Supporting Consumer Services in a Deterministic Industrial Internet Core Network

Ted H. Szymanski

A convergence is occurring in the networking world. Industrial networks currently provide deterministic services in robotic factories and aircraft, while the best effort Internet of Things provides best effort services for consumers. The author argues that a convergence should occur, and that a future converged industrial Internet of Things (IIoT) should support both best effort and deterministic services, with very low latency and jitter.

ABSTRACT

A convergence is occurring in the networking world. Industrial networks currently provide deterministic services in robotic factories and aircraft, while the best effort Internet of Things provides best effort services for consumers. We argue that a convergence should occur, and that a future converged Industrial Internet of Things (IIoT) should support both best effort and deterministic services, with very low latency and jitter. This article presents the design of a deterministic IIoT core network consisting of many simple deterministic packet switches configured by an SDN control plane. The use of deterministic communications can reduce router buffer sizes by a factor of ≥ 1000, and can reduce end-to-end latencies to the speed of light in fiber. A speed-of-light deterministic core network can have a profound impact on virtually all consumer services such as multimedia distribution, e-Commerce, and cloud computing or gaming systems. Highly aggregated video streams can be delivered over a deterministic virtual network with very high link utilization (≤ 100 percent), very low packet jitter (≤ 10 μs), and zero congestion. In addition to improving consumer services, a converged deterministic IIoT core network can save billions of dollars per year as a result of significantly improved network utilization and energy efficiency.

INTRODUCTION

The existing best effort Internet of Things (BE-IoT) suffers from congestion and provides inefficient best effort service for consumers. It provides no guarantees for the bandwidth, delay, or jitter of a consumer's Internet connection(s), and it is typically overprovisioned to operate at light loads, to reduce delay, jitter, and packet loss rate. This over-provisioning costs service providers several billions of dollars per year in excess capital costs and energy costs, and large delays still occur frequently during times of congestion. As a result of congestion, the BE-IoT cannot support the demanding machine-to-machine (M2M) communications required in robotic factories, airplanes, and space craft. We argue that the future converged industrial Internet of Things (IIoT) should support both best effort services for consumers and deterministic services for M2M communications, where the end-to-end delay, jitter, and packet loss rate can be deterministically bounded.

General Electric (GE) coined the term Industrial Internet to acknowledge the growing importance of connecting industrial machines rather than humans. Industrial automation will use the IIoT to enable a new wave of robotic manufacturing, by interconnecting industrial sensors, control systems, and robots. GE envisions that the transformation to industrial automation may impact the world on the same scale as the industrial revolution of the 19th century. It estimates that industrial automation may increase worldwide GDP by $15 trillion by 2030 by reducing costs and waste, and improving manufacturing processes. GE also estimates that the IIoT may control about $82 trillion of industrial GDP by 2030, representing about half of the world's GDP. In March 2014, five companies (GE, Cisco, AT&T, IBM, and Intel) formed the Industrial Internet Consortium to advance the technologies, and in June 2015 the Consortium included 160 companies, indicating strong industrial support.

Reduction of the large Internet latencies has received significant attention lately. In 2013, the Association of Computer Manufacturers (ACM) and the Internet Society held a workshop on reducing Internet latencies, which concluded that unnecessary delays should be removed from every layer of the protocol stack [1]. A recent ACM paper, "The Internet at the Speed of Light," shows that Internet latencies are typically 10× to 100× larger than the minimum delays due to the speed of light in fiber [2]. They argue that a speed-of-light Internet would be a "technological leap" forward that could fundamentally transform computing. For example, a speed-of-light Internet could transform cloud computing or gaming systems, both multi-billion-dollar industries, by dramatically increasing the size of the reachable population (of machines or people) given a fixed latency. In 2014, the International Telecommunication Union (ITU) began to explore the impact of a Tactile Internet network, with a goal to reduce end-to-end latencies to 1 ms. They argue that the Tactile Internet would add a new dimension to human-machine interaction and revolutionize M2M interaction [3].

Large Internet latencies lead to very high costs in the e-Commerce industry. In 2014, global e-Commerce revenue was about US$1.2 trillion.

The author is with McMaster University.

0163-6804/16/$25.00 © 2016 IEEE

A 100 ms latency penalty can reduce sales for Amazon by 1 percent, and similar figures have been reported for Bing and Google [2]. Amazon's sales revenues were US$89 billion in 2014, and a 1 percent loss represents over US$1 billion in 2015. According to Akamai, a quick page load time is a key factor in a consumer's loyalty to an e-Commerce site, as 40 percent will wait no longer than 3 s before abandoning the site.

Large Internet latencies also lead to very high costs in the financial services industry. A 1 ms increase in latency can reduce revenue by US$100 million per year for firms performing high frequency automated stock trading. Internet latencies can be reduced by deploying new fiber, but the cost is prohibitive. For example, the cost of deploying new fiber under the Arctic Circle to reduce the London-to-Tokyo latency by 60 ms is US$1.5 billion.

According to Sandvine Networks, large-scale video distribution from services such as YouTube and NetFlix currently consumes about 50 percent of the continental U.S. core bandwidth at peak times. This figure is expected to rise to potentially 90 percent in the future.

We believe that a convergence of the best effort and deterministic communications paradigms into a single unified network should occur. This article first presents the design of a deterministic IIoT core network based on a network of simple deterministic packet switches controlled by a software defined networking (SDN) control plane [4]. The packet switches can operate at layer 2 or 3, as shown in Fig. 1. Our SDN control plane can program thousands of deterministic virtual networks (VNs) into the IIoT core to distribute highly aggregated video streams, as shown in Fig. 2. Our deterministic IIoT design has three unique features:

- It can provably operate all Internet links at 100 percent loads.
- It can simultaneously reduce end-to-end transport delays to the speed of light in fiber.
- The complexity of scheduling traffic through the switches with low jitter is not NP-Hard [4].

We show that the ability to operate the future deterministic core network at 100 percent capacity can lead to potential capital cost savings of US$37 billion per year.

This article is organized as follows. We discuss the evolution to a converged IIoT network. We present the design of a deterministic packet switch for layer 2 or 3. We present a deterministic U.S. core network and its performance. We explore the distribution of aggregated video over the converged IIoT. We conclude the article.[1]

The Evolution to Deterministic Services

The existing BE-IoT poses several challenges for industrial automation and the consumer services industry. The BE-IoT suffers from congestion, which causes:

- Excessively high end-to-end delays potentially as large as 50–500 ms
- Potentially high packet loss rates of 5–50 percent, unless the network is significantly overprovisioned [4]

The Internet Engineering Task Force (IETF) acknowledges that overprovisioning lowers the utilization of the BE-IoT infrastructure to typically below 50 percent [5, 6]. The IETF has ruled out overprovisioning as a means to achieve deterministic services, and aims to achieve at least 50 percent link utilizations for deterministic traffic flows in the future Internet.

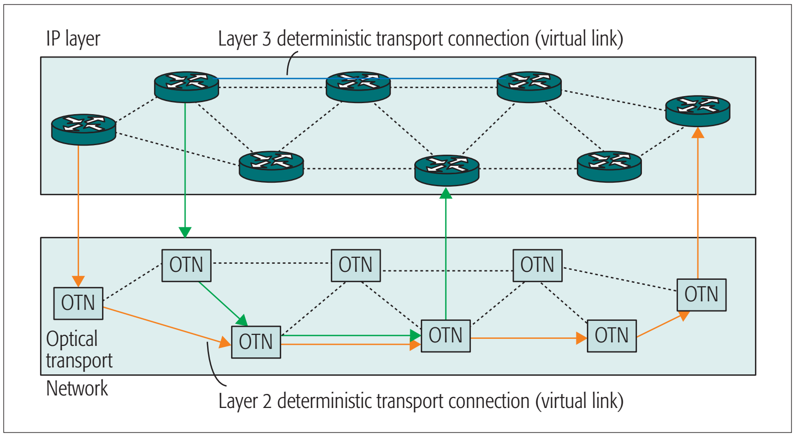
IP layer
Layer 3 deterministic transport connection (virtual link)
OTN
OTN
OTN
OTN
OTN
OTN
OTN
Optical transport Network
Layer 2 deterministic transport connection (virtual link)

Figure 1. A layer 3 network of IP routers, with a layer 2 optical transport network (OTN) underlay. "Deterministic transport connections" (virtual links) can be embedded into each layer.

Current BE-IoT routers typically use a bandwidth-delay product buffer sizing rule, which provides buffers for about 250 ms of data per IO port to provide congestion control for worst case scenarios [4, 7]. A router with 400 Gb/s links has buffers for about 100 Gbits of data per IO port (in the worst case), equivalent to about 8.3 million maximum-size IP packets. Referring to Fig. 2 and assuming 400 Gb/s links, the BE-IoT router in Chicago will have buffers for about 32 million IP packets. These large buffers increase BE-IoT router complexity, costs, power consumption, and failure rates, and play a key role in the BE-IoTs excessive delays during times of congestion.

ATM and MPLS-TE Core Networks

A deterministic traffic flow is immune to congestion and interference from all other traffic flows, and can also be called a guaranteed rate (GR) or constant bit rate (CBR) flow. In the 1990s the international community developed the asynchronous transfer mode (ATM) standard with CBR service (in principle). Unfortunately, the problem of scheduling CBR traffic flows through an input-queued packet switch with minimum delay and jitter is a well-known NP-Hard problem; see [4, 8–10]. The ATM standard did not solve the NP-Hard switch scheduling problem and could not provide a true deterministic service [4]

The ATM standard was eventually abandoned. Multiprotocol label switching with traffic engineering (MPLS-TE) was developed shortly thereafter to offer improved service. However, the MPLS-TE standard also did not solve the NP-Hard switch scheduling problem and could not provide a true deterministic service [4].

MPLS-TE exists today, but it does not provide a true deterministic service for industrial automation, robotic manufacturing, and mission-critical M2M communications [4].

[1] A related video, "A Speed of Light Deterministic Industrial Internet of Things," can be found at https://youtu.be/cXA0HEjKRPY

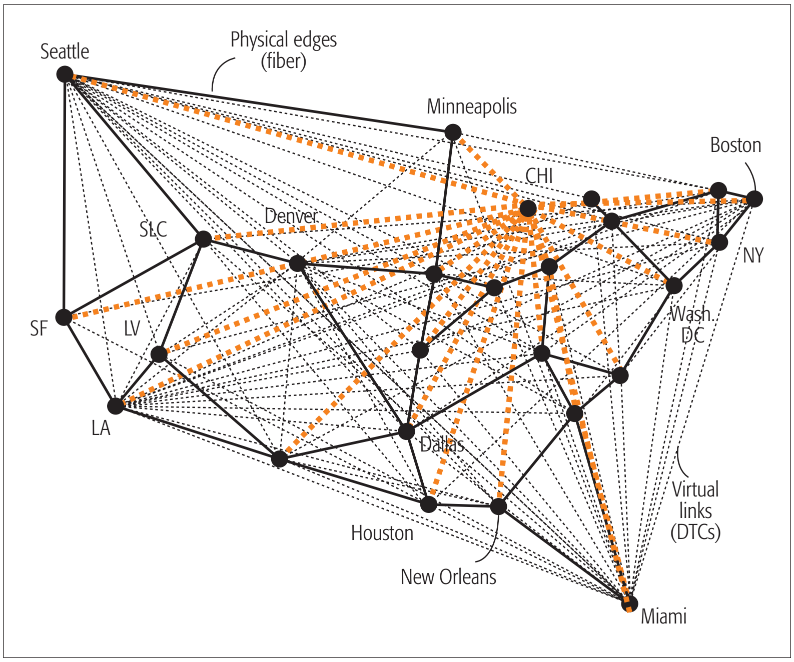

Figure 2. A deterministic U.S. IIoT core network with virtual links originating at six cities (Seattle, Los Angeles, Denver, Chicago, Boston, and Miami). A virtual network supporting video distribution from Chicago is highlighted (in red).

Existing Industrial Networks

Proprietary industrial networks offering low-latency deterministic M2M services have existed for years in industrial automation and the avionics industry. However, the proprietary nature of these networks has increased costs and limited wide-scale deployment. For example, Airbus developed the patented Avionics Full-Duplex Switched Ethernet (AFDX) network for the Airbus 380. The A380 requires over 500,000 m of control wires. Unfortunately, in 2006 wiring problems (the wires were a few inches too short) delayed the A380 project, leading to cost overruns of €2 billion. The proprietary nature of the control wires meant that low-cost replacements were not readily available. In 2014, wiring problems delayed the new U.S. Air Force KC-46 refueling tanker project leading to cost overruns of US$1.5 billion. The IEEE recently developed the deterministic Ethernet standard to provide an open low-cost standard to support both best effort and deterministic M2M services, to avoid similar problems recurring.

The Deterministic Ethernet Access Network

To address the need for deterministic M2M services in factories, vehicles, trains, planes, and audio/video applications, the IEEE developed the 802.1Q standard for Deterministic Ethernet to provide both deterministic and best effort services on a single Ethernet link [11]. The standard requires that a packet must be delivered within a deterministic time bound, but for flexibility it does not specify any scheduling algorithms. Typically, an application will explicitly reserve times for data transmissions on the Ethernet broadcast medium to achieve a deterministic delay bound. The IEEE standard requires that all applications on the broadcast medium are synchronized, potentially to within nanoseconds or microseconds of accuracy, to avoid packet collisions. The IEEE also added 3 bytes to the basic Ethernet packet size to allow for the identification of 16 million virtual networks.

Layer 2 networks are usually small, and are typically interconnected with service provider bridges and backbone bridges. Traditionally, a layer 3 IP core network interconnects "islands" of smaller layer 2 networks. The IEEE is currently looking at the requirements for providing deterministic services in larger layer 2 networks, such as bridged and switched Ethernet networks and rings, to support real-time M2M services. However, the introduction of switches significantly complicates the provisioning of deterministic services, since the problem of scheduling deterministic traffic flows through one switch with minimum delay and jitter is NP-Hard in general [10]. Our scheduling algorithms can also be used to program deterministic layer 2 bridges.

In Fig. 1, our layer 2 OTN can span a continent, where each simple deterministic packet switch can ideally fit on a field programmable gate array (FPGA). The layer 2 switches must obey strict deterministic packet forwarding schedules and could use any transport-oriented packet format, for example, the Deterministic Ethernet or carrier Ethernet packet formats. Hence, the network in Fig. 1 can be viewed as IP-over-Carrier-Ethernet-over-dense wavelength-division multiplexing (DWDM).

The Wireless TSCH Access Network

The IETF has created a Working Group, 6TiSCH, to incorporate IEEE's time synchronized channel hopping (TSCH) wireless standard into the IP infrastructure [12]. The standard will provide deterministic real-time M2M services for "last-mile" wireless access networks supporting IPv6. The TSCH standard allows a wireless node to explicitly reserve time slots for transmission on several frequency-based channels. The transmissions of a traffic flow will typically experience extensive frequency hopping to mitigate the effects of wireless fading and interference in any one frequency. However, for flexibility the standard does not specify the scheduling algorithms to be used.

IETF Activities in Deterministic Networks

In October 2015, the IETF approved the Deterministic Networking Group to explore the feasibility of adding deterministic services to the BE Internet network (as a work in progress). The IETF has published a draft Deterministic Networking Problem Statement [5] and a Deterministic Forwarding Per Hop Behavior (PHB) [6] draft for use with the differentiated services (DiffServ) service model. These drafts specify an abstract model rather than a detailed technical solution. The drafts require that the packets in a deterministic flow must receive deterministic service in each Internet router, but for flexibility they do not specify the router architecture or any routing/scheduling algorithms. The IETF drafts require that all routers are synchronized, to within 10 ns–10 μs of accuracy. This tight synchronization is a potential problem for a deterministic network that spans a continent as shown in Fig. 2 (our approach solves this problem).

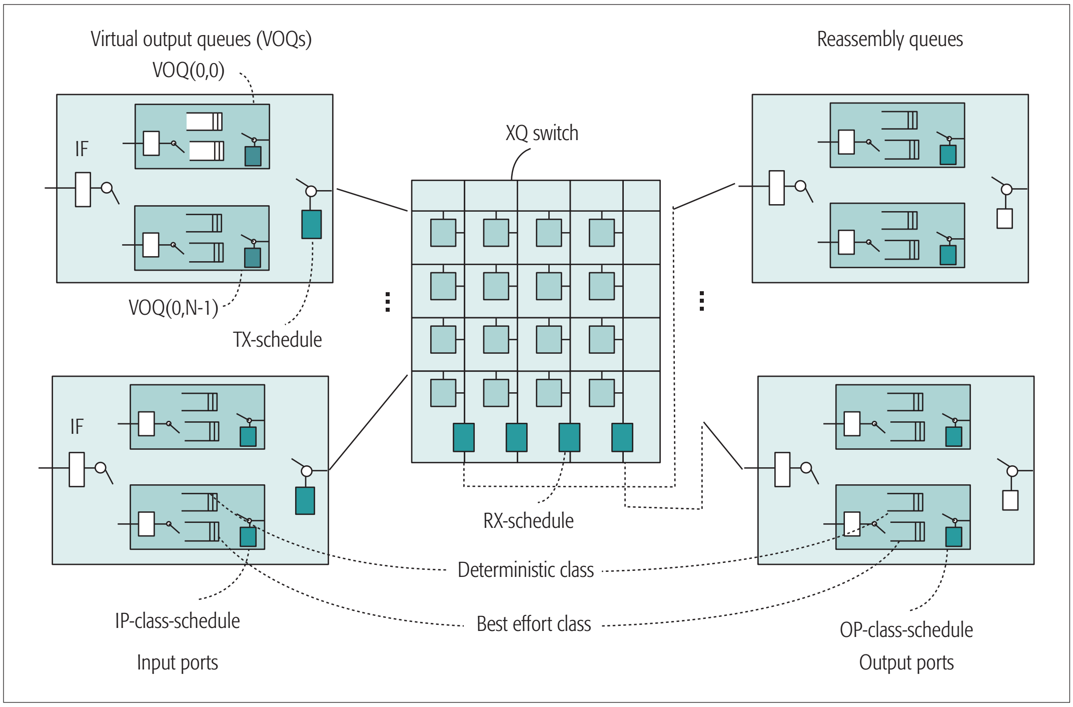

Figure 3. Basic deterministic switch with combined input, crosspoint and output queueing (CIXOQ).

In late 2015, the Industrial Internet Consortium published a draft Industrial Internet Reference Architecture. The first draft identifies the most important architectural issues and it is "broad rather than deep." The architecture does not mention deterministic communications or time-synchronization requirements, but it does discuss the use of prioritization to achieve better service for M2M flows.

The IETF has proposed several use cases for deterministic communications, including:

- Professional audio over the Internet
- Deterministic radio access networks
- Deterministic mobile networks
- Deterministic control for utilities such as the smart power grid

The existing power grid distributes vast amounts of power over a network of high-voltage transmission lines. The ability to increase transmission line utilizations by 10 percent can lead to potential capital cost savings of several billion dollars [13]. However, the future smart power grid will require a very fast control system. According to the IETF, jitter of less than 250 μs and end-to-end delays of less than 4–10 ms are needed [13]. The deterministic U.S. IIoT network shown in Fig. 2 can meet a 10 ms delay constraint over distances of about 2000 km, and the jitter is less than 10 μs.

The Industrial Internet and Tactile Internet Projects

In late 2015, the Industrial Internet Consortium published a draft Industrial Internet Reference Architecture. The first draft identifies the most important architectural issues and is broad rather than deep. The architecture does not mention deterministic communications or time synchronization requirements, but it does discuss the use of prioritization to achieve better service for M2M flows. In this article, we argue that deterministic communications offers several benefits over best effort communications using prioritization, and present a deterministic Industrial Internet core network.

In 2014, the ITU began a project on the Tactile Internet to describe a future Internet network with exceptionally low end-to-end latency, and high availability, reliability and security, for applications including industrial automation and smart transportation systems. This project does not mention deterministic communications or time synchronization requirements. The Tactile Internet project has the same goals as the Industrial Internet project.

A Deterministic Packet Switch

Packet switches use several types of queueing, including input queueing (IQ), output queueing (OQ), and combined input and output queueing (CIOQ). An $N \times N$ OQ switch can achieve 100 percent throughput with minimum delay; however, it requires an internal speedup of N to remove all contention for output ports, which increases costs and power use. Large OQ switches are intractable and are rarely used [4].

An IQ or CIOQ switch can achieve 100 percent throughput with no internal speedup, or with a small Internet speedup of typically 2 or 4. However, complex scheduling algorithms are needed to schedule the packets through the switch with 100 percent throughput and without contention [8, 9]. The problem of scheduling deterministic traffic flows through an IQ or CIOQ switch with no speedup, 100 percent throughput, and minimum delay and jitter is NP-Hard; see [4, 10].

Figure 3 illustrates a simple packet switch that supports deterministic traffics flows with 100 percent throughput, and deterministic delay and jitter guarantees [14]. The switch adds crosspoint queues (XQs) to the basic CIOQ switch, yielding a combined input, crosspoint, and output queueing (CIXOQ) switch. The majority of buffering occurs at the input ports (IPs) and output ports (OPs); the XQs are very small, and exist only to simplify the scheduling algorithms. Variable-size Internet packets arrive at the IPs, and are typically fragmented into fixed sized cells (with 64 or 128 bytes) for transmission through the switch. The variable-size Internet packets are reassem-

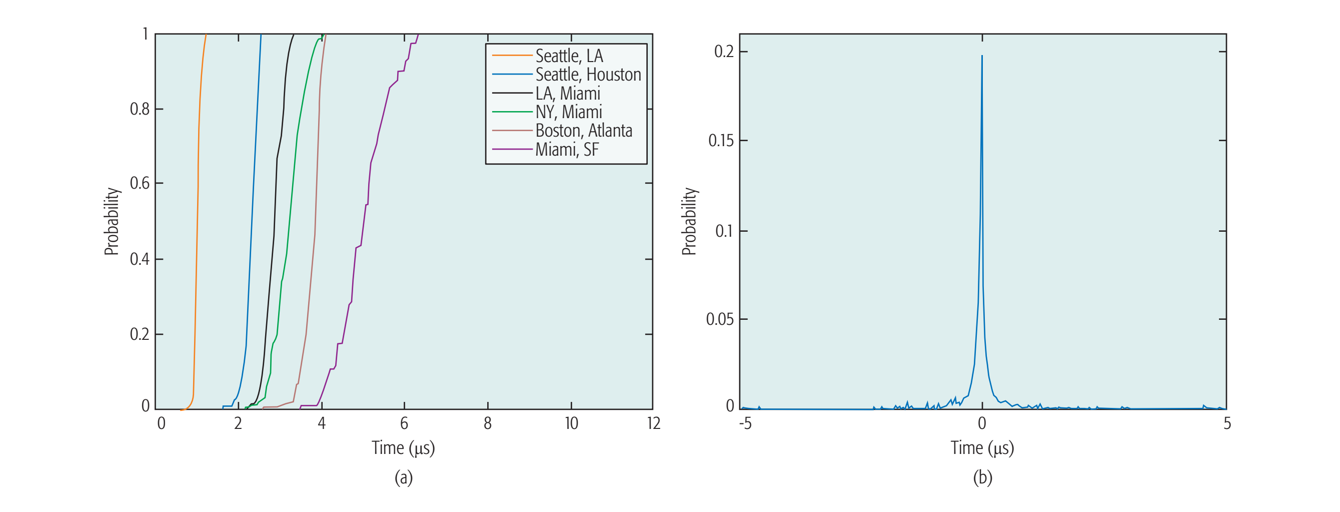

Figure 4. a) End-to-end queueing delay CDF for selected flows in the USA backbone; b) jitter distribution for all flows in the U.S. backbone.

bled at the output side of the switch. In an $N \times N$ switch, each IP has N virtual output queues (VOQs), where VOQ(i,j) stores data which arrives at IP(i) and is going to OP(j). Each VOQ in Fig. 3 supports two prioritized traffic classes, the deterministic and best effort classes.

However, a VOQ can be partitioned to support many prioritized traffic classes, including the three existing DiffServ traffic classes, expedited forwarding (EF), assured forwarding (AF), and DE), and a new deterministic class. Another new traffic class can also be created to handle short TCP/IP control packets (i.e., TCP acknowledgment [ACK] packets and socket open/close connection packets) with expedited guaranteed rate (GR) service.

Let each $N \times N$ CIOQ or CIXOQ switch have an $N \times N$ matrix of guaranteed traffic rates to be supported between the input and output ports. Reference [4] presents a very fast recursive scheduling algorithm, which can schedule the transmission of packets through a CIOQ switch with near-minimal delay and jitter. The algorithm mathematically recursively decomposes the $N \times N$ matrix of guaranteed traffic rates to achieve a very low-jitter transmission schedule with 100 percent throughput.

When the XQs are added to the CIOQ switch, as shown in Fig. 3, the scheduling algorithm is simplified. Each row of the $N \times N$ traffic matrix can be processed in isolation to compute a TX-Schedule for each IP. (Each row of the $N \times N$ matrix is a $1 \times N$ vector, which can be processed using the recursive scheduling algorithms in [4, 14].) At each IP, the TX-Schedule identifies a VOQ to be serviced for each time slot of a scheduling frame. The TX-Schedule provides each IP with a guaranteed rate of transmission, from the VOQs into the XQs of the switch. Each column of the $N \times N$ traffic matrix can also be processed to compute an RX-Schedule for each OP. The RX-Schedule specifies the XQ to be serviced in each column of the XQ switch for each time slot of a scheduling frame. The RX-Schedule provides each OP with a guaranteed rate of reception from the XQs into the OQs of the switch. When an IP receives service in a time slot, an *Input-Class-Schedule* can specify the traffic class or traffic flow to be serviced. At the OPs, once packets are reassembled, an optional *Output-Class-Schedule* can specify the traffic class or traffic flow to be serviced.

The Deterministic Schedules

The switch in Fig. 3 can reserve time slots for the transmissions of every deterministic traffic flow in a scheduling frame with F time slots. A scheduling frame length of $F = 1024$ can allocate bandwidth in increments of 0.1 percent of the line rate. Using a 400 Gb/s line rate and $F = 1024$, each time slot reservation will reserve 400 Mb/s. Given a traffic rate matrix, the schedules can be computed in microseconds. The traffic demands for deterministic flows change relatively slowly, over seconds or minutes, and hence the schedules can be computed once and stored in lookup tables and reused until the traffic demands change. The shaded boxes in Fig. 3 represent lookup tables. The routers do not need to be synchronized to microseconds of accuracy, as these schedules can be circularly rotated by arbitrary amounts and still retain the deterministic delay and jitter bounds. This ability to circularly rotate schedules is very important, since all the routers or switches in the U.S. core network in Fig. 2 need not be synchronized.

A Deterministic U.S. IIoT

Figure 2 illustrates a deterministic U.S. IIoT core network, with 26 nodes (cities) and 86 edges. The bold lines represent optical fiber links between cities. The dotted lines represent congestion-free deterministic transport connections (DTCs) between cities.

Our SDN control plane can program many virtual networks (VNs) into layer 3, as shown in Fig. 2. A VN is composed of many virtual links (VLs), where each VL represents a DTC, which passes through many routers. A router views a DTC as a congestion-free one-hop VL between remote cities, as shown by the dotted lines in Fig. 2. Our SDN control plane can configure the deterministic connections in layer 3 by configuring each router with several deterministic forwarding schedules. Packets of a DTC will be forwarded along a fixed path of Internet routers using these deterministic forwarding schedules, resulting in near-minimal buffer sizes and queueing latencies. Our SDN control plane uses a Max-Flow Min-Cost routing algorithm [15], without relying on sub-optimal best effort IP routing algorithms. It can create single-path or multi-path DTCs, with redundancy for improved reliability.

Our SDN control plane can also embed many VNs and VLs into an optional layer 2 underlay network of simple packet switches called the optical transport network (OTN), as shown in the bottom part of Fig. 1. Each VL in layer 2 can bypass several IP routers in layer 3, which will significantly improve energy efficiency. Current IP routers consume between 5 and 10 nJoules per bit transmitted, while state-of-the-art layer 2 packet switches consume about 250 pJoules per bit, resulting in an energy savings of a factor of about 30. The use of deterministic connections in layers 2 and 3 can help meet the aggressive energy efficiency targets specified by the Greentouch consortium (www.greentouch.org).

In Fig. 2, our SDN control plane programmed 300 VLs into the IIoT. Six cities selected at random (Seattle, Los Angeles, Denver, Chicago, Boston, and Miami) each have 50 VLs, with 25 VLs going to/from the other cities. Each of these six cities can reach any other city over a one-hop VL. (In Fig. 2, it is straightforward to embed a fully connected network where every pair of cities is interconnected with two VLs.) An IP router can also use VLs in its Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP) routing algorithms to support best effort traffic. These routing algorithms often minimize the number of hops, and they can be modified to use the VLs, which are viewed as one-hop logical connections between cities.

Experimental Results with 92 Percent Loads

In our tests, a scheduling frame with 1024 time slots was used. Each time slot was sufficient to transmit a maximum-size IP packet over an edge. Assuming 400 Gb/s edges and 1500-byte IP packets, a time slot consists of 30 ns. (A 400 Gb/s edge may consist of 4 parallel 100 Gb/s channels, in which case a time slot consists of 120 ns).

The IIoT network performance is deterministic, and was determined using three methods, which were all in agreement;

1. Reference [4] presents theoretical bounds on the end-to-end latencies and jitter.
2. A software simulator was developed to simulate the deterministic system.
3. An FPGA hardware testbed was developed where 26 simple routers were synthesized onto an Altera FPGA, and the performance was measured in hardware.

The hardware testbed can transmit packets at a rate exceeding 400 million packets/s. The hardware testbed and software simulator yield identical deterministic results.

Figure 4a illustrates the cumulative distribution function (CDF) of the end-to-end queueing delay between several cities, expressed in time slots. This figure does not include the fiber latency. The queueing delays in Fig. 4a are all less than 10 μs. Using standard single-mode fiber, the speed of light is about 200 km/ms. Consider the VLs between Los Angeles and Miami. The length of the fiber between these cities is at least 3800 km, depending on the physical path. The fiber latency is therefore 19 ms. The end-to-end queueing delay along the VL (≤ 10 μs) is over 1000 times smaller than the end-to-end fiber delay (≥ 19 ms).

Figure 4b illustrates the probability distribution of the jitter of the packets leaving a VL (averaged over all VLs in the U.S. network). The jitter is defined as the time difference of two consecutive departing packets in a given VL minus the ideal time between packets in the VL. According to Fig. 4b, most packets are delivered with a jitter ≤ 1 μs. According to theory, given a VL with a provisioned rate of 10 Gb/s and maximum-size IP packets, the maximum jitter is about 1.2 μs [4]. These jitter times, measured in microseconds, are exceptionally small when compared to the end-to-end fiber delays in the U.S. backbone network, measured in milliseconds.

According to our testbed the switch at Chicago buffers less than 50 packets, even at 93 percent average link loads. A BE-IoT router at Chicago with 400 Gb/s links would have a worst case buffer size of about 32 million packets [4, 7]. The use of deterministic packet switching, combined with our very low-jitter scheduling algorithm, has reduced the worst case buffer sizes by a factor exceeding 100,000 times.

Large-Scale Video Distribution

In this section, we explore large-scale video distribution over the future deterministic IIoT using simulations. Assume a Netflix data center in Chicago distributes video to several cities. Our SDN control plane can program a VN into the IIoT to support video distribution, with VLs from Chicago, to all other cities, as shown in Fig. 2. Netflix has about 35 million subscribers in the United States, and alone accounts for about 33 percent of the U.S. download bandwidth in peak hours. Each VL will typically carry between 1000 and 100,000 video streams, depending on the time of day. The provisioned rate of the VLs can be updated by an autonomic controller in the SDN control plane every 15 minutes (or as needed).

Video encoders typically use the standard three-level group of pictures (GOP) format. Each GOP consists of a large independent (I) frame, followed by several optional smaller predictive (P) frames, where several small bi-predictive (B) frames may exist between the P frames.

Figure 5a explores the aggregation of multiple low-bit-rate single-layer scalable video coding (SVC) video streams for mobile devices (i.e., tablets and smartphones). No buffering to smooth the traffic is assumed in Fig. 5. A single-layer SVC video stream called "Gandhi" is used, with a G16B15 GOP format. The G16B15 GOP format has one independent I frame followed by 15 smaller B frames. It has 53,968 frames, with a screen size of 352 × 288 pixels, a rate of 7.5 frames/s, with an average bit rate of 18.5 kb/s. To generate multiple video streams for aggregation, the same video stream was circularly rotated by a random amount before aggregation. Referring to the top row in Fig. 5a, the single video stream

Netflix has about 35 million subscribers in the USA, and alone accounts for about 33 percent of the USA download bandwidth in peak hours. Each VL will typically carry between 1,000 and 100,000 video streams, depending upon the time of day. The provisioned rate of the VLs can be updated by an autonomic controller in the SDN control-plane every 15 minutes (or as needed).

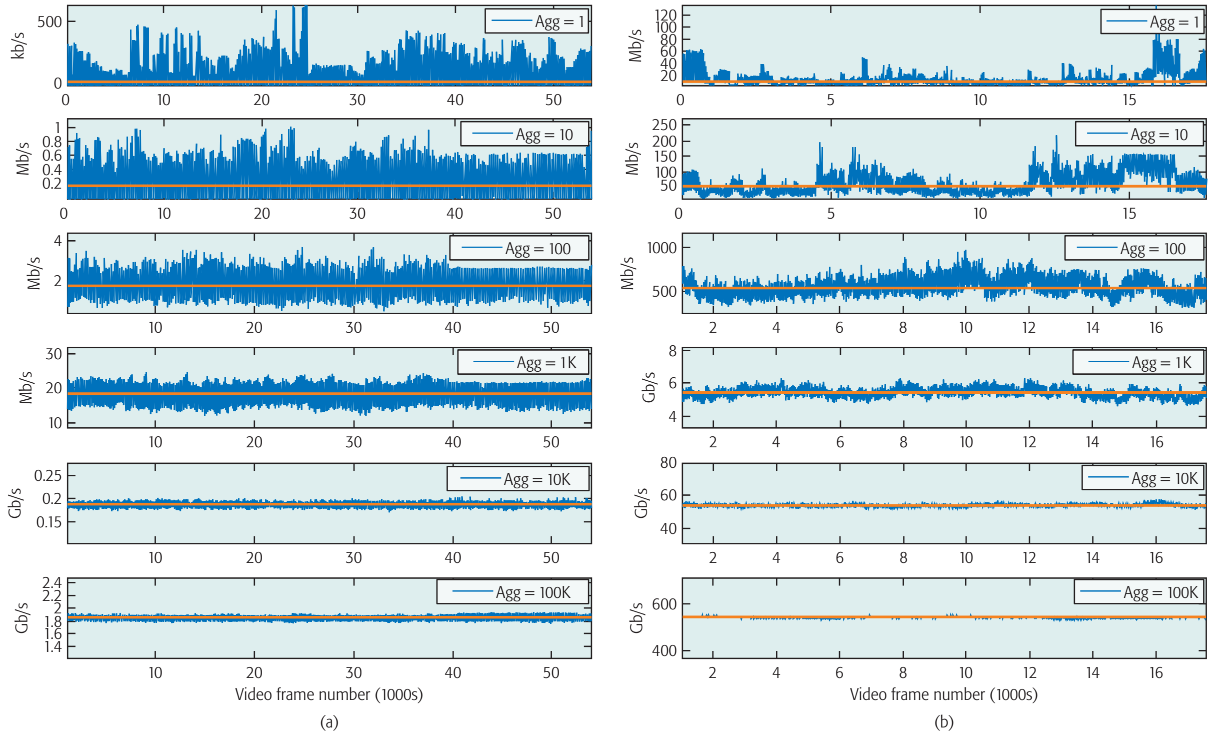

Figure 5. Instantaneous bandwidth vs. video frame number, for various degrees of aggregation: a) SVC video "Gandhi"; b) V9P 4K UHD video "Tears of Steel."

is quite bursty, with a mean bit rate of 18.5 kb/s and a peak bit rate of about 500 kb/s. When 10K streams are aggregated, the mean bit rate is 185 Mb/s and the peak bit rate is about 190 Mb/s, a considerable reduction in burstiness. We have aggregated several other SVC videos and observed the same behavior. A VL can transport thousands of SVC video streams with very high link utilizations, given the reduction in burstiness that occurs with aggregation.

Figure 5b explores the aggregation of multiple ultra high definition (UHD) 4K video streams for home TV. There are several encoders for UHD video, including the H.264, H.265, and VP9 encoders. A single video, the UHD VP9 4K "Tears of Steel" video, is used, with a GOP format of G24B0. It has 17,952 frames, with a screen size of 4096 × 1744 pixels, a rate of 24 frames/s, and an average bit rate of 5.414 Mb/s. To generate multiple video streams for aggregation, the same video stream is circularly rotated by a random amount. Referring to the top row of Fig. 5b, the single video stream is quite bursty, with a mean bit rate of 5.4 Mb/s and a peak bit rate of about 120 Mb/s. When 10K streams are aggregated, the mean bit rate is 54 Gb/s and the peak bit rate is about 56 Gb/s, a considerable reduction in burstiness. A VL can transport thousands of UHD video streams with very high link utilization, given the reduction in burstiness that occurs with aggregation.

The northbound traffic leaving a data center going to a remote city represents the aggregation of thousands of video streams. A token-bucket-based video shaper queue (VSQ) can be used at each data center to further smooth an aggregated stream before transmission over a VL. Our simulations indicate that an aggregated stream of 1000 videos (or more) can be delivered with queueing delays in the VSQ of ≤ 2–4 ms, with link utilizations of 95 percent [15]. In other words, significant overprovisioning is not needed. The southbound traffic arriving at a destination data center from the core network represents the aggregation of thousands of video streams. A video playback queue (VPQ) can be used to demultiplex the smoothed aggregated stream into multiple bursty video streams for distribution over a local area network. This playback queue will incur a similar small delay of typically 2–4 ms [15]. In the continental U.S. network shown in Fig. 2, the queueing delays in the VSQ and VPQ are much smaller than the fiber latencies.

The IETF has ruled out the use of overprovisioning to support deterministic services, and states that link utilizations of at least 50 percent should be supported for deterministic traffic in the future Internet [5, 6]. According to 2013 and 2014 annual reports, the annual sales of best effort hardware (routers, switches, wireless nodes) from Cisco, Huawei, Ericsson, and Alcatel-Lucent can be estimated at US$22, US$23, US$14.2, and US$15 billion, respectively, for a total of US$74 billion annually. Assuming a 50 percent link utilization, half of this annual hardware expenditure is effectively unused, and the unnecessary capital costs of underutilized networks can reach US$37 billion annually. Hence, it is desirable to achieve higher utilizations, well above 50 percent and approaching 100 percent, for deterministic traffic. This article demonstrates the technologies to achieve up to 100 percent utilization for deterministic traffic flows, using simple low-cost CIOQ or CIXOQ switches, which can lower the excess capital costs and energy costs of a future deterministic core network significantly.

Conclusion

The Internet network has used an inefficient best effort communications paradigm for the last 40 years, incurring excessive delays, capital costs, and energy costs. This article proposes a deterministic Industrial Internet of Things core network consisting of many simple deterministic packet switches controlled by an SDN control plane. Our SDN control plane can program thousands of deterministic virtual networks into the core network to provide each consumer service with its own dedicated congestion-free VN with exceptionally low latency and jitter. Highly aggregated video streams can be delivered over the continental United States with very low end-to-end latency determined by the speed of light in fiber, with jitters less than 10 μs, and with up to 100 percent link utilizations. By achieving 100 percent link utilizations rather than the 50 percent targeted by the IETF, the proposed deterministic network can save potentially US$37 billion in capital costs annually. A speed-of-light deterministic core network can have a profound impact on virtually all consumer services such as multimedia distribution, e-Commerce, and cloud computing or gaming. It can also pay for itself quickly, due to its significantly improved utilization and energy efficiency. The deterministic technologies proposed in this article can also provide deterministic services in metro area networks, data center networks, and supercomputer networks. We believe that a future converged deterministic Internet of Things that combines the best effort and deterministic communications paradigms can fundamentally transform computing and consumer services in the 21st century.

A speed-of-light deterministic core network can have a profound impact on virtually all consumer services such as multimedia distribution, e-Commerce, and cloud computing or gaming. It can also pay for itself quickly, due to its significantly improved utilization and energy efficiency.

References

[1] M. Ford, "Workshop Report: Reducing Internet Latency 2013," *ACM SIGCOMM CCR*, vol. 44, no. 2, Apr. 2014, pp. 80–86.

[2] A. Singla *et al.*, "The Internet at The Speed of Light," *ACM Hotnets 2014*, Oct. 2014, Los Angeles, CA, pp. 1–7.

[3] G. Fettweis *et al.*, "The Tactile Internet," ITU-T Technology Watch Report, Aug. 2014, pp. 1–24.

[4] T.H. Szymanski, "An Ultra Low Latency Guaranteed-Rate Internet for Cloud Services," *IEEE Trans. Networking*, vol. 24, no. 1, Feb. 2016, pp. 123–36.

[5] N. Finn and P. Thubert, "Deterministic Networking Problem Statement (04)," IETF Internet Draft, Standards Track, Oct. 19, 2015, pp. 1–17.

[6] S. Shah and P. Thubert, "Deterministic Forwarding PHB (04)," IETF Internet Draft, Aug. 30, 2015, pp. 1–8.

[7] S. Iyer, R. R. Kompella, and N. Mckeown, "Designing Packet Buffers for Router Linecards," *IEEE Trans. Networking*, vol. 16, no. 3, June 2008, pp. 705–17.

[8] V. Anantharam *et al.*, "Achieving 100% Throughput in an Input Queued Switch," *IEEE Trans. Commun.*, vol. 47, no. 8, 1999, pp. 1260–67.

[9] W.J. Chen, C-S. Chang, and H-Y. Huang, "Birkhoff-von Neumann Input Buffered Crossbar Switches for Guaranteed-Rate Services," *IEEE Trans. Commun.*, vol. 49, no. 7, July 2001, pp. 1145–47.

[10] I. Keslassy *et al.*, "On Guaranteed Smooth Scheduling for Input-Queued Switches," *IEEE/ACM Trans. Networking*, vol. 13, no. 6, Dec. 2005, pp. 1364–75.

[11] IEEE 802 Tutorial, "Deterministic Ethernet: 802.1 Standards for Real-Time Process Control, Industrial Automation, and Vehicular Networks," Nov. 12, 2012, pp. 1–72.

[12] D. Dujovne *et al.*, "6TiSCH: Deterministic IP-Enabled Industrial Internet (of Things)," *IEEE Commun. Mag.*, vol. 52, no. 12, Dec. 2014, pp. 36–41.

[13] P. Wetterwald and J. Raymond, "Deterministic Networking Utilities Requirements," IETF Internet Draft, June 30, 2015, pp. 1–26.

[14] T. H. Szymanski, "Crossbar Switch and Recursive Scheduling," U.S. Patent 9042380B2, issued May 26, 2015, pp. 1–36.

[15] T. H. Szymanski, "Max-Flow Min-Cost Routing in a Future Internet with Improved QoS Guarantees," *IEEE Trans. Commun.*, vol. 61, no. 4, Apr. 2013, pp. 1485–97.

Biography

Ted H. Szymanski (teds@mcmaster.ca) completed his Ph.D. degree at the University of Toronto. From 2001 to 2011, he held the Bell Canada Chair in Data Communications at McMaster University. Previously, he was a professor at Columbia University and its Center for Telecommunications Research, and McGill University and the Canadian Institute for Telecommunications Research. He participated in a 10-year research program within the Networks of Centers of Excellence of Canada, which demonstrated a free-space intelligent optical backplane using photonic packet-switches with about 1000 optical channels. Contributors included Nortel Networks (Ericsson), Newbridge Networks (Alcatel), Lockheed-Martin/Sanders, and McGill, McMaster, Toronto, and Heriot-Watt Universities. His group also demonstrated the first FPGA with optical IO, using the U.S. DARPA/Lucent/Coop smart-pixel foundry service. His interests include security, energy efficiency, deterministic communications, and the industrial Internet of Things.