Analysis of TSN for Industrial Automation based on Network Calculus

Jiayi Zhang, Lihao Chen, Tongtong Wang, Xinyuan Wang
dept. IP Technology Research
Huawei Technologies Co., Ltd.
Beijing, China
{zhangjiayi11, lihao.chen, tongtong.wang, wangxinyuan}@huawei.com

***Abstract*—Time-Sensitive Networking (TSN) Ethernet is becoming a primary industrial networking technology, since it provides bounded latency capability with broad reaching ecosystem, supporting both real-time communications and non-time-critical communications in a factory. In TSN, QoS-based shaping is a core technology to provide bounded latency. Network calculus is a theory which aims at worst-case performance analysis in the network. In the paper, latency upper bound of TSN for industrial automation scenario is analyzed based on Network Calculus. Credit-based Shaping (CBS) combined with Strict Priority (SP) is used as the specific QoS-based shaping mechanism of TSN. Relating factors that influence the latency upper bound of TSN are also studied.**

***Index Terms*—Time-Sensitive Networking (TSN), Network Calculus (NC), Industrial Automation network, Bounded Latency, Quality of Service (QoS)**

I. Introduction

Recently, the discussion of TSN has become more and more enthusiastic in industries. People began to think that TSN might be the evolution direction for future industrial automation networks. The traditional Ethernet is known as "Best Effort" (BE) which does not provide bounded latency guarantee. Various Industrial Ethernet techniques (such as EtherNet/IP, PROFINET, EtherCAT, POWERLINK) have been developed to meet different communication requirements, but the openness and interoperability of these Industrial Ethernet techniques are not as good as Ethernet, and some related proprietary technologies bring high costs in hardware and software. TSN techniques perfectly combine the advantages of traditional Ethernet and Industrial Ethernet: providing good openness and interoperability, as well as bounded latency and bandwidth guarantee.

TSN techniques consist of multiple standards defined in IEEE 802.1 group. Some standards have already been released and others are still ongoing. Generally, by using either time-based scheduling or QoS-based shaping, TSN can provide better Quality of Service (QoS) on bounded latency for different traffic types. Time-based scheduling like Time-Aware Shaping (TAS) defined in IEEE 802.1Qbv, relying on network-wide time synchronization, can ensure bounded queuing delay and bounded delay jitter. However, network-wide time slot planning problem should be NP-hard, when network scales up and the number of streams is large. The other approach, QoS-based shaping, runs asynchronously over network, like Credit-Based Shaping (CBS) defined in IEEE 802.1Qav, also efficiently guarantee end-to-end latency in Audio-Video Bridging (AVB) networks.

Network Calculus (NC) is a theory for deterministic queuing delay analysis in networked systems. It can derive relationship between single-node delay and parameters of shapers and schedulers to provide a guideline for network configuration and resource reservation, also it helps to calculate end-to-end latency for each flow or traffic class to check whether delay or buffering requirements are met. In this paper, latency upper bound of TSN for industrial automation scenario is analyzed based on Network Calculus.

Network calculus theory and results can be referred in [1], [2]. In [3], delay bound for CBS is derived. In [4], it compares many NC algorithms to calculate delay bound for flows in network, which requires trade-off between tightness and complexity. In [5], [6], a method with per-class shaping is proposed, which can be used to calculate delay bound for TSN, and is referred in IETF draft on calculating bounded latency for deterministic network (DetNet) [7]. In terms of application, NC has been successfully used in evaluation and verification of Avionics Full-Duplex Switched Ethernet (AFDX) [8], [9]. For application in Industrial Automation network, [10] utilizes NC to analyze switched Ethernet for industrial performance requirement. In [11], a system is proposed that uses NC calculated delay to configure shapers in PROFINET. In this work, we would like to make use of the latest proposed NC method in [5] to analyze latency bound for Industrial Automation network that using TSN QoS-based shaping.

The intention of this paper is to design a model (including network topology, flows, the characteristic of these flows, and the method of using TSN technologies) based on real industrial automation use cases, and to make use of network calculus and to calculate the upper bound on latency (i.e., the worst-case latency) of the flows, and to analysis the influencing factors by comparing the simulation results. By demonstrating the latency upper-bound analysis, results show that using QoS-based method can meet transmitting latency requirement in industrial automation network, which indicates an alternate solution to provide latency guarantee besides time-based scheduling.

The paper is organized as following. Section I is an introduction for background, motivation and related work. Section

978-1-7281-0303-7/19/$31.00 ©2019 IEEE

II introduces basic characteristics of TSN and background of industrial networks. The section III discusses fundamental concepts of network calculus and mathematical model for latency bound calculation. Section IV analyzes specific industrial scenarios with network calculus and simulation results. Section V gives conclusion and future work.

II. Description of TSN for Industrial Automation

A. Network Architecture of Industrial Automation

The industrial automation network has a hierarchical structure, as shown in Fig. 1 [12]. On the Operational Technology (OT) side, devices are connected by the network within each functional block and blocks are connected via routers, switches, or PLCs (Programmable Logic Controller, may act as a gateway).

The implementation of industrial automation network needs to satisfy specific requirements, including network requirement (e.g., bounded latency and reliability), working environment requirements (e.g., temperature, humidity, and electromagnetic compatibility), and security requirements. Therefore, Industrial Ethernet protocols (or Real-Time Ethernet, Ethernet-based fieldbus) were developed and used. Some of these protocols redefined the Ethernet MAC layer.[1] In addition, with the advent of era of intelligent manufacturing, new challenges are rising for industrial automation and its network, such as flexible data access, easy or zero network configuration, etc.

B. Typical Traffic Types in Industrial Automation

There are many different traffic types in the industrial automation network. The behavior as well as requirement of these traffic types differs. One possible classification is shown in Tab. I [13].

C. Basic TSN Features for Frame Queuing and Forwarding

Formerly, the goal of IEEE 802.1 is to specify how the MAC service is supported by bridged networks, the principles of operation of those networks, and the operation of MAC bridges and VLAN bridges. As there is neither latency nor bandwidth guarantee provided, this kind of network is called as "best-effort" network.

IEEE 802.1 TSN Task Group is started in 2012 by renaming the AVB Task Group. TSN task group defines a set of technical standard, providing new queuing and forwarding mechanisms. IEEE 802.1Qav defines Credit-Based Shaping (CBS), which provides bounded latency for specific traffic classes by bandwidth resource reservation. IEEE 802.1Qbv defines Time-Aware Shaping (TAS), which enables the queue gate state control (i.e., open or close) based on timeslot. IEEE 802.1Qch defines Cyclic Queuing and Forwarding, which combines Per-Stream Filtering and Policing (PSFP) with Time-Aware Shaping. IEEE 802.1Qcr defines Asynchronous Traffic Shaping (ATS), which forwards frames based on per-flow state. All these mechanisms are potential candidates to provide better queuing and forwarding services than "best-effort" in industrial networks. IEC/IEEE joint project P60802 is defining a TSN profile standard for industrial automation.

D. TSN for Industrial Automation Examples

TSN will facilitate future Industrial Automation network by providing an open, interoperable, and standardized layer-2 network with bounded latency, bandwidth guarantee and high reliability. Different types of traffic may be transmitted together on one physical link, while all latency requirements are met. The case in Fig. 2 is designed to support up to 9 different types of traffic sharing the red-colored backbone links. [14]

III. Network Calculus

A. Basic Concepts

To provide deterministic latency guarantee to different type of services in TSN Industrial network, a method is needed to calculate the worst-case (upper) bound on latency of flows. The per-hop delay bound of a TSN relay node is influenced by hardware implementation, link rate and queuing mechanism. Here, we do not consider preemption and implementation related factors, but focusing on the queuing and forwarding delays. Network Calculus (NC) is used to analyze worst-case delay and backlog of the network [2].

A TSN relay node can be abstracted as Fig. 3 shown. Considering a flow at output port of the TSN node, $R(t)$ is the cumulative arrival data until time t. For any time period t, the incremental arrival data is constrained by an arrival curve $\alpha(t)$, that

$$R(s+t) - R(s) \leq \alpha(t), \quad \forall\ s \geq 0,\ t \geq 0 \tag{1}$$

In TSN, traffic of input flow is constrained by an arrival curve, which can be given by traffic specification (TSpec).

The queuing and forwarding method used in this node can be characterized as service curve $\beta(t)$, which describes the minimal service capability. The service curve $\beta(t)$ is defined as below, if the accumulative input data $R(t)$ and output data $R^*(t)$ of the node satisfies

$$R^*(t) \geq R \otimes \beta(t) = \inf_s \{R(s) + \beta(t-s)\}, \quad \forall\ 0 \leq s \leq t \tag{2}$$

where $\inf_s\{X\}$ calculates the infimum, or greatest lower bound of X with respect to s, and operator $\otimes$ calculates convolution defined in min-plus algebra [2].

The output flow $R^*(t)$ is constrained by output bound α^*, which can be used as arrival curve of next hop.

$$\alpha^*(t) = \sup_u \{a(t+u) - b(u)\}, \quad \forall\ u \geq 0 \tag{3}$$

By calculating the maximum vertical deviation between arrival curve $\alpha(t)$ and service curve $\beta(t)$, one can obtain the backlog bound as

$$B = \sup_t \{\alpha(t) - \beta(t)\} \tag{4}$$

[1]In the 7-layer / 5-layer OSI model, the second layer is the data link layer and MAC is its sub-layer. Ethernet is one of the most widely used MAC protocols.

[2]Network Calculus, in the perspective of cumulative data, is based on min-plus algebra $\{\mathbb{R} \cup \{+\infty\}, \min, +\}$

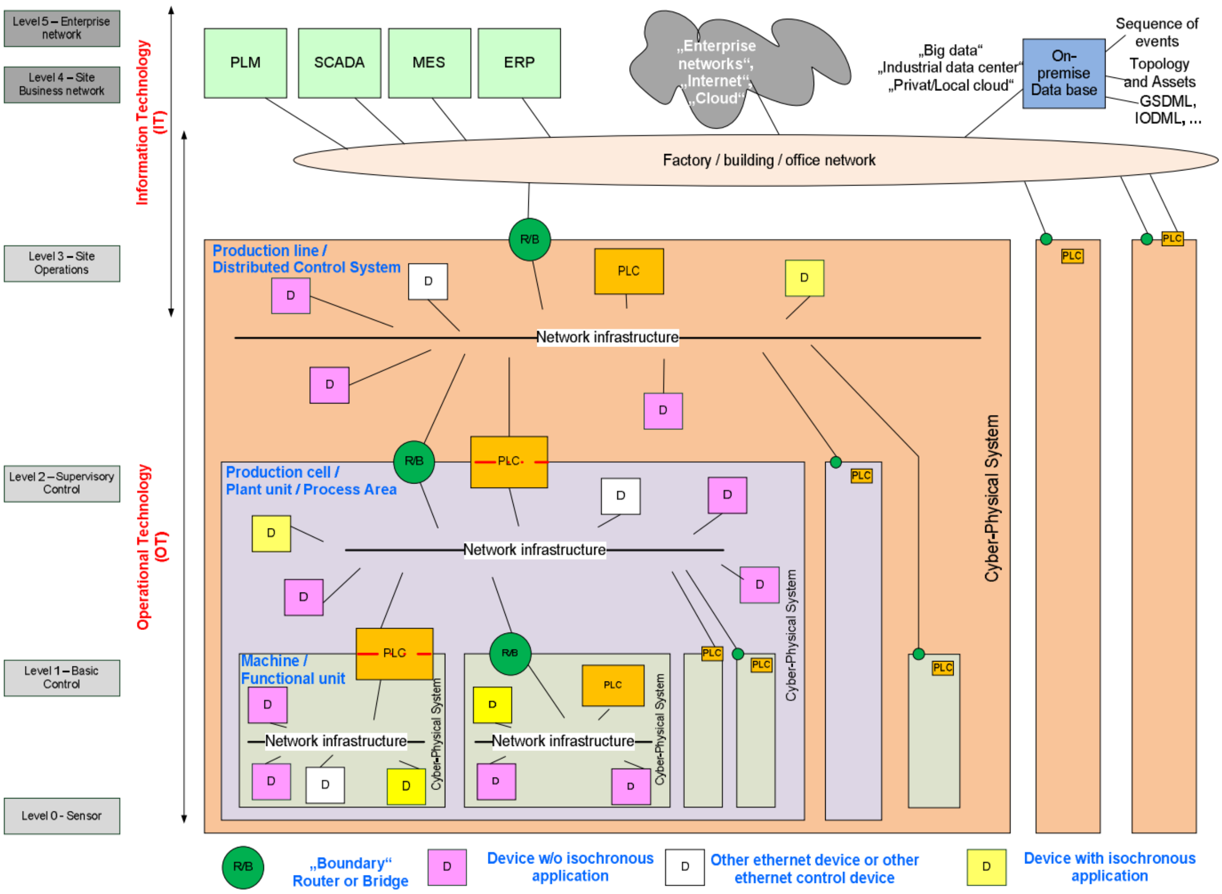

Fig. 1. Hierarchical structure of Industrial Automation [12].

TABLE I
A CLASSIFICATION OF TRAFFIC TYPES [13]

Types	Periodicity	Typical period	Synchronized to network	Data delivery guarantee	Tolerance to interference	Tolerance to loss	Typical application data size	Criticality
Isochronous	Periodic	<2ms	Yes	Deadline	0	None	Fixed: 30-100 Bytes	High
Cyclic	Periodic	2-20ms	No	Latency	⩽ latency	1-4 Frames	Fixed: 50-1000 Bytes	High
Events	Sporadic	n.a.	No	Latency	n.a.	Yes	Variable: 100-1500 Bytes	High
Network Control	Periodic	50ms-1s	No	Bandwidth	Yes	Yes	Variable: 50-500 Bytes	High
Config & Diagnostics	Sporadic	n.a.	No	Bandwidth	n.a.	Yes	Variable: 500-1500 Bytes	Medium
Best Effort	Sporadic	n.a.	No	None	n.a.	Yes	Variable: 30-1500 Bytes	Low
Video	Periodic	Frame Rate	No	Latency	n.a.	Yes	Variable: 1000-1500 Bytes	Low
Audio/Voice	Periodic	Sampling Rate	No	Latency	n.a.	Yes	Variable: 1000-1500 Bytes	Low

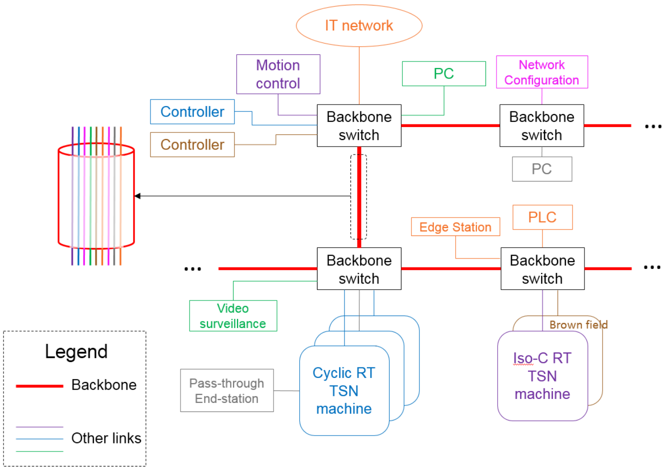

Fig. 2. An example of multi-traffic transmission on one backbone link

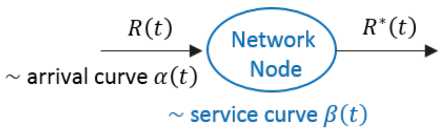

Fig. 3. Two-port network model of a TSN relay node

where $\sup_t\{X\}$ calculates the supremum, or minimum upper bound of X with respect to t. The buffer space at a node should be no less than the backlog bound B, in order to guarantee zero congestion loss.

By calculating the maximum horizontal deviation between arrival curve $\alpha(t)$ and service curve $\beta(t)$, one can obtain the delay bound as below

$$D = \sup_s\{\inf_t\{t \geq 0 \mid \alpha(s) \leq \beta(s+t)\}\} \tag{5}$$

Figure 4 shows an example of arrival curve, service curve, backlog bound, and delay bound.

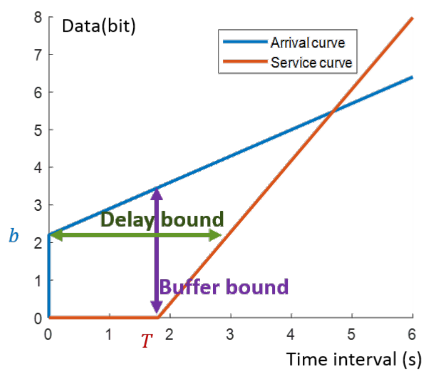

Fig. 4. Computation of backlog bound and delay bound.

Note that in the formula of delay bound (5), the service curve $\beta(t)$ may reflect either per-hop forwarding method that a network node offers to a flow, or concatenation of service curves on multiple nodes in a path. Consider a flow traverses two nodes, with service curve $\beta_1(t)$ and $\beta_2(t)$ sequentially. Then concatenation of these two nodes offers a service curve $\tilde{\beta}(t)$ as below

$$\tilde{\beta}(t) = \inf_s\{\beta_1(s) + \beta_2(t-s)\}, \forall\ 0 \leq s \leq t \tag{6}$$

By using arrival curve and end-to-end concatenated service curve, Formula (5) gives an end-to-end worst-case delay bound over a path.

In summary, network calculus demonstrates that, given input traffic constraint as arrival curve and forwarding method modeled as service curve, the worst-case delay bound can be obtained. This worst-case delay bound can be used as the latency guarantee for a deterministic service in the network.

B. *Per-hop Shaping Model*

In this subsection, we introduce a method to calculate worst-case delay bound for a flow in the network. Section III-A theoretically describes how to calculate upper-bound of end-to-end delay of a path, however, the calculation would be difficult or impossible when aggregation happens [16], because burstiness of a flow is influenced by the burstiness of other flows at the same output port (concurrent flows), and the increased burstiness happens at every hop along the path. For this problem, [5], [6] and IETF DetNet draft [7] proposed a method that reshape the traffic at every hop, which enables the calculation of worst-case delay bound.

The reshaping component is named as Interleaved Regulator (IR) [5], which is a per-class queuing method. Implementation of IR can be briefly described as below,

- Packets of flows from a same class enter a same FIFO queue of IR.
- Only packet at head of the FIFO queue is checked. This packet will be transmitted if it satisfies traffic constraint.

By using IR, traffic constraints can be guaranteed at each output port in the network to control burstiness caused by flow aggregation.

With IR, delay within a node includes two main parts: queuing and forwarding delay D_Q and regulating delay D_R [3]. It is proved in [5] that, IR does not increase end-to-end worst-case latency bound, so the latency bound can be calculated as the sum of per-hop queuing and forwarding delay.

In IEEE 802.1 AVB/TSN, Credit-Based Shaping (CBS) is a per-class shaping method, that flows belong to a same class will share a same queue. Two Stream Reservation (SR) traffic classes are defined: SR Class A and SR Class B, whose available transmission bandwidth is controlled by a credit function [17]. CBS is initially invented for audio and video steam transmission with bounded latency requirement. CBS can be combined with strict priority (SP) transmission selection, and used in many time-critical applications. A typical combination of CBS+SP traffic class is provided in Tab. II, that Priority

[3]In general, delay at a TSN node should also considers: 1) Output Delay, 2) link delay, 3) preemption delay, 4) processing delay. Details can be referred to [7].

1 (highest) is assigned to network control data traffic (CDT), Priority 4 (lowest) is assigned to best-effort (BE) or traditional Ethernet traffic, and Priority 2 and 3 are assigned for SR class A and B respectively.

TABLE II
TRAFFIC CLASS DEFINITION

Priority	Traffic Class	Traffic constraint
1	CDT	Token bucket
2	SR Class A	
3	SR Class B	
4	BE	

Assuming the arrival traffic for Priority 1, 2 and 3 is constrained by token bucket. Service curve of CBS for SR class A and B is given in [6] as

$$\beta^l(t) = R^l(t - T^l), \quad l \in \{\text{A, B}\} \tag{7}$$

where for SR Class A

$$R^A = \frac{I^A(c - r_h)}{c} \tag{8}$$

$$T^A = \frac{L^{\bar{A}} + b_h + \frac{r_h L^{\bar{h}}}{c}}{c - r_h} \tag{9}$$

for SR Class B,

$$R^B = \frac{I^B(c - r_h)}{c} \tag{10}$$

$$T^B = \frac{L^{BE} + L^A + \frac{L^{\bar{A}} I^A}{c - I^A} + b_h + \frac{r_h L^{\bar{h}}}{c}}{c - r_h} \tag{11}$$

where,

c	link rate
I^A, I^B	idle slope for SR Class A and SR Class B
b_f	burst of the current flow f
r_h, b_h	sum of rate, sum of burst for CDT traffic
$L^{\bar{h}}$	maximum packet length with priority lower than CDT
L^l	maximum packet length of flows with priority l *
$L^{\bar{l}}$	maximum packet length of flows with priority lower than l *

*$l \in \{A, B, BE\}$

For CDT traffic, service curve has similar rate-latency form as (7), whose parameters can be straight-forward derived as,

$$\beta^{CDT}(t) = R^{CDT}(t - T^{CDT}) \tag{12}$$

$$R^{CDT} = c \tag{13}$$

$$T^{CDT} = \frac{b_h - b_f + L^{\bar{h}}}{c} \tag{14}$$

Considering non-BE flows have token bucket constraint. The arrival curve for each CBS traffic class is given as

$$\alpha^l(t) = b^l + r^l t, \quad l \in \{A, B\} \tag{15}$$

where for SR Class A and B, burst b^l is the sum of bucket size for all flows with the same class traversing the same port. For CDT, arrival curve of a flow f would be $\alpha^{\text{CDT}}(t) = b_f + r_f t$.

According to network calculus, the delay bound for one hop (5) from node i to j can be specifically given as

$$D_{i,j} = T^l + \frac{b^l - L_{\min,f}}{R^l} + \frac{L_{\min,f}}{c} \tag{16}$$

where the last term represents output delay, $L_{min,f}$ is the minimum packet size for flow f. Then, end-to-end worst-case delay bound is the sum of per-hop delay bound along the path.

It should be noted that, the shaping behaviour of IR may increase delay to a packet, however, it is proved that IR does not increase worst-case latency bound in a feed-forwarding FIFO network [5]. This property is helpful to give analytical latency guarantee to time-sensitive flows.

IV. SIMULATION

A. Network Description

In this section, we use IR-based model to calculate latency bound for an example industrial automation network. The topology of the network used for simulation is shown as Fig. 5. Flow information is described in Tab. III, in which 4 industrial automation traffic types are mapped to specific traffic classes and priorities at the bridge. Priority 1 flows have the highest priority which are forwarded prior to all other flows. Priority 2 flows and priority 3 flows use Credit-Based Shaping of SR class A and SR class B respectively. Priority 4 flows, i.e. best-effort flows, have the lowest priority. Store-and-forward is used and preemption is not enabled for this simulation.

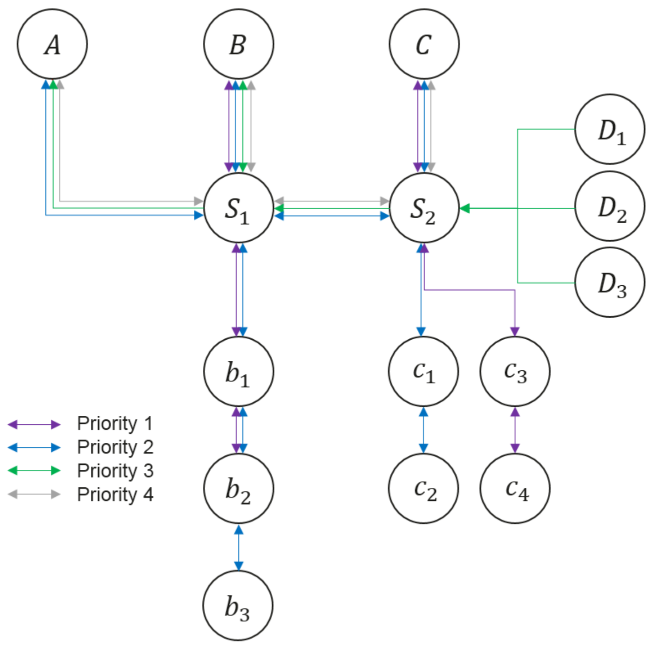

Fig. 5. Simulation topology

Simulation topology and flow settings can be related to the industrial automation network characteristics and typical traffic types. In Fig. 5, node A, B, C, D, S and node b, c map to different level of devices in the hierarchical structure of industrial automation network in Fig. 1. For example, in

TABLE III
FLOW DESCRIPTION

Flow path	Traffic Type	Forwarding	Priority
p_1: B, S_1, b_1 p_2: B, S_1, b_1, b_2 p_3: b_1, S_1, B p_4: b_2, b_1, S_1, B p_5: C, S_2, c_3 p_6: C, S_2, c_3, c_4 p_7: c_3, S_2, C p_8: c_4, c_3, S_2, C	Isochronous	CDT (SP)	1
p_9: B, $_1$, S_2, C p_{10}: C, S_2, S_1, B p_{11}: A, S_1, B p_{12}: B, S_1, A p_{13}: A, S_1, S_2, C p_{14}: C, S_2, S_1, A p_{15}: C, S_2, c_1 p_{16}: C, S_2, c_1, c_2 p_{17}: c_1, S_2, C p_{18}: c_2, c_1, S_2, C p_{19}: b_3, b_2, b_1, S_1, B p_{20}: B, S_1, b_1, b_2, b_3	Cyclic	SR Class A (CBS)	2
p_{21}: D_1, S_2, C p_{22}: D_2, S_2, S_1, A p_{23}: D_3, S_2, S_1, B	A/V	SR Class B (CBS)	3
p_{24}: A, S_1, B p_{25}: B, S_1, A p_{26}: A, S_1, S_2, C p_{27}: C, S_2, S_1, A	BE	BE (SP)	4

Fig. 5, node B and node C could be PLCs, node b_1, b_2, c_3, c_4 could be Drives, node b_3, c_1, c_2 could be IO devices, node D_1, D_2, D_3 could be sensors (e.g., surveillance cameras), node A could be a higher-level controller (e.g., a supervisory PLC or an industrial PC), and node S_1 and S_2 could be industrial switches. Node B connects and controls b_1, b_2, b_3, while node C connects and controls c_1, c_2, c_3, c_4. Control loops exist between PLC and Drives with Isochronous traffic flow communication, and between PLC and IO devices with Cyclic traffic flow communication. Cyclic traffic flows also exist between node A, B, and C. Audio/Video flows are sent from D_1, D_2, D_3 to A, B, C, respectively. Best-effort flows exist between A and B, A and C, e.g., firmware and software updating, programming distribution, log uploading.

B. Parameters

The network configurable parameters are,

1) Bandwidth per link $c_{i,j}$,
2) Percentage of bandwidth reserved for SR class A per output port $I^A_{i,j}$
3) Percentage of bandwidth reserved for SR class B per output port $I^B_{i,j}$.

where subscript i, j stand for the parameter for output port on node i towards node j. In the following simulation, we let $I^A_{i,j} = 50\% c_{ij}$, $I^B_{i,j} = 25\% c_{ij}$. Assume that the core link rate is 1 Gbps, between node S_1 and S_2, from S_1 to A, B, and from S_2 to C. The rate of other links is 100 Mbps.

Consider the flows with Priority 1, 2 and 3 have traffic constraints of token bucket, the parameters for every flow f are required,

4) Maximum packet size L^f_{max}, or maximum packet size of a batch of packets for per-batch packet forwarding,
5) Minimum packet size L^f_{min},
6) Maximum rate r^f. For periodic traffic, average rate could be defined by typical period T^f, as $r^f = L^f_{max}/T^f$

In the following simulation, flow parameters, or traffic specification (TSpec), are set referring to the typical flow parameters given in Tab. I. For Isochronous flows between node B, b_1 and b_2, i.e. on flow path $p_1 \sim p_4$, a sender transmits at most L_{max} =0.8kb data during $T = 2$ ms. For Isochronous flows between node C, c_1 and c_2, i.e. on flow path $p_5 \sim p_8$, a sender transmits at most $L_{max} = 0.8$ kb data during $T = 1$ ms. For Cyclic traffic, a sender transmits at most $L_{max} = 8$ kb data during $T = 10$ ms. Audio and Video (A/V) rate is no more than $r = 1$ Mbps, maximum packet size is $L_{max} = 12$ kb. For BE traffic, maximum packet size is also 12kb, without constraint for rate. The minimum packet size L_{min} for Priority 1 to 4 is 0.8kb, 0.4kb, 8kb and 0.24kb respectively.

With these parameters, by using delay bound calculating method (16), the theoretical worst-case delay upper-bound can be computed for every flow, in a network with CBS+SP forwarding.

C. Usage of Bandwidth

Consider two flow sets with different number of flows. Set 1 has 116 flows in total, while each flow path of priority 1 shown in Tab. III has 3 independent flows, and each flow path of priority 2, 3, 4 has 4, 4, 8 flows, respectively. Flow set 2 has doubled number of flows than set 1. Worst-case delay bound of the two flow sets are shown in Fig. 6, where one flow is shown for each path [4].

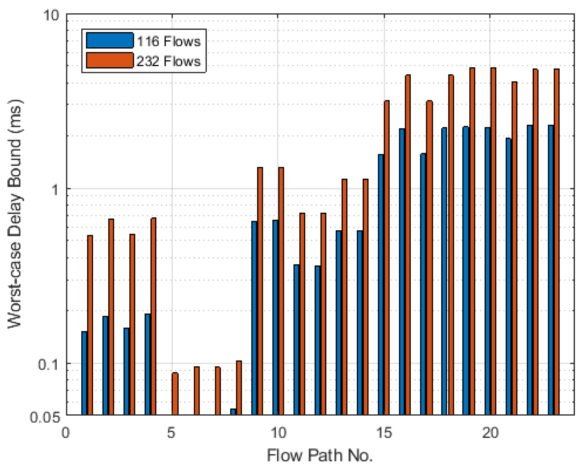

Fig. 6. Worst-case delay bound comparison between different number of flows.

Bandwidth usage $\gamma_{i,j}^{l}$ of output port of every traffic priority can be analyzed by $\gamma_{i,j}^{l} = \frac{\sum_f r_{i,j,f}}{c_{i,j}}$, which counts for all the flows f traversing the output port on node i towards node j, for priority $l \in$ {CDT, A, B, BE} separately. The bandwidth usage percentage of each priority on each port is shown in Fig. 7. The corresponding relationship between the port's number and the ports are shown in the Tab. IV, where a port i, j stands for the output port on node i towards node j.

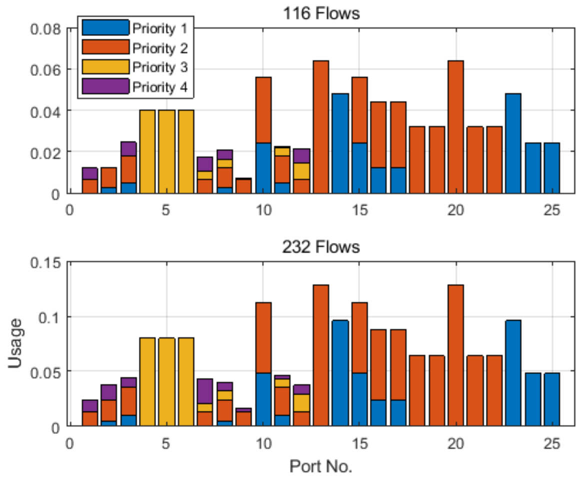

Fig. 7. Port usage comparison between different number of flows.

TABLE IV
OUTPUT PORT WITH PORT NUMBER

Port No.	Output Port						
1-6	A,S_1	B,S_1	C,S_2	D_1,S_2	D_2,S_2	D_3,S_2	
7-12	S_1,A	S_1,B	S_1,S_2	S_1,b_1	S_2,C	S_2,S_1	
13-18	S_2,c_1	S_2,c_3	b_1,S_1	b_1,b_2	b_2,b_1	b_2,b_3	
19-25	b_3,b_2	c_1,S_2	c_1,c_2	c_2,c_1	c_3,S_2	c_3,c_4	c_4,c_3

By using the network calculus analysing model, one can explicitly derive delay bound of each flow on a network with CBS+SP forwarding, as shown in Fig. 6. In this case, when the number of flows is 232, end-to-end worst-case delay bound is no more than 1 ms for Priority 1, and no more than 10 ms for Priority 2 and 3. With this analysis method, it can be checked whether adding more flows violates the traffic requirement in delay, as well as from the usage point of view, which port is more sensitive to addition of flows with specific class.

There is space for future research and analysis, e.g., will a better result be achievable if the bandwidth reservation percentage of each SR class are more properly configured, and what if the topology becomes more complex and the number of flow continues growing.

D. Introducing Offset to Periodic Traffic

To provide latency guarantee under QoS-based shaping network, the design phase claims that worst-case delay bound should not be greater than the data delivery guarantee of transmitting delay. According to network calculus, the worst-case bound indicates a minimum service and worst-case burstiness, that on an output port, all concurrent flows with higher or same priority than observed flow have maximum burstiness simultaneously, meanwhile packets of the observed flow stands at the end of the transmitting queue. Considering such a rarely-happened worst case, the calculated delay bound inherently leads to a pessimistic result and over-provisioning of network resource.

To reduce the pessimism and improve the network usage, we borrow the idea of sender scheduling, that introduces transmitting time offsets among senders. Preshaping is a similar method, which control transmitting pace based on E2E delay optimization [19]. By this means, the worst-case burstiness is dispersed while average bandwidth remains unchanged.

Consider the example in IV-A. Suppose each flow path of priority 1 has 10 flows, and each flow path of priority 2, 3, 4 has 8, 8, 10 flows, respectively. Basically, every senders operates synchronized and transmits packets at the same epoch and with the corresponding transmitting cycle. The worst-case delay bound of every flow is shown in Fig. 8 in the blue blocks [4]. A sending offset, with the length of half the sending cycle, can be imposed on half number of the flows which have the same sending cycle and the same path. The result is shown in Fig. 8 in the red blocks.

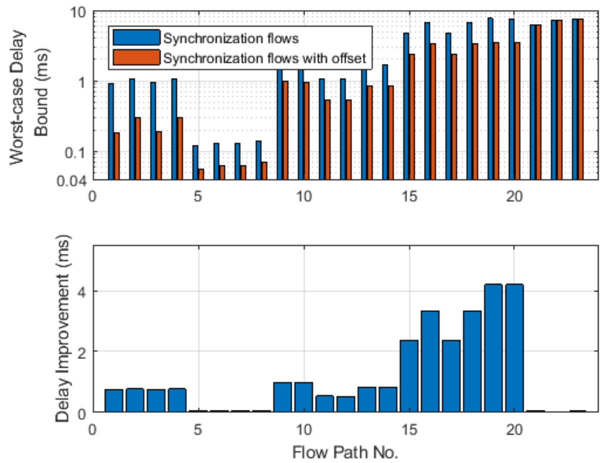

Fig. 8. Delay bound comparison before and after introducing sender offset.

With the offset imposed, worst-case delay of flows with Priority 1, 2 and 3 are improved by as much as several milliseconds, as shown in subplot 2 in Fig. 8. Noted, we do not calculate delay bound for BE flows in TSN, because traffic constraint cannot be committed at sender for those background traffic, thus arrival curve cannot be obtained and NC latency bound cannot be calculated.

To interpret this improvement, consider a flow pair, including a half-period offset sub-flow and a zero-offset sub-flow with the same path as well as other parameters. Theoretically,

[4]Flows with the same path and the same parameters obtain the same worst-case delay bound by network calculus. For simplicity, Fig. 6 and Fig. 8 show one flow for each flow path in Tab. III, without BE flows.

a flow pair has the same average rate comparing with two original sub-flows, but only half the burstiness. A flow pair has the same average rate and same burstiness comparing with a single aggregated flow with double of one original sub-flow's average rate. The relation between two original sub-flows, a flow pair, and a aggregated flow is shown in Fig. 9. By substituting aggregated flow parameters into (9), (11) and (14), the sum of burstiness for concurrent flows on the same port b_h, b^l is reduced by half. Therefore the delay in (16) is reduced, which is in accordance with results in Fig. 8. It is reasonable to have further improvement if we let the aggregated flow include more than two sub-flows with different offsets.

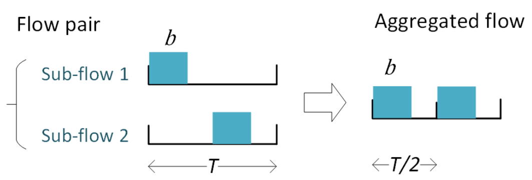

Fig. 9. Original sub-flows, a flow pair, and a aggregated flow.

Results imply that, for example, if a set of PLCs run multiple control loop applications, and several flows are forwarded through the same ports, it is better not to let all application cycles start at same time. Introducing constant offset to cyclic senders would reduce the worst-case delay, which is valid for both time-based scheduling and QoS-based shaping.

V. Conclusion

An industrial automation network model is analyzed where different types of traffic are queued and forwarded by a combined CBS+SP mechanism. Network calculus is used to evaluate worst case end-to-end latency bound for flows, where interleaved regulator is apply to facilitate theoretical analysis. Results of worst-case latency bound show that typical latency requirements in industrial automation network can be satisfied if the network is properly configured. The impact of different bandwidth usage and traffic sending schemes on the performance are also compared.

This work indicates that, network calculus can be used to give theoretical upper bound, which is vital to introduce asynchronous QoS-based shaping method (like CBS) to industrial automation network with transmitting latency requirement.

Future work will focus on expanding the usability of network calculus based TSN performance analysis, e.g., to analyze the effect of shaping parameters and to support more queuing and forwarding mechanisms (use a single mechanism as well as combinations), to support variable traffic specifications, and to support calculation of more flows.

References

[1] C. S. Chang, Performance guarantees in communication networks. Springer Science Business Media, 2012.

[2] J-Y. Le Boudec and P. Thiran, "Network calculus: a theory of deterministic queuing systems for the internet", Vol. 2050, Springer Science I& Business Media, 2001.

[3] J. A. R. De Azua and M. Boyer, Complete Modelling of AVB in Network Calculus Framework, in the 22nd ACM Intl Conf. on Real-Time Networks and Systems (RTNS), NY, USA, 2014, pp. 5564.

[4] S. Bondorf, P. Nikolaus, J. B. Schmitt, "Quality and Cost of Deterministic Network Calculus Design and Evaluation of an Accurate and Fast Analysis" , In Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 2017), 2017.

[5] J-Y. Le Boudec, A Theory of Traffic Regulators for Deterministic Networks with Application to Interleaved Regulators, IEEE/ACM Transactions on Networking 26.6 (2018): 27212733.

[6] E. Mohammadpour, E. Stai, M. Mohiuddin, and J.-Y. Le Boudec, "End-to-end Latency and Backlog Bounds in Time- Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping", In 2018 30th International Teletraffic Congress (ITC 30), vol. 2, pp. 1-6, 2018.

[7] N. Finn, J-Y. Le Boudec, E. Mohammadpour, et. al., DetNet Bounded Latency, Internet Engineering Task Force (IETF), Work in Progress, Mar 2019.

[8] M. Boyer and C. Fraboul, "Tightening end to end delay upper bound for AFDX network calculus with rate latency FIFO servers using network calculus," 2008 IEEE International Workshop on Factory Communication Systems, Dresden, 2008, pp. 11-20.

[9] M. Boyer, N. Navet, X. Olive, and E. Thierry. "The pegase project: Precise and scalable temporal analysis for aerospace communication systems with network calculus." In International Symposium On Leveraging Applications of Formal Methods, Verification and Validation, Springer, Berlin, Heidelberg, 2010, pp. 122-136.

[10] J. Georges, T. Divoux, and E. Rondeau, "Confronting the performances of a switched Ethernet network with industrial constraints by using the network calculus". Int. J. Commun. Syst., Vol. 18, pp. 877-903, 2008.

[11] S. Kerschbaum, K.-S. Hielscher, and R. German. "The need for shaping non-time-critical data in PROFINET networks." In 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), pp. 160-165. IEEE, 2016.

[12] IEC/IEEE P60802, Use Cases IEC/IEEE 60802, http:// www.ieee802.org /1/files/public/docs2018/60802-industrial-use-cases-0918-v13.pdf, p. 9.

[13] A. Ademaj, Industrial Automation Traffic Types and their Mapping to QoS/TSN Mechanisms, http://www.ieee802.org/1/files/public /docs2018/60802-ademaj-traffic-type- characterization-1118-v01.pdf, p. 3.

[14] L. Chen, Multi-traffic transmission in industrial backbone network, http://www.ieee802.org/1/files/public/docs2018/60802-chen-multi-traffic-transmission-on-backbone-0918.pdf.

[15] Y. Jiang, "A Basic Result on the Superposition of Arrival Processes in Deterministic Networks", 2018, [Online]. Available: http://arXiv:1804.10973

[16] J. C. R. Bennett, K. Benson, W. F. Courtney, J.-Y. Le Boudec, "Delay jitter bounds and packet scale rate guarantee for expedited forwarding", IEEE/ACM Trans. Netw., vol. 10, no. 4, pp. 529-540, Aug. 2002.

[17] IEEE 802.1, "IEEE 802.1Q-2018 - IEEE Standard for Local and Metropolitan Area NetworksBridges and Bridged Networks", IEEE WG 802.1, July 2018, http://www.ieee802.org/1/.

[18] W. Steiner, "Deterministic Ethernet for Real-Time and Critical Applications", 12th IEEE World Conference on Factory Communication System - Communication in Automation (WFCS 2016), Portugal, May, 2016

[19] N. Navet, J. Migge, J. Villanueva, M. Boyer, Pre-shaping bursty transmissions under IEEE802. 1Q as a simple and efficient QoS mechanism. SAE International Journal of Passenger Cars-Electronic and Electrical Systems. 2018 Apr 3;11(2018-01-0756).