Road space rationing on the Internet: Pay more to get better throughput in best-effort networks

Katsushi Kobayashi
Advanced Institute for Computing Science, RIKEN, Kobe, JP
ikob@riken.jp

ABSTRACT

In discussions on network neutrality, cost-sharing imbalances between heavy- and light-use subscribers are considered as threats to Internet sustainability. The reason for this is that although exponential traffic growth requires ISPs to invest in network bandwidth, ISPs cannot obtain more revenue from Internet access charges because of flat-rate pricing. We propose a network architecture to realize a "pay more to get better service" policy while preserving flat-rate pricing and best-effort Internet architecture. The architecture is inspired by "road space rationing," a concept applied successfully to travel demand management in big cities such as Athens, Mexico City, and Sao Paulo. The architecture divides the entire network into a large number of network slices, with subscribers divided into the same number of groups. The subscribers of each subgroup normally have access only to the corresponding slice. When better throughput is required, subscribers pay for more accessible network slices. We discuss how to implement the architecture with existing Internet protocols.

1. INTRODUCTION

The exponential growth of Internet traffic remains persistent because of continually emerging network applications such as e-mail, Web, P2P, and VoD. Such growth requires ISPs to continuously invest in their network infrastructures. However, especially in developed countries, the growth of ISP revenue will not sufficiently compensate for this investment. The reasons for this are that the number of subscribers is unlikely to increase because of higher Internet penetration, and that subscription fees per subscriber remain the same because of flat-rate ISP pricing. According to network usage analysis in Japan, a small number of subscribers account for the most traffic on the Internet. For example, only the top 4% of subscribers use more than 2.5 GB/day (or 230 kbps), accounting for 75 and 60% of total uploading and total downloading traffic, respectively [6]. The bandwidth for many other subscribers is very small compared to their access circuits, e.g., the modal download traffic on fiber-to-the-home (FTTH) networks is only 94 MB/day (or 8.7 Kbps), while subscribers' traffic has grown. Therefore, the current cost-share balance between heavy- and light-use subscribers is highly unequal. This imbalance makes it difficult to convince subscribers of the need to increase flat-rate ISP prices, even to support traffic growth. Without ISP investment, it is obvious that network quality will only becomes worse. Furthermore, the Internet could collapse because of shortage of bandwidth capacity in handling traffic growth. These unbalanced and unsustainable cost-sharing problems have been pointed out in network neutrality discussions. Thus, we need a system to provide subscribers with the incentive to share more of the network cost.

For instance, some Japanese ISPs decided to apply a 15-30 GB/day upstream traffic limit for residential access to protect their network quality [1]. The ISPs have offered another service for business subscribers to those customers who want to send more data with a 10 to 100 times higher fee. However, since only a small number of subscribers use the limit, the contribution to redress the imbalance is marginal. Furthermore, even if an ISP offers some differentiation between residential and business subscribers, the differentiation is not ensured throughout the Internet. Service differentiation strategy and its pricing is expected to increase the revenue in Internet access service. Furthermore, the larger the number and more granular differentiation classes are provided, the more subscribers can be satisfied with maximizing the profit of ISPs. For the purpose of sustainable Internet growth, pricing should reflect the usage share of bottleneck that is not only in terms of bandwidth but also in terms of investment. For instance, if most of a bottleneck is subscriber access, such as CATV or ADSL, traffic volume pricing metered by access is appropriate. On other hand, if the bottleneck is not access, such as transit of Tier-1 or access to a popular Web service in an FTTH environment, the volume meter does not reflect the real cost/benefit. In addition, a pricing mechanism valid for inter-domain traffic is expected.

Diffserv provides a differentiated service using a tag in each packet [5]. However, this mechanism cannot contribute to best-effort traffic consuming the most of network bandwidth with numerous Web-based applications. Because Diffserv is designed to obtain higher revenue from traffic, such as VoIP and VPN, other than best-effort traffic. Diffserv's effective area is limited, i.e., to its own ISP and consolidation, because quality of service (QoS) provisioning across several ISPs is still a challenge.

We believe a network architecture that accomplishes differentiation throughout the Internet is key to ensure Internet sustainability. The primary objective of this paper is to establish a network architecture in which traffic from subscribers paying more wins out over that from others paying less, in the form of "pay more to get better service." We propose a novel Internet QoS architecture inspired by road space rationing (RSR) used in managing travel demands. This architecture can be constructed with minimal modifications of existing networks by discussing its implementation. We also discuss how this QoS architecture fits with a Flexible Arrays of Inexpensive Network (FAIN), a parallel Internet architecture we have proposed previously [11]. The rest of the paper is organized as follows. Section 2 presents the original RSR concept and its application on the Internet. Section 3 shows the FAIN architecture we previously proposed. Section 4 discusses how to construct RSR on FAINs. Section 5 discusses operational and implementation issues. We conclude the paper in Section 6.

2. ROAD SPACE RATIONING IN TRAFFIC DEMAND CONTROL AND ITS APPLICATION TO THE INTERNET

In the context of travel demand management in cities, RSR is a well known and effective strategy for reducing vehicle traffic and air pollution with only a small implementation cost. Vehicles are restricted from accessing a designated urban area or city center based on the last digits of the license number on pre-established days and during certain periods (typically the peak hours). When the restriction is based on two digits, a 20% reduction in traffic is theoretically expected. RSR has brought about valuable effects on large cities such as Athens, Mexico City, and Sao Paulo. Although other cities have considered RSR, it has not been implemented because of a fundamental drawback. That is, the wealthy can avoid license number-based restrictions by purchasing a second or third car. This kind of advantage of wealthy founded in RSR is that the Internet QoS architecture has really been expected ever, i.e., "pay more to get better service."

In the following sections, we discuss how the RSR concept applies to the Internet. In general, RSR is a strategy to equally divide both demand and supply into the same number of groups and to allocate each divided sub-supply to a sub-demand. Hence, in the case of traffic demand management, vehicles, considered as demand, are partitioned into two groups by the last digits of the license number. Road space as supply, or more precisely the product of road and time, is divided along the time axis into slots of alternating days. For Internet RSR, the packet rate is considered as demand, and the network bandwidth as supply. Packets can be easily partitioned into an appropriate number of groups using IP header data, which is similar to vehicle license numbers. However, a time-slot division policy for the original RSR on a network is impractical. This is because the temporal resolution of alternating network usage days is much too coarse. Furthermore, only two sub-groups is clearly insufficient to redress the greater than 100-fold imbalance in Internet traffic. On the other hand, a larger number of time-slots reduces the accessible time share for each group. Thus, the network should be partitioned with an alternative approach that does not worsen subscribers' experience, and can provide a large number of groups. The idea that partitioning a network into thousands of sub-network components without protocol modification have been presented, for example, in DynaBone and FAIN [19, 11].

3. FLEXIBLE ARRAYS OF INEXPENSIVE NETWORKS

The FAIN architecture is a packet-switch network similar to the present Internet, and comprises a large number of virtually sliced networks and multiplexer-demultiplexer (MUX-DEMUX) edges such as DynaBone [11, 19]. Since unnecessary synchronization points on the network core are eliminated, a FAIN accomplishes a "parallel-in-global" design infrastructure that prevents a "parallel-in-a-box" architecture.

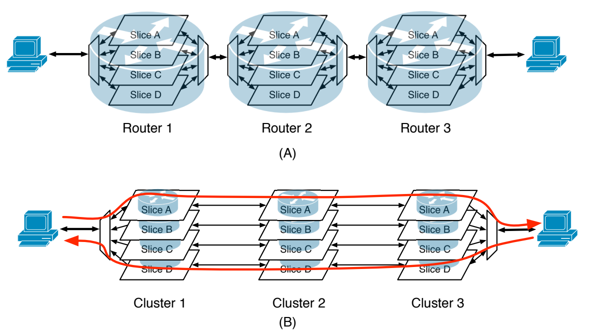

Figure 1: (A) Network comprising "parallel-in-a-box" routers. (B) Network based on "parallel-in-global" architecture.

In FAIN architecture, an entire network is virtually sliced into multiple networks (Fig. 1). All network components except end systems, i.e., routers and links, are sliced into the same number of virtual components. A single control plane, i.e., a routing table, is shared among all the slices (Fig. 2). The network topology of the virtual slice does not differ from that of other slices or from that of the original network before slicing. The data plane of a virtual network slice comprises the virtual forwarding engine of the router and the virtual link, both of which are isolated from the other slices.

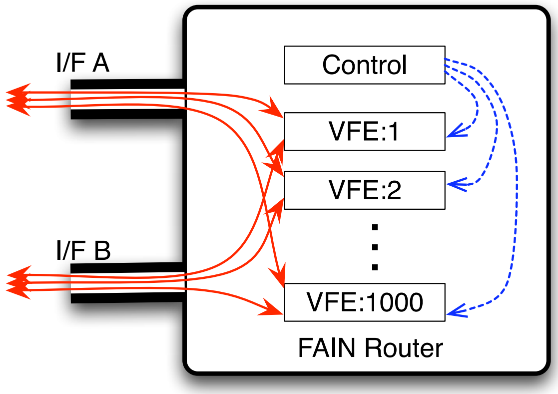

Figure 2: Routing cluster architecture in FAIN router. Both physical interfaces A and B contain a thousand logical links. The FAIN router comprises routing controller and a thousands virtual forwarding engines (VFEs). The solid arrows on left side represent data plane flows, and dotted ones on right side represent interactions between controller and VFEs.

Therefore, a packet should not be passed across different slices. Virtualized FAIN device arrays are allocated to one or more physical components by taking into account the performance and resources of the forwarding engine sub-array to router chassis and the virtual links to a physical one. No synchronization mechanism is required between the physical FAIN devices because the virtual slices are isolated. Simply adding physical device units can increase the FAIN component capacity. To achieve flexible matching between the virtual forwarding engines and the physical components, a large number of slices should be used.

All FAIN network edge routers provide every slice access to the accommodating subscribers because they act as MUX-DEMUX components. With a packet from a network to a subscriber, the FAIN edge ignores the slice identifier of an egress packet and simply forwards it to the subscriber. Conversely, with a packet from a subscriber to a network, the FAIN edge router dispatches an ingress packet on a virtual slice in accordance with the ISP's policy, as mentioned later.

To provide the end-to-end slices of a FAIN, a slice identification field is required in the packet header of the network layer. The slice identifier enables subscribers assigned to more than one slice to dispatch a stream to an appropriate slice based on their own policy, e.g., to maximize use of the assigned slice sub-array. To maximize the flexibility of logical slice allocation to the physical device, a FAIN is required to support a large number of slices, as in an Internet RSR. We have proposed using a 20-bit IPv6 flow label field to specify the FAIN slice identifier [11]. A 20-bit identifier can support one million FAIN network slices. It creates up to one million times accessible slice number differentiation. Note that according to the IPv6 specification, the flow label value is chosen by the end-host [7]. However, for RSR on the Internet, this value should be provided at the subscriber edge because ISPs compel subscribers to restrict accessible slices, as discussed later. This difference only affects the behavior of routers supporting a FAIN. End-hosts and existing routers can be used without any modification.

4. FAIRNESS AMONG NETWORK SLICES

The basic concept of RSR for the Internet is that a subscriber with access to a greater number of slices should receive higher throughput, considering each slice has equal priority at any point in the Internet. In addition, ISPs expect more revenue by controlling the number of accessible slices according to the amount paid by a subscriber. To achieve this, the following function modules are required at the router, subscriber edge, and end-host:

Router: Provide per-slice fairness on every FAIN router.

Subscriber edge: Apply ISP and subscriber contract at ISP-subscriber edges.

End-host: Transport stack or application to use more than one slice on the subscriber side.

We next discuss how to satisfy the above requirements.

4.1 Router: Per-slice fairness

Per-slice fairness can be implemented by allotting each FAIN slice to a dedicated queue using fairness-queuing mechanisms, such as fair queuing, round robin, and deficit round robin [15, 18]. Put another way, our approach is to provide a stochastic fairness system that is effective in the Internet. Stochastic fair queuing (SFQ) has been extensively studied [17]. However, unlike our approach, existing SFQ cannot provide global fairness because it works individually on each router. Justice, proposed by Eriksson et al., propagates the weight of weighted fair queuing (WFQ) to all routers along the path [8]. While Justice provide a fairness system throughout the Internet, it cannot contribute to achieve "pay more for better service" in the global Internet.

4.2 Subscriber Edge: Applying contract

Throughput differentiation is achieved with a number of accessible slices. An ISP provides only a single slice to a residential subscriber. For this type of subscriber, the ISP-subscriber edge router rewrites the slice identifier field in all ingress packets. According to this approach, a residential subscriber is not required to manage FAIN slices. This also allows an ISP to dynamically change slice identifier allocation with its own policy, e.g., load balance.

For a subscriber with access to more than one slice, the edge router should check the slice identifier field to determine whether to follow the contract. An end-host of the subscriber can set the slice identifier field of outgoing packets to deal with his/her own policy and accessible slices. Contrary to single-slice subscribers, subscribers with access to multiple slices must manage slice identifiers on their own. However, multiple-slice access can expand the design space for best-effort Internet architectures, as described below.

4.3 End-host: Using multiple slices

The simplest way to use multiple slices is to assign the transport connections to one of the accessible slices. If a subscriber needs to differentiate between each transport connection, he/she should control the number of connections in each slice, e.g., by limiting the number of connections in each slice for premier customers. For a stalled throughput, the connection can be switched to another slice only by the subscriber. The second approach is to give higher priority to critical connections using multiple slices. The subscriber sends duplicated or FECd packet streams into dispersed slices and decreases the packet loss rate with the number of slices. Furthermore, the subscriber can assign different priority levels to different connections by varying the number of slices. Another possible approach using multiple slices is to simply assign the packets in a connection among the slices. Although the datagram throughput is expected to be proportional to the number of slices, the transport throughput will worsen. The reason for this is that adding a slice also increases the packet loss rate because of per-slice fairness. Consequently, if the subscriber needs a higher data rate, a multiple connections approach, such as Grid FTP in application layer, MPTCP, or CMT-SCTP, should be chosen. It should be noted that all the above-mentioned *advanced* multiple-slices approaches include negative effects, such as wasting bandwidth, increasing out-of-order potential, or requiring additional efforts of end systems. Also note that even in the simplest way to assign connections to a slice, a well designed network is required, e.g., the subscriber's access must be faster than the bandwidth of each slice. However, subscribers who can purchase more than one slice can handle the multiple-slices to take into account both the positive and negative effects.

Although the modules above mentioned work together, the functions do not require mutual conversation among them. For example, the allowed slice identifiers for each user are unchanged for a short period and are specified by the contract. The cost of their implementation might be limited because each is an individual module. Although subscriber differentiation is achieved, it is only applied to ingress traffic from the subscriber side. That is, there is no control mechanism for egress traffic to the subscriber assuming that he/she can receive traffic from any slice in a FAIN. However, because P2P applications use bandwidths in both directions for exchanging data, a strong correlation was found between upstream and downstream volumes, corresponding to the ingress and egress traffic in FAIN edges [16]. Consequently, it might be effective for P2P traffic management to regulate only ingress traffic.

4.4 Multiple connections

Numerous multiple-connection transport studies have been reported [9, 4]. We can categorize these studies according to their objectives. One line of research aims to establish multiple connections along with a single path to achieve higher end-to-end throughput. This is effective under high-bandwidth delay product situations to prevent round trip time (RTT) unfairness, even if it is greedy under other connections. Another set of studies seeks to establish connections associated with all available paths to use bandwidth resources. The single-path approach is still effective even for a single-slice access subscriber. However, given that an inter-slice fairness mechanism protects other slices from this traffic, a multiple-slice access subscriber can expect higher throughput than a single-slice subscriber. The multi-path approach is attractive not only for its bandwidth throughput but also for its resiliency against infrastructure failures. However, the multi-path approach has been a challenge, not because of the transport issue but because of routing. It is difficult to ensure multi-path diversity in the current Internet. If an ISP allows forwarding table growth with additional topology information, a complementary topology approach can provide path diversity using multi-topology routing [10, 13, 14]. This complementary approach satisfies customers who requires faster convergence than the existing routing and faster recovery mechanisms [12]. In the complementary approach, ISPs make a set of routing tables that are disjoint from one another by manipulating the cost metrics. ISPs assign end-to-end slices to each of the complementary table sets. Subscribers can establish multiple connections on multiple paths simply by associating connections with the complementary slices. ISPs can provide complementary topology as a special service only for contracted subscribers by checking the slice identifier at the ISP-subscriber edges.

5. OPERATION AND IMPLEMENTATION ISSUES

This section discusses issues in RSR for the Internet from the viewpoint of operation and implementation of operational networks.

There is an insufficient number of queues on existing routers, e.g., from less than ten to several thousands in current models because of resource limitations [2, 3]. This is obviously poor compared with one million (or 20 bits) using IPv6 flow label. In this case, it is possible to assign more than one slice to a single queue. For instance, the router aggregates slices only using the upper bits as three bits for eight queues. Furthermore, even if the number of queues is different between neighbor routers, assignment can be done individually. The subscriber can expect differentiation up to the number of implemented router queues. Any subscriber can decide how many slices to purchase taking into accont the "effective" number of queues along the utilizing paths. Note that the number of queues along every path cannot be determined because packets traverse across varying paths, several ISPs, and different components. However, the "effective" number of queues can be estimated by the infrastructure situation at the time of purchase.

RSR on the Internet is effective across multiple ISPs, not only a single ISP, because a FAIN slice identifier is unchanged. In peering between FAIN and legacy networks, the border router dispatches a slice identifier to packets from legacy to FAIN domains, the same as a subscriber edge. For ISPs, transit and peering payments are the most important issues. The current pricing is a flat rate, usage basis, or a combination of the two. A FAIN creates additional design space since these pricings can be used in it. For instance, a transit ISP can offer discounted prices to lower tier ISPs by limiting the number of slice accesses. The transit ISP border router overwrites the slice identifier according to the contract. As a result, all subscribers of the lower tier ISP only receive limited throughput with an affordable access charge. This approach breaks the FAIN end-to-end slice architecture. However, it is not a critical modification because the end-to-end slice identifier does not involve an existing middlebox that differs from IP address and transport port number. Even if a FAIN end-to-end slice must be preserved, an ISP should use a tunnel technique such as IP-in-IP or GRE, instead of overwrite.

6. CONCLUSION

We presented the concept of a differentiation mechanism into Internet service business consistent with a "pay more for better service" policy, borrowing from the RSR concept used in traffic demand control. We also discussed how to incorporate the concept into our previously proposed FAIN architecture. The mechanism would help resolve the cost-sharing problem between heavy- and light-use subscribers raised in network neutrality discussions. Of course, the inter-slice fairness only applies an instantaneous rate. It cannot directly redress unbalance in the amount of traffic between subscribers. However, differentiation according to the number of slices can satisfy the heavy-subscribers who continuously generate traffic with any application. We believe the differentiation is a strong incentive for pay higher service classes who can pay more. Moreover, millions of differentiations create new design space even only effective in instantaneous.

Congestion control and management are the most interesting for the industry such as CONEX in IETF. RSR is compatible with any type of congestion control and management, since it is orthogonal to exiting congestion controls. If a link is congested, users in other slices sharing the link are not aware of the congestion because every slice is isolated. This may prevent beneficial co-operation between users. However, users in the same slice can detect congestion, and can co-operate to prevent congestion. On other hand if the slice is consumed by only one user at a congested point, other user traffic is protected, and benefits from slice isolation. Determining which approaches are better for users and ISPs is for future work.

Current Internet infrastructure comprises many Ethernet switches. To map RSR slices to L2 switch queues, the slice identifier information should be put down into L2 headers. A protocol modification in L2 is required to map a full 20-bit IPv6 flow label field. Another possibility is to map RSR slice identifiers to 802.1q VLAN tags. Current Ethernet switches can provide dedicated queues per VLAN(s).

We did not present any reliable evidence supporting our concept, that is, how it is better (or worse) than current Internet architecture and ISPs. For example, it is important to compare RSR with current pricing varieties, such as pay for amount of traffic, traffic-cap, or granular flat-rate for different access bandwidths, i.e., 20 USD/Mo. for 10 Mbps and 30 USD/Mo. for 20 Mbps. Instead, we shed light on the problem and expanded the design space of current best-effort Internet architectures.

Acknowledgments

The author thanks the PFLDNeT'10 reviewers for useful input.

7. REFERENCES

[1] NTT Communications: New Bandwidth Controls to Help Ensure OCN Network Quality. http://www.ntt.com/release_e/news08/0006/0625.html, 2008.

[2] Cisco CRS-3 Modular Services Card (Line Card). Cisco Systems, 2010.

[3] Junos® OS, Class of Service Configuration Guide, Release 10.3. Juniper Networks, 2010.

[4] E. Altman, D. Barman, B. Tuffin, and M. Vojnovic. Parallel tcp sockets: Simple model, throughput and validation. In *Proceedings of the IEEE INFOCOM*, 2006.

[5] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. RFC2475: An Architecture for Differentiated Service. *RFC Editor United States*, 1998.

[6] K. Cho, K. Fukuda, H. Esaki, and A. Kato. Observing slow crustal movement in residential user trafc. In *Proceedings of the 2008 ACM CoNEXT conference*. ACM New York, NY, USA, 2008.

[7] S. Deering and R. Hinden. RFC2460: Internet Protocol, Version 6 (IPv6) Specification. *Internet RFCs*, 1998.

[8] J. Eriksson, M. Faloutsos, and S. Krishnamurthy. Justice: Flexible and enforceable per-source bandwidth allocation. *NETWORKING 2005*, pages 1206–1218, 2005.

[9] T. Hacker, B. Athey, and B. Noble. The end-to-end performance effects of parallel TCP sockets on a lossy wide-area network. In *Parallel and Distributed Processing Symposium., Proceedings International, IPDPS 2002, Abstracts and CD-ROM*, pages 434–443, 2002.

[10] A. Hansen, A. Kvalbein, T. Cicic, S. Gjessing, and O. Lysne. Resilient routing layers for recovery in packet networks. In *Dependable Systems and Networks, 2005. DSN 2005. Proceedings. International Conference on*, pages 238–247.

[11] K. Kobayashi. Flexible arrays of inexpensive network (FAIN): toward global parallelism in the internet to satisfy future traffic growth. In *Proceedings of the 2008 ACM CoNEXT Conference*, pages 1–6. ACM, 2008.

[12] A. Kvalbein, A. Hansen, T. Cicic, S. Gjessing, and O. Lysne. Fast IP network recovery using multiple routing configurations. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–11. IEEE, 2007.

[13] Y. Liao, L. Gao, R. Guerin, and Z. Zhang. Reliable Interdomain Routing Through Multiple Complementary Routing Processes. In *Proceedings of the 2008 ACM CoNEXT conference*. ACM New York, NY, USA, 2008.

[14] M. Motiwala, M. Elmore, N. Feamster, and S. Vempala. Path splicing. 2008.

[15] J. Nagle. On packet switches with infinite storage. *Communications, IEEE Transactions on*, 35(4):435–438, 1987.

[16] L. Plissonneau, J. Costeux, and P. Brown. Analysis of peer-to-peer traffic on ADSL. *PAM2005 (LNCS3431)*, pages 69–82.

[17] N. Shacham and P. McKenney. Packet recovery in high-speed networks using coding and buffer management. In *INFOCOM'90. Ninth Annual Joint Conference of the IEEE Computer and Communication Societies.'The Multiple Facets of Integration'. Proceedings., IEEE*, pages 124–131. IEEE, 1990.

[18] M. Shreedhar and G. Varghese. Efficient fair queueing using deficit round-robin. *IEEE/ACM Transactions on Networking (TON)*, 4(3):385, 1996.

[19] J. Touch, G. Finn, Y. Wang, and L. Eggert. DynaBone: dynamic defense using multi-layer Internet overlays. In *DARPA Information Survivability Conference and Exposition, 2003. Proceedings*, volume 2, pages 271–276. IEEE, 2003.