MPLS – Is the Emperor Clothed?

Geoff Huston
August 2001

Multi-Protocol Label Switching, or MPLS, is an approach to switching in IP networks that has generated considerable interest in recent times. MPLS has the characteristics of a hybrid switching architecture, using a label-switching technique borrowed from the connection-oriented switching used in ATM and adapting it to control connectionless forwarding in IP networks.

Label switching is an approach to packet switching where the packet switch uses the incoming packet's header label to make both a local switching decision and also to swap in a new header label that is used to control the actions of the next packet switch. A sequence of such switching decisions and associated label swap operations forms a path across the network. The proponents of this technology point to the ability of label swapping to reduce the size of the packet header field used for switching decisions. In IP's model of destination address-based switching the IP router must look up a table of all known destination address prefixes in order to make a local forwarding decision. Using label swapping, the number of labels in use at any interior label switch can be as low as the number of exit switches, allowing the lookup tables to be far smaller and the label size to be considerably shorter than an IP address. The advantage claimed by this approach to IP forwarding is that the label switching operation can be faster than an IP address lookup in an address table that may have more than 100,000 entries, either by allowing a network's interior switches to be faster and cheaper.

To give a flavour of the way MPLS operates in an IP network, the basic approach used in MPLS for IP is to modify the interior routing protocol so that each interior router associates a label plus a switching decision with each interior route. This association of a label with an interior route is passed to the router's immediate neighbours. Routes to those IP addresses that are external to the network are associated with the best exit router by the routing protocol. When an IP packet enters the MPLS network the destination address of the packet is associated with the IP address of the appropriate exit router. This interior IP address is associated with a label, and an MPLS header is placed on the packet. This table allows the router to make a local forwarding decision. When the packet is received by the next router, this router already has an installed table of labels, where the incoming label is used to make the next forwarding decision, together with the new label to use. Upon exit from the network the MPLS header is removed from the packet, and the original IP packet is then passed out from the network.

The intent of this approach is to take the strengths of both IP routing and connection-oriented forwarding. IP routing allows the network to dynamically detect and route around failures in the network. Connection-oriented label swap forwarding allows for very efficient and fast forwarding decisions.

MPLS is also useful in supporting IP-based Virtual Private Networks. VPNs present a unique set of challenges to IP due to the requirement to ensure that the traffic in each VPN remains wholly within the VPN and that each VPN operates as a private address realm, allowing the same address to be used in distinct VPNs without clashing. This means that if the same provider router is used to forward packets from all VPNs, then the provider router must be able to support a forwarding table for each VPN, and associate each packet with a unique VPN on its transit across the network. Here's where MPLS can really assist. On entry to the network a packet is associated with a VPN and the per-VPN forwarding table is used to determine the exit router and exit interface. A label is pushed onto the packet which describes the exit interface and a second label is pushed onto the packet which describes the label switched path to the exit router. The outer label allows the packet to the switched to the correct exit router, and the inner

label allows the exit router to select the correct outbound interface. Simple, right? Well maybe there are some issues with this approach.

While many ISPs are announcing plans to support MPLS VPNs in their network, others are sounding a cautionary note. It was recently reported in the industry press that researchers at AT&T, network operations expert Randy Bush and security expert Steve Bellovin, have said that MPLS has serious management issues and the MPLS-based VPN approach cannot scale and has potential security and privacy concerns.

Strong stuff indeed. Lets look at these claims and see whether the MPLS emperor really has no clothes.

MPLS in and of itself adds very little to the basic IP network architecture. Its a questionable claim to assert that because labels are shorter than addresses, then switching using short labels means faster switching than switching on IP addresses. At its most basic level all you've done is put one level of indirection between routing and forwarding and introduced a source of potential problem in terms of maintaining a precise alignment between the vector of labels and a hop-by-hop forwarding path. We've already seen IP switching become embedded in chips, and switching speeds for IP packets with address-based lookup in the forwarding table are certainly no slower than switching based on label swapping. So the claim that MPLS gets you improved switching speed is not visible in current vendor products.

The management issues this label switching architecture presents are indeed serious. In an MPLS network you are now relying on the routing protocol to accurately propagate both reachability information and also undertake an associated task of label distribution at the same time. Routing systems are still far from perfect these days, and, as any network operator can attest, their failure modes can be devilishly tricky. When you introduce a second level of indirection between routing and forwarding through the use of a label switching mechanism, the failure modes are potentially far more complex. However this is perhaps something that will be addressed as the technology matures and we gain more experience in operating large MPLS networks, so to claim that the stability of MPLS networks will always pose additional problems for network operators is questionable.

One of the potential roles for MPLS is in the area of traffic engineering, or TE. IP routing protocols tend to discover single best paths, leading to a situation where traffic tends to concentrate on a small set of backbone links. MPLS can be used to configure multiple paths between the pairs of network ingress switches and egress switches, allowing the ingress switch to place entire flows across a diverse set of MPLS paths. In an ideal world this would allow each ingress element to use the full diversity of the network, evenly loading a diverse collection of network paths that lead to a common egress. Its a fine objective, but as usual there's some gap between desire and practical operational management. The bulk of traffic in most IP networks is TCP traffic, and TCP itself is an elastic application, adapting to network congestion by reducing its sending rate. If the traffic sources already adapts to network congestion, then the utility model of TE overlayed across a network of elastic sources presents a strange feedback problem - when is a path 'overloaded'? When you consider that we already have a rich collection of active queue management techniques that are intended to maintain a steady state of a fully occupied link with an average queue length which can be maintained within a pair of preset minimum and maximum length thresholds then the concept of a 'congested' link is a curious one. There may be something in MPLS TE for IP, but it seems that we've managed to get hung up on MPLS itself and the various control structures for TE support without understanding how to apply this to real network states. Its not that MPLS TE may not be useful, but we have yet to understand the interaction between the various end-to-end conversations that are attempting to react to transient congestion, the active queue management systems in each router that are attempting to react to transient congestion and MPLS TE which is also attempting to react to transient congestion.

One of the major roles for MPLS is in supporting VPNs, and here's where the differences in the perspectives on MPLS become quite obvious. There's no doubt that VPN service providers want to enter the world of value-added service provision, and with MPLS VPNs the intent is to say "we can value-add routing on top of basic packet transmission services". The customer simply connects private edge networks to the provider's network and the provider manages both the routing and transmission elements to integrate this site into the VPN - the customer effectively outsources both transmission and routing to the provider. Neat stuff. The problem is one of scaling, as usual. If you see these forms of VPNS as the replacement for all forms of retail transmission services, such as Frame Relay PVCs, ATM PVCs, and even T1 leased lines, then there is no doubt that you have to worry about VPN Provider Edge routers with hundreds, or maybe thousands, of attached individual client networks. And, of course, because the provider is undertaking the routing for the VPN customer, then there is no effective customer-side route aggregation that is assisting here. In this case the size of the routing tables in the PE routers becomes a real issue. and right now the concept is that the provider's VPN routing system is pumping out a complete set of per-customer routing information out to each Provider Edge router. This is not a good idea if scaling is your concern.

Can we make MPLS VPNs scale? Its not looking good as the brute force approach of outsourcing all the routing to the provider's edge loads up the edge routers. So, if the attempts to bury the complexity of per-VPN forwarding decisions into the provider's routing system have scaling problems, then maybe the approach is to look at Layer 2 VPNs. In this approach the MPLS path across the provider's network is viewed as a virtual circuit between two customer sites. This approach attempts to unload the entire routing function back out to the customer edge and allows for less information in the provider edge, which in turn allows the system to scale. But in this model you've lost the value-add proposition of the customer outsourcing routing to the provider. So while this approach has better scaling properties for the provider, the ability for the provider to cost effectively operate a managed VPN service is far more dubious, as now the provider has to manage hundreds, if not tens of thousands of customer edge routers, as well as managing a collection of layer 2 virtual circuits that are MPLS overlays across the provider's backbone network.

The reported security concerns were that if the provider misconfigures the MPLS VPN setup then one customer's traffic is inserted into another customer's network. This is not just a property of MPLS, and indeed in any form of managed shared network, misconfiguration of this sort can cause havoc. I cannot see that MPLS has altered this proposition in any tangible way.

Doubtless there will be more MPLS networks built, and many of them in the VPN market, but to label this approach as doomed reminds me of what they told us about IP networks a decade or so ago, and how IP networks were a management headache, insecure and with no reliable level of service quality, and that customers wanted more from their provider. So, in general, I'm of the view that its early days to say that the emperor has no clothes and we should back out of MPLS as a building block for VPNs. Equally, I'm not convinced that MPLS will truly scale to match the size of the future VPN market, but you may never know that until it breaks!