id (*vid*) space onto which both physical identifiers as well as higher layer addresses/names are mapped. VEIL completely eliminates *network-wide flooding* in both the *data* and *control* planes, and thus is highly scalable and robust. In this poster, we present a Click Modular Router based implementation of VEIL, which constructs a scalable "plug-&-play" layer-2 network which has many advanced features such as *mutlipath-routing*, *"flooding-free"* address resolution and seamless support for *mobility* of end hosts connecting to the network. Finally, the current implementation runs on top of real hardware, which can be deployed with several other traditional Ethernet switches, or as a complete network of VEIL switches alone. Therefore, it can be readily deployed to create feature-rich, large-scale Ethernet networks for various needs.**

I. Motivation

The explosive growth of the Internet has enabled a wide range of diverse devices to interconnect and communicate with each other through a variety of disparate technologies. While serving as the universal "glue" that pieces together various heterogeneous physical networks, the Internet Protocol (IP) suffers from certain well-known shortcomings. E.g., need for careful and extensive network configurations, relatively poor support for mobility and so forth. In contrast, layer-2 technologies such as Ethernet are largely "plug-&-play" where hosts are equipped with persistent MAC addresses and Ethernet switches automatically learn about host addresses and location, adapt to changes in network topology as well as host mobility and perform packet forwarding seamlessly with minimal operator configuration and intervention. Because of this simple "plug-&-play" semantics, today's *switched* Ethernet technology has led to rapidly expanding *large* and *dynamic* networks such as large data centers and Metro Ethernet, with up to tens of thousands switches and millions of hosts.

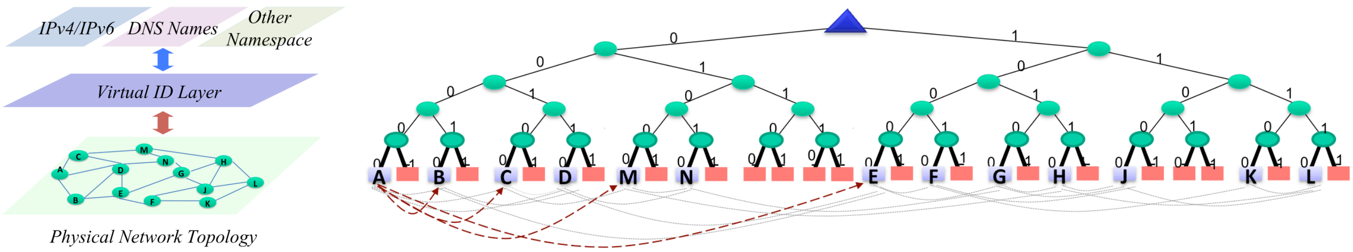
On the other hand, the unprecedented scale as well as the demanding efficiency and robustness requirements of these new large and dynamic (layer-2) networks also pose revolutionary challenges to the Ethernet technology that was originally intended for use in small local area networks. For instance, the *network-wide flooding* often resorted to by Ethernet switches to locate end hosts and forward packets whose locations are not learned, significantly reduces network capacity. The spanning tree algorithm used to avoid forwarding loops not only results in sub-optimal forwarding paths, but is also slow in adapting to changes in the network topology. To address these challenges, several solutions [3]–[6] have been proposed, of which, SEATTLE [3] is closest in spirit to our work, in that, both utilize DHT (distributed hash table) techniques for scalable and efficient address look-up and resolution. However, SEATTLE employs the OSPF-style shortest path routing in layer 2. It therefore not only requires network-wide flooding in the control plane for building routing tables, but also suffers from the same scalability and robustness limitations plaguing shortest-path routing.

II. Overview of VEIL

We present *VEIL*—a novel *Virtual (Ethernet) Identifier (Id) Layer* for creating large scale "layer-2" networks with several advanced features such as multi-path routing, seamless mobility support etc. The key objective is two-fold: i) Address the challenges facing the traditional Ethernet, while retaining its "plug-&-play" feature; ii) Provide a uniform *convergence* layer (or "logical link layer" using the ISO OSI parlance) to support a large, dynamic *and heterogeneous* (layer-2) network that is capable of connecting hundreds of thousands or more *diverse* physical devices which includes not only Ethernet-equipped devices, but also non-Ethernet devices such as 802.16-based sensors, blue-tooth devices etc., in a *scalable* and *robust* fashion. The proposed VEIL architecture is a shim layer that operates under the traditional network layer (e.g., IPv4/IPv6) and above the ("native") link layer/physical layer such as Ethernet, 802.11 Wireless LANs, etc. VEIL uses a topology-aware, structured *vid* space as the basis for efficient and scalable (DHT-style) object look-up/address resolution, routing and forwarding. VEIL completely eliminates *network-wide flooding* in both the *data* and *control* planes. As a result, VEIL is highly scalable, robust and at the same time offers better support for multi-homing and mobility.

In the following, we provide an overview of the three key components of VEIL: *vid space construction and vid assignment*, *VIRO routing*, and *vidlookup and forwarding*. See [1], [2] for a detailed description of VEIL and the VIRO routing protocol and the different components.

• ***vid* space construction and *vid* assignment:** The *vid* space construction and *vid* assignment to VEIL-switches can



(a) The *vid* layer

(b) *vid* space as a *virtual binary tree*: the grey dotted lines denote physical connectivity, the red dashed lines indicate the logical neighbors at *different* levels of node A; the red boxes represent the *unused vid*'s.

Fig. 1. Overview of VEIL.

be performed in either a centralized or distributed fashion during the bootstrapping phase. In this process, each VEIL-switch is assigned a 32-bit long *switch-vid*. Once the *vid* space has been constructed, when a new VEIL-switch joins the network, its *vid* is assigned based on the subtree (and its neighbors in the subtree) that it is attached to. End-host (*vid*)s are dynamically assigned at the time they join the network. When an end-host is connected via either a wired or wireless link to a VEIL-switch in the network, it is assigned an *extended vid* consisting of the 32-bit *vid* of the *switch-vid* plus a randomly assigned 16-bit local id. Thus the host-device has a unique 48-bit long *vid* which is compatible with the Ethernet MAC header, and used instead of real MAC address while communicating with other host devices. Therefore, it ensures backward compatibility with Ethernet based protocols, and does not require any modifications to end-host devices to work with the VEIL network architecture.

• **VIRO Routing Protocol:** Taking advantage of the topology-aware structured *vid* space, *VEIL* employs a DHT-style routing algorithm (VIRO [1]) to build routing tables at each node so as to maintain network-wide connectivity and perform end-to-end data delivery. In VIRO, routing tables are constructed *piece-meal-wise* using the *vid* logical distance instead of physical distance (e.g., hop counts). As in any DHT routing algorithm, a *pub-sub* or *publish-&-query* mechanism is used by each node to publish and query relevant routing information to build routing entries at each level using a round-by-round, bottom-up procedure. As a result, VIRO completely eliminates *network-wide flooding* in both the data plane (unlike the Ethernet switching algorithm) and *control plane* (unlike OSPF and other shortest path routing algorithms). Furthermore, because of the natural *hierarchical* structure of the *vid* space, routing information regarding far-away part of the network is automatically aggregated using the *vid* prefixes. Hence the routing table size is of the order of $O(\log N)$, where N is the number of nodes in the network, as opposed to $O(N)$ (as in the case of OSPF). Unlike OSPF, VIRO switches need not maintain the topology of the entire network due to the structured *vid* space and hence, changes in network topology do not need to be flooded globally. Due to the aggregate routing information maintained by switches, failure of a link or switch node can be *localized*, without affecting nodes in far-away parts of the network. Furthermore, path and topology diversity can be easily exploited in VIRO by using multiple routing entries. Hence, failure of one routing entry (a nexthop node or link) does not affect network-wide reachability.

• ***vid* Lookup and Forwarding:** The third major component of VEIL is (*vid*) *lookup and forwarding*. VEIL uses IPv4/IPv6 addresses as a persistent identifier for the host devices, which is then mapped to *vid* space in a completely host-agnostic way by capturing ARP query packets and replying with the *vid* of the host as an ARP reply instead of its MAC address. VEIL performs address/name resolution and *vid* look-up by building (standard) DHT look-up mechanisms on top of the same *vid* space. Routing and forwarding between VEIL-switches is performed using (*vid*)s except at the network *edge* where physical addresses/logical addresses/names are needed (between a VEIL-switch and an end-host) to locate individual end-hosts or when data/services are to be delivered.

In summary, 1) VEIL is highly scalable, robust; 2) it decouples routing from addressing, and thus is namespace-independent; 3) it provides seamless and efficient support for multi-homing, mobility and access control; 4) VEIL localizes failures, and provides *built-in* mechanisms for fast rerouting and load-balancing; and 5) VEIL can be readily extended to enable multiple (logical) topologies or multiple virtualized networks on top of the same physical network substrate to further enhance network robustness or service isolation.

Acknowledgment

We gratefully acknowledge the support of our sponsors. The work is supported in part by the National Science Foundation grants CNS-0626808 and CNS-1017092.

References

[1] S. Jain, Y. Chen, Z.-L. Zhang, and S. Jain. Viro: A plug & play, scalable, robust and namespace independent virtual id routing for future networks. In *Tech report*. http://networking.cs.umn.edu/veil/viro.pdf.

[2] S. Jain, Y. Chen, Z.-L. Zhang, and S. Jain. VEIL: A Plug-&-Play Virtual (Ethernet) Id Layer for Below IP Networking. In *First IEEE Workshop on Below IP Networking in conjunction with GLOBECOM*. IEEE, 2009.

[3] C. Kim, M. Caesar, and J. Rexford. Floodless in seattle: a scalable ethernet architecture for large enterprises. *SIGCOMM*, 2008.

[4] R. Perlman et al. Rbridges: transparent routing. In *IEEE INFOCOM*, volume 2, pages 1211–1218. IEEE, 2004.

[5] T. L. Rodeheffer, C. A. Thekkath, and D. C. Anderson. Smartbridge: a scalable bridge architecture. In *SIGCOMM*, pages 205–216, New York, NY, USA, 2000. ACM.

[6] S. Sharma, K. Gopalan, S. Nanda, and T. Chiueh. Viking: A multi-spanning-tree Ethernet architecture for metropolitan area and cluster networks. In *INFOCOM*, volume 4, 2004.