IEEE iWQoS Workshop, San Hose, California, Aug. 2011

# Future Internet Video Multicasting with Essentially Perfect Resource Utilization and QoS Guarantees

T.H. Szymanski, Bell Canada Chair, Dept.ECE, McMaster University, Canada, teds@mcmaster.ca

***Abstract*—The multicasting of aggregated digital video over a proposed Future Internet backbone network with essentially perfect throughput, resource-utilization and QoS guarantees is summarized. The Future Internet routers require only minor modifications to the existing router designs. Buffers in existing internet routers are partitioned into 2 traffic classes which can co-exist, the *Essentially-Perfect QoS* class and the *Best-Effort* class, i.e., no new buffers are required. Each router includes an FPGA-based *Scheduler Lookup Table* for the essentially perfect QoS class. RSVP-TE is used to provision the multicast trees in an MPLS-TE network. Each router computes an essentially-perfect transmission schedule for all its QoS-enabled traffic flows, which never experience interference or congestion. (This integer-programming scheduling problem is a long-standing unsolved problem.) The Best-Effort traffic is scheduled using the usual Best-Effort schedulers. It is shown that thousands of bursty self-similar video streams can be multicast across the proposed Future Internet with essentially-perfect link efficiencies and QoS guarantees. The technology (i) can be added into new Internet routers with minimal cost (i.e., a few FPGAs); (ii) it allows for the co-existence of the *Essentially-Perfect QoS* and the usual *Best-Effort* traffic classes; (iii) it is compatible with the existing IEFT DiffServ and MPLS-TE service models; (iv) it allows for Internet link efficiencies as high as 100%, and (v) it can reduce Internet router buffer and power requirements significantly.**

***Index Terms*—Future Internet, Quality of Service, essentially perfect QoS, video, multicast, self-similar, buffer sizing, scheduling, low-jitter, link efficiency, power efficiency**

## I. INTRODUCTION

THE Internet network is emerging as a universal platform for delivering new services. However, it suffers from many technical challenges including a reliance on significant over-provisioning, poor resource-utilization and Quality of Service (QoS) guarantees, and high power consumption. The inefficient use of Internet resources is estimated to cost hundreds of millions of dollars in excess energy costs per year and to contribute a noticeable amount of world green-house gasses. To address these problems, the international research community is exploring the *'Future Internet Network'* [1][2], and is open is both *evolutionary* and *revolutionary* changes.

The theory for a *Future Internet* network which can provide *essentially-perfect* link-efficiencies, resource-utilization and QoS guarantees for all *'smooth'* traffic flows, for all admissible network loads $\leq$ 100% of capacity, *has already been established* in [3,6]. A *'smooth'* traffic flow is defined as one which does not exhibit excessively bursty characteristics, and *'essentially-perfect QoS'* is defined as service where the *'Service Lead or Lag'* (SLL), relative to a perfectly-scheduled flow with the same rate, is bounded by at most K packets (please see [3,6] for details). Extensive simulations on the real topologies shown in Fig. 1a, representing about 5 years of work and about $350K of funding, are presented in [3-7].

978-1-4577-0103-0/11/$26.00©2011 IEEE

The proposed Future Internet network can support 2 classes of traffic which can co-exist, a new *Essentially-Perfect QoS* traffic class and the usual *Best-Effort (BE)* traffic class. The Future Internet routers use a new and fast scheduling algorithm which guarantees essentially-perfect service to every admissible traffic flow in the QoS class [3,6]. This integer-programming scheduling algorithm has remained unsolved for several decades (see [3,6,11,12]), and it is one key innovation of the proposed Future Internet. In contrast, scheduling flows to minimize jitter in one IQ switch is NP-Hard [12]. Using the algorithm in [3], all QoS traffic flows will never experience interference or congestion, and the queueing delay per router will be relatively small. The technology can be added to new routers with minimal cost, i.e., a few FPGAs per router.

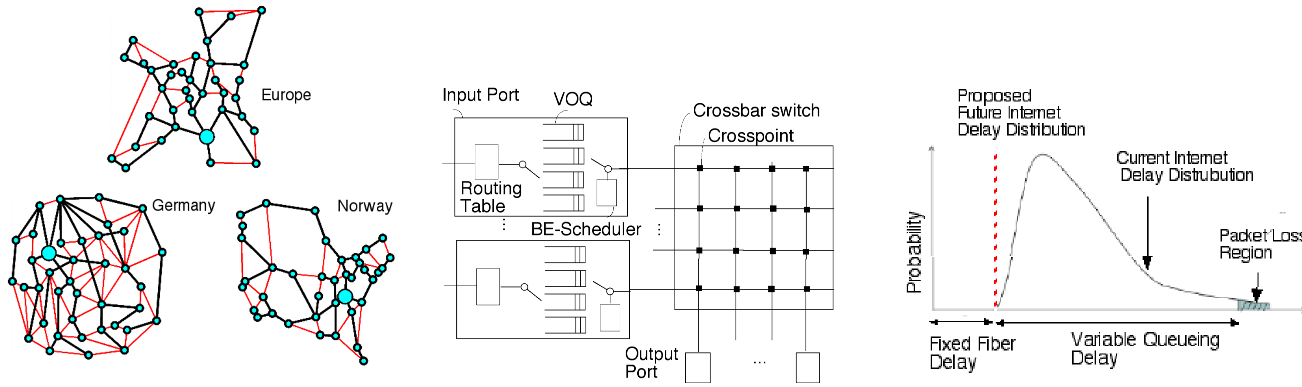
The multicasting of aggregated video over the proposed Future Internet is explored. A real H.264 video trace available at [14] is aggregated to generate 50 hours of data for aggregations of 10, 100, ..., 1M video streams. These aggregated streams are used to quantify the link-utilization and QoS guarantees of the proposed Future Internet network. It is shown that aggregated self-similar video streams can be delivered across the proposed Future Internet, achieving up to 100% link-efficiencies and resource-utilization, and simultaneously achieving essentially-perfect end-to-end QoS guarantees, for all QoS-enabled traffic. The same techniques can be applied to Metropolitan and Local Area Networks, and all-optical networks.

Section 2 of this extended abstract summarizes the technology. Section 3 summarizes the experimental results.

## II. INTERNET ROUTERS

Several topologies for Internet backbone networks are shown in Fig. 1a, including the Germany, Norway and European backbone networks [8]. The bold nodes represent the roots of IP multicast trees which reach all other nodes, and each bold link represents an edge in a multicast tree. The multicast trees can distribute 1,000s of aggregated video channels to every destination node, thereby replacing the traditional cable and satellite video distribution systems.

Current Internet routers use a *'Bandwidth-Delay-Product'* rule to determine the buffer sizes for each link [9,10]. Each link requires a buffer of size $O(C \cdot RTT)$, where $C$ is the link capacity and $RTT$ is the round-trip-time. The buffers can exist at the input side, the output side or within the router. An *Input-Queued* (IQ) router with buffers at the input side is shown in




(a) (b) (c) Delay

Fig. 1. (a) Internet backbone topologies (b) Internet packet/cell end-to-end delay variation.

Fig. 1b. Each input port in an NxN switch has N *Virtual Output Queues* (VOQs). The routers use a *Best-Effort* (BE) scheduler to schedule packet transmissions from the VOQs through the switch. An IP router operating at 40 Gbps with a mean RTT of 250 millisec. requires buffers of 10 Gbits or $\approx$ 1 million IP packets per input port. A distribution for the typical end-to-end delay across the current Best-Effort Internet is shown in Fig. 1c [6]. A packet can encounter congestion on each link in an end-to-end path, and the *worst-case delay on each link* can approach the RTT $\approx$ 250 millisec. A major problem of the current Internet is illustrated in Fig. 1c, i.e., the statistical delay spread and the lack of mathematically-provable delay and QoS guarantees for end-to-end traffic flows.

Existing Internet routers use *Best-Effort* (BE)-schedulers as shown in Fig. 1b, which cannot achieve rigorous QoS guarantees or 100% throughput. It is often said that *'a chain is only as strong as its weakest link'*, and one weak link in Internet QoS is the BE-schedulers found in current Internet routers. As a result, existing Internet routers are typically over-provisioned for real-time traffic and operate at a small fraction (i.e., 25-50%) of peak capacity and resource-utilization (i.e., link utilization).

Fig. 2a illustrates a proposed input port design, which allows QoS-enabled traffic to co-exist with the usual BE traffic. In each input port, the VOQs are logically partitioned into 2 classes, the QoS-VOQs and the BE-VOQs. These VOQs may reside in the same physical memory, but they are logically shown to be distinct in Fig. 2a. Each input port has two software-loadable *Scheduler Lookup Tables* (SLTs). The SLT(1) selects either the QoS-VOQs or the BE-VOQs for service. The SLT(2) selects one particular QoS-VOQ for service, from the set of QoS-VOQs. The SLTs can have the length of a scheduling frame, i.e., $F$=1,024 time-slots. No extra buffers are required, since Internet routers already maintain very large buffers. The SLT entries can be easily computed in software periodically using the scheduling algorithm in [3] in a network processor. The only new hardware required is the software-loadable SLTs and some logic, which can easily fit on a small FPGA per input-port.

The schedules for QoS traffic may be updated (or recomputed) when the RSVP-TE signalling protocol adds or removes a QoS traffic flow from a switch. As observed by researchers at Bell Labs., the switch schedules can be updated relatively slowly [12], perhaps 100 times per second.

## III. Multicasting Aggregated Video Streams

In this section, the aggregation of 10 up to 1M high-definition (HD) H.264 video streams is explored. The Mars-to-China trace available at [14] is a 30 minute HD 1920x1080 video trace in the H.264/AVC format, with a mean bit rate of 4.85 Mbps and a compression ratio of 154. The ratio of the peak to minimum frame size is 4,086, indicating a very high degree of burstiness.

Copies of the Mars-to-China video stream are aggregated at a source node, and shaped into a low-jitter stream in a *Video Shaper Queue* (VSQ), i.e., a token bucket shaper with a relatively small token bucket. The VSQ can be modelled as a discrete-time discrete-state $D^x/D^y/1$ queueing system, with deterministic inter-arrival and inter-departure times D (please see [5-7] for details). Batches of video frames arrive and depart from the VSQ at deterministic times (i.e., D=1/30 sec. per batch) with batch size distributions $x$ and $y$ respectively. The arriving batch-size distribution $x$ is determined from the statistics of the arriving video streams. The departing batch-size distribution $y$ is determined from the provisioned service-rate of the multicast tree, and the bound on the SLL of the VSQ at the source node. The batch inter-arrival/departure times $D = 1/30$ sec. can be adjusted to any value i.e., $D = 1$ millisec, but the results do not change significantly. Video streams are recovered at each destination node, from the low-jitter aggregated stream, in a *Video Playback Queue* (VPQ) [5-7].

Fig. 2b illustrates the mean VSQ delay vs. the excess bandwidth provisioned in the multicast tree, for numerous degrees of aggregation. Each point reports the mean result over 100 measurements. The 95% confidence intervals are also drawn. Fig. 2c illustrates the mean VPQ delay vs. the excess bandwidth provisioned in the multicast tree, for numerous

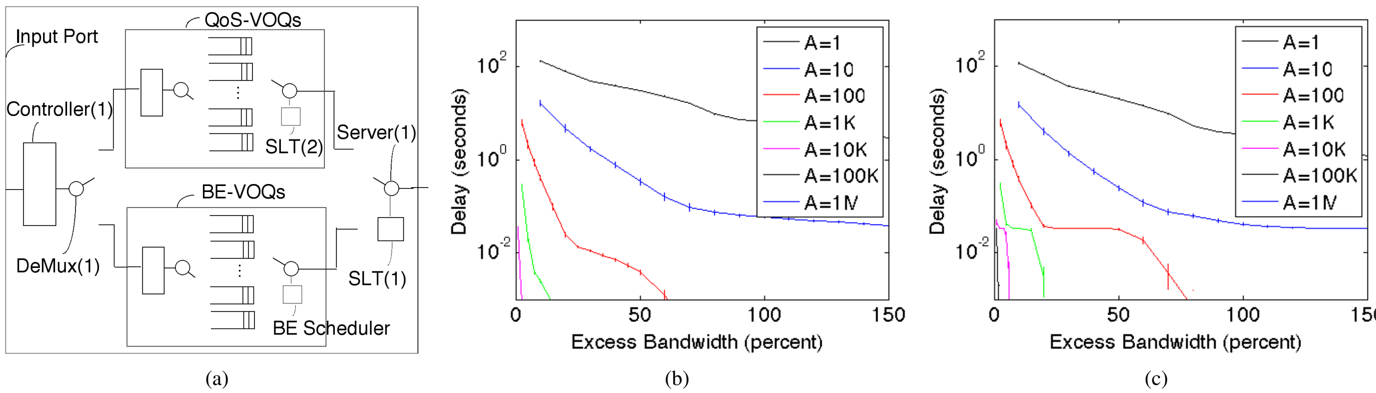



Fig. 2. (a) new Input Port design (b) SQ delay (c) PBQ delay.

degrees of aggregation. For all levels of aggregation, the VSQ and VPQ delays drop rapidly as the excess bandwidth increases above zero.

Table 1 illustrates the various queueing delays in an end-to-end path traversing 20 routers, using the proposed QoS technology. The aggregation of 1,000 video streams each requiring 4.85 Mbits per second (Mbps) will require an aggregate bandwidth of 4.85 Gbps. To achieve a small delay, an excess bandwidth of 4% is provisioned in the multicast tree, so the provisioned rate is 1.04x4.85=5.044 Gbps. Let the mean IP packet size be 1500 bytes. According to theorems 1-4 stated in [6], the mean queueing delay along an end-to-end path of 20 routers is $\leq$ 20*2*2.4 $\mu$sec $\leq$ 0.1 millisec. Given the aggregation of $10^3$ video streams, a provisioning of 4% excess bandwidth will result in a VSQ delay of $\approx$ 23 millisec, a VPQ delay of $\approx$ 65 millisec, and a small cumulative mean queueing delay within the Internet routers of $\leq$ 0.1 millisec. An aggregation of $10^5$ or more video streams can achieve total end-to-end queueing delays $\leq$ 3 millisec with 1% excess bandwidth, thereby achieving a link-efficiency of $\geq$ 99%.

TABLE I
END-TO-END DELAY BOUNDS FOR AGGREGATED H.264/AVC VIDEO STREAMS.

| Channels | EXBW | Hops | RQ Delay | SQ Delay | PBQ Delay |
|---|---|---|---|---|---|
| | | | (millisec) | (millisec) | (millisec) |
| 1 | 100% | 20 | 49.5 | 50 sec | 50 sec |
| 10 | 50% | 20 | 6.6 | 0.5 sec | 0.5 sec |
| 100 | 25% | 20 | 0.8 | 20 | 20 |
| 1,000 | 4% | 20 | $\leq$ 0.1 | 23 | 65 |
| 10,000 | 2% | 20 | $\leq$ 0.01 | 1.9 | 33 |
| $10^5$ | 1% | 20 | $\leq$ 0.01 | $\leq$ 1 | $\leq$ 1 |

For the backbone topologies shown in Fig. 1a, the buffer requirements per router can be reduced by factors of 100 - 10,000 if all traffic uses the QoS-enabled traffic class. Each QoS-enabled traffic flow requires $\leq$ 2 packet buffers per router on average [3-7]. The new technology allows networks to achieve higher link efficiencies and resource-utilizations, i.e., backbone links carrying real-time traffic can operate at close to 100% capacity.

According to theory presented in [3-7], this technology will allow for the design of a new generation of Internet routers which can offer essentially-perfect QoS guarantees and resource utilizations, using for example RSVP-TE and MPLS-TE, and which can be smaller, faster and consume less power. Content providers can provision multimedia streams with QoS guarantees, using relatively small excess bandwidths. Internet backbone capacities can be increased and energy costs can be reduced. To the best of our knowledge, no other theory exists for achieving *Essentially-Perfect end-to-end QoS* guarantees in packet switched networks without speedup, i.e., see [3,6,9-14]. Our extensive simulations reported in [3-7] are all in agreement with the theory. If any readers have interesting routing or scheduling problems, please email this author.

## REFERENCES


[1] European Future Internet Portal, www.future-internet.eu, 2011
[2] GENI "Global Environment for Network Innovations - Exploring Networks of the Future", www.genie.net.
[3] T.H. Szymanski, "A Low-Jitter Guaranteed-Rate Scheduling Algorithm for Packet-Switched IP Routers", IEEE Trans.Comm., Nov. 2009.
[4] T.H. Szymanski, "Method and Apparatus to Schedule Packets Through a Crossbar Switch with Delay Guarantees", US Patent Application.
[5] T.H. Szymanski and D. Gilbert, "Internet Multicasting of IPTV with Essentially-Zero Delay Jitter", IEEE Trans. Broadcasting, March 2009
[6] T.H. Szymanski and D. Gilbert, "Provisioning Mission-Critical Telerobotic Control Systems over Internet Backbone Networks with Essentially-Perfect QoS", IEEE JSAC, Vol. 28, No. 5., June 2010
[7] T.H. Szymanski and D. Gilbert, "Design of an IPTV Multicast System for Internet Backbone Networks", Int. Journal Digital Multimedia Broadcasting, Vol. 2010, Article 169140
[8] S. Orlowski, R. Wessaly, M. Pioro, A. Tomaszewski, "SNDlib 1.0 - Survivable Network Design Library", Networks, Oct. 2009.
[9] R.S. Prasas, C. Dovrolis, M. Thottan, "Router Buffer Sizing for TCP Traffic and the Role of the Output/Input Capacity Ratio", IEEE Trans. Networking, 2009.
[10] S. Iyer, RR. Kompella, N. Mckeown, "Designing Packet Buffers for Router Linecards", IEEE Trans. Networking, Vol. 16, No. 3, June 2008
[11] W.J. Chen, C-S. Chang. and H-Y. Huang, "Birkhoff-von Neumann Input Buffered Crossbar Switches", IEEE Trans. Comm., Vol. 49, No. 7, July 2001
[12] I. Keslassy, M. Kodialam, T.V. Lakshman and D. Stilliadis, "On Guaranteed Smooth Scheduling for Input-Queued Switches", IEEE/ACM Trans. Networking, Vol. 13, No. 6, Dec. 2005
[13] A. Lombardo, G. Schembra, and G. Morabito, 'Traffic Specifications for the Transmission of Stored MPEG Video on the Internet', IEEE Trans. Multimedia, Vol. 3, No. 1, March 2001.
[14] P. Seelingm, M. Reisslein, and B. Kulapa, Network performance evaluation using frame size and quality traces of single layer and two layer video", IEEE Comm. Surveys, Vol. 6, No. 3 2004.