Network Calculus-based Modeling of Time Sensitive Networking Shapers for Industrial Automation Networks

Xiaoyu Liu[1,2,3], Chi Xu[1,2,*], Haibin Yu[1,2]
[1]Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2]Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
Email:{liuxiaoyu1, xuchi, yhb}@sia.cn

***Abstract*—With the interconnection of distributed devices from different industrial automation networks, heterogeneous traffics with different Quality of Service (QoS) requirements increase dramatically. Shapers in Time-Sensitive Networking (TSN) enable the coexistence of different industrial traffics within a single network infrastructure. In this paper, with full consideration of the time-criticality of different industrial traffics and shaping mechanisms, we propose a network calculus-based modeling scheme for the widely-used Credit Based Shaper (CBS) and Time Aware Shaper (TAS). Then, we analyze the QoS of industrial traffics with different real-time characteristics by CBS and TAS, respectively. To evaluate the accuracy of network calculus-based modeling scheme, we compare our theoretical analysis with simulations, wherein a compactness function is defined to describe QoS compactness. Simulation results verify the effectiveness of proposed modeling scheme, by which we further validate both CBS and TAS can satisfy the QoS of traffics for different industrial automation scenarios.**

***Index Terms*—time sensitive networking, credit based shaper, time aware shaper, quality of service (QoS), network calculus, industrial automation networks**

I. Introduction

With the development of Industry 4.0, a large number of distributed intelligent devices are interconnected with each other, which significantly enhance the heterogeneous traffics in industrial automation networks. In particular, time-sensitive traffics have more stringent requirements on network bandwidth and QoS. Meanwhile, industry expects to apply a general and standard communication technology to make horizontal and vertical integration of manufacturing and business sectors. Ethernet is proven to be a flexible, highly scalable and open standard network technology in business. However, original Ethernet is not suitable for real-time and safety-critical applications in manufacturing. Fortunately, layered structure of Ethernet stack allows us to make improvements for different needs. As a result, various industrial Ethernet technologies with real-time extension are developed for industrial automation, e.g., PROFINET, EtherCAT, EPA, etc. However, these technologies are heterogeneous and proprietary, wherein special mechanisms are defined to guarantee time-criticality, and therefore their interoperability is difficult. To guarantee QoS for time-critical applications, the Time-Sensitive Networking (TSN) Task Group [1] is established to standardize real-time and safety-critical enhancements for Ethernet in 2012. Specifically, TSN, based on standard Ethernet, focuses on data link layer to provide a universal underlying architecture for high level protocols and applications. Thus, TSN is regarded as an enabling technology for the horizontal and vertical communication unification in industry. Currently, TSN standards can be roughly divided into three sections:

- ***Time synchronization*** establishes the foundation for the clocked end-to-end transmission of traffics with real-time requirements.
- ***Fault tolerance, resource reservation and path selection*** ensure that the latency and reliability requirements are met for real-time control traffics.
- ***Shaping and scheduling*** allow the joint transmission of traffics with real-time requirements as well as traditional best-effort traffics in a single network infrastructure.

Fully considering the characteristics of heterogeneous traffics in industry, we mainly focus on shaping and scheduling in this paper. With shaping and scheduling, TSN classifies traffics into several traffic-classes based on the time-criticality. TSN proposes several shapers in different mechanisms: time aware shaper (TAS) [2], credit based shaper (CBS) [3], peristaltic shaper (PS) [4], burst limiting shaper (BLS) [5] and so on. Modeling for input and output characteristics of traffics across shapers is an effective way to prove their correct behavior in temporal domain. [6][7] give a complete and accurate network calculus-based modelling of stream-reservation (SR) traffics in CBS, however the model of time-triggered traffics is in absence. [8][9] make delay analysis of SR traffics considering the time-triggered traffics in TAS based on [6]. However, the overlapping among different priority traffics is ignored.

* Corresponding author

[10][11] make formal worst-case timing analysis of BLS and PS by compositional performance analysis, then derive the blocking interference between traffics with different priorities. [12][13][14] model TSN shapers by different simulator, and analyze the QoS of real-time traffics. However, the simulation results are not compared with the formal analysis.

In our paper, we focus on two basic shapers with different triggered mechanisms, i.e., CBS and TAS. We use network calculus to analyze the performance of time-critical industrial traffics in CBS and TAS, wherein arrival and service curves are derive. To the best of our knowledge, this is the first work using network calculus to model and analyze multiple traffics for industrial automation networks. Then, we analyze the QoS of different time-critical indutrial traffics by network calculus and simulation. Especially, we define a compactness function to verify the QoS compactness of network calculus-based theoretical analysis and simulation results. Finally, simulation results fit well with proposed theoretical analysis. The main contributions of our work are as follows: A network calculus-based modeling scheme of TSN shaper for industrial traffics is proposed. Time-triggered traffics are considered in CBS, and the overlapping interference of time-slot among different priorities industrial traffics are considered in TAS.

This remaining of this paper is organized as follows: Section II presents the system model. Section III introduces basic theory of network calculus and the formal analysis of QoS. Section IV and Section V analyze the characteristics of traffics in CBS and TAS and establish the network calculus-based models of CBS and TAS, respectively. Section VI compares the QoS of traffics in different shapers in formal by network calculus. Section VII defines a compactness function and performs extensive simulations to verify the accuracy of proposed network calculus-based models and analyze the QoS of various traffics in CBS and TAS. Section VIII concludes that network calculus-based modeling scheme is effective, and CBS as well as TAS are suitable for different industrial automation network scenarios.

II. System Model

A TSN network consists of a set of end nodes (ENs) and switches (SWs), connecting via physical links. The physical links are full duplex, allowing communication in both directions, and the network can be multi-hop. The output port of a SW is connected to one EN or the input port of another SW. The applications running in ENs communicate via flow (i.e., aggregation of the same classes traffics), which have a single source and may have multiple destinations. An EN could send multiple-classes flows to the network and receive multiple-classes flows from other ENs, as shown in Fig. 1. Our model uses the input constraints of ENs and the output constraints of SWs, which are not related to the topology but only to the concatenation of SWs. The concatenation is determined by routing. The routing is not our research work in this paper, and we assume that the routing has already been determined.

According to the time-criticality in real industrial automation networks, we mainly consider the the following three

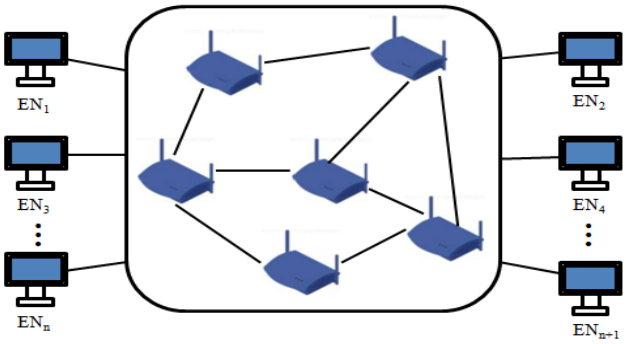

Fig. 1. System Model.

kinds of traffics: **CT** traffics, represent the periodic real-time traffics such as control traffics; **AV** traffics, represent the non-periodic real-time traffics such as audio and video traffics; **BE** traffics, represent non-periodic non-real-time traffics such as monitoring and configuration traffics.

To analyze the timeliness, we focus on the the end-to-end delay, which is one of the most concerned QoS. More specifically, the end-to-end delay consists of propagation delay, transmission delay, queuing delay and hardware delay.

$$T_{dd} = t_{prop} + t_{tran} + t_{que} + t_{hdwa}, \tag{1}$$

where T_{dd} denotes the total end-to-end delay, t_{prop}, t_{tran}, t_{que} and t_{hawa} denote propagation delay, transmission delay, queuing delay and hardware delay, repectively.

As ENs and SWs are connected through Ethernet physical links, the propagation delay and the transmission delay are constant, which depends on the physical medium (e.g., fiber, twisted pairs, etc.) and the size of the frame of **CT**, **AV** and **BE**. Hardware delay is very small that can be generally ignored. Queuing delay, which is variable depending on different flow characteristics, is the most critical factor affecting total end-to-end delay.

III. Network Calculus Preliminary

Network calculus is a theory by means of the min-plus convolution operation algebra, and is designed for the deterministic performance analysis of network communication, such as the worst-case bounds on latency and backlog of traffics [15]. It is used to construct the arrival and service curve models of the traffics from source ENs and passing SWs.

An arrival curve $\alpha(t)$ is a constraint on the arrival process $R(t)$ of one flow, where $R(t)$ denotes the input cumulative function counting the total data bits of the flow that has arrived at the node up to time t. Thus, $R(t)$ is constrained by $\alpha(t)$ if

$$R(t) \leq \inf_{0 \leq u \leq t} \{R(u) + \alpha(t-u)\} = (R \otimes \alpha)(t), \tag{2}$$

where $\otimes$ denotes min-plus convolution. A typical arrival curve model is usually a leaky bucket model given by

$$\alpha_{r,b}(t) = \begin{cases} rt + b, t \geq 0, \\ 0, t < 0, \end{cases} \tag{3}$$

where r denotes the average rate of the flow over a long time and b denotes the maximum burst tolerance of the flow.

A service curve $\beta(t)$ models the service of SW. Assume that $R^*(t)$ is the output cumulative function that counts the total data bits of the flow departing from the network node up to time t.

$$R^*(t) \geq \inf_{0 \leq u \leq t} \{R(u) + \beta(t-u)\} = (R \otimes \beta)(t). \quad (4)$$

A typical service curve model is a rate-latency model given by

$$\beta_{R,T}(t) = R\left[t - T\right]^+, \quad (5)$$

where R denotes the service rate, T denotes the service latency and the notation $[x]^+$ is equal to x if $x \geq 0$ and 0 otherwise.

In order to describe the service of shapers provided to flows, strict service curve of shapers is defined below.

During any backlogged period of duration Δt, i.e., $R^*(t) < R(t)$, the maximum number of bits that a shaper can provide to flow satisfies:

$$R^*(t) - R(t) \leq \sigma(t), \quad (6)$$

where $\sigma(t)$ denotes the strict service curve of shapers.

The concatenation of network components, especially the concatenation of SWs, will change service rate and service delay of service curve. The total service curve of the concatenation of SWs is formulated as the service curve of network system. According to [15], the total rate-latency service curve is defined as follows:

$$\begin{aligned} \beta_{R,T}(t) &= \beta_{R_1,T_1}(t) \otimes \beta_{R_2,T_2}(t) \otimes ... \otimes \beta_{R_n,T_n}(t) \\ &= \beta_{\min(R_1,R_2,...,R_n),\sum_{i=1}^{n}(T_1,T_2,...,T_n)}(t), \end{aligned} \quad (7)$$

where $\min(r_1, r_2, ..., r_n)$ is the minimum service rate of all concatenate SWs and $\sum_{i=1}^{n}(T_1, T_2, ..., T_n)$ is the sum of service delays of all concatenate SWs.

Assuming that the flow constrained by $\alpha(t)$ traverses the switched network offering $\beta(t)$. The latency experienced by the flow in the network is bounded by the maximum horizontal deviation between the graphs of $\alpha(t)$ and $\beta(t)$.

$$h(\alpha, \beta) = \sup_{s \geq 0} \{\inf \{\tau \geq 0 | \alpha(s) \leq \beta(s+\tau)\}\}. \quad (8)$$

Key notations related to network calculus-based models in this paper are summarized in Table 1.

TABLE I
SUMMARY OF KEY NOTATIONS

C	Physical link rate
$l_{CT,AV,BE}^{max}$	the max bit size of CT,AV,BE frame
$\alpha_{CT,AV,BE}^{CBS}$	Arrival curve of CT,AV,BE in CBS
$b_{CT,AV,BE}^{CBS}$	max burst bits of a class of traffics of CT,AV,BE in CBS
$r_{CT,AV,BE}^{CBS}$	average bit rate of a class of traffics of CT,AV,BE in CBS
σ_{SR}^{CBS}	strict service curve guaranteed to SR-Classes in CBS
R_{SR}^*	output cumulative function of SR-Classes
Δt_{SR}^+	transmission duration of SR-Classes
Δt_{SR}^-	waiting duration of SR-Classes
C_h^{SR}	upper bound of credit of SR-Classes
$idSl_{SR}$	idleSlope of SR-Classes
$sdSl_{SR}$	sendSlope of SR-Classes
$C_h^{A,B}$	upper bound of credit of SR-A or SR-B
$C_l^{A,B}$	lower bound of credit of SR-A or SR-B
$idSl_{A,B}$	idleSlope of SR-A or SR-B
$sdSl_{A,B}$	sendSlope of SR-A or SR-B
$\alpha_{CT,AV,BE}^{TAS}$	Arrival curve of CT,AV,BE in TAS
$b_{AV,BE}^{TAS}$	max burst bits of a class of traffics of CT,AV,BE in TAS
$r_{AV,BE}^{TAS}$	average bit rate of a class of traffics of CT,AV,BE in TAS
σ_{CT}^{TAS}	strict service curve guaranteed to CT in TAS
σ_{AV}^{TAS}	strict service curve guaranteed to AV in TAS
t_{AV}^{start}	start time of transmission time-slot of AV
t_{AV}^{stop}	stop time of transmission time-slot of AV
t_{BE}^{start}	start time of transmission time-slot of BE
t_{BE}^{stop}	stop time of transmission time-slot of BE
d_{AV}	the delay of AV
D_{AV}	the max delay of AV
$L_{CT,AV,BE}^{gb}$	the guard band size of CT,AV,BE
P_{CT}	the cycle period of CT in TAS
$T_{CT}^{TAS,CBS}$	the service latency of CT in TAS or CBS
$T_{AV}^{TAS,CBS}$	the service latency of AV in TAS or CBS

IV. NETWORK CALCULUS-BASED MODELING FOR CBS

A. Flows in CBS

CBS schedules two classes of flow, denoted as SR-Class flow and best-effort flow, while SR-Class flow has higher priority than best-effort flow. To indicate whether or not SR-Class flow could be scheduled, the credit is defined. C_h and C_l are the upper and lower bound of the credit that SR-Class flow can reach. The credit is accumulated at a constant rate called *idleSlope* (*idSl*) when (i) the SR-Class flow is waiting for transmission; (ii) no more frames of SR-Class flow is waiting but the credit is negative. The credit is decremented at a constant rate called *sendSlope* (*sdSl*), when the SR-Class flow is transmitted [3]. The relationship of *idleSlope* and *sendSlope* is given by

$$sdSl = C - idSl. \quad (9)$$

Taking two types of SR-Classes as examples (e.g., SR-A and SR-B where SR-A has higher priority than SR-B), the process of shaping is shown in Fig. 2.

B. Arrival and Service Curves in CBS

CT and **AV** flows is classified as SR-Classes while **CT** has higher priority than **AV**, **BE** flow is classified as best-effort class. CBS guarantees bound on delay for SR-Classes and avoids starvation of best-effort class. Arrival curves of **CT**,

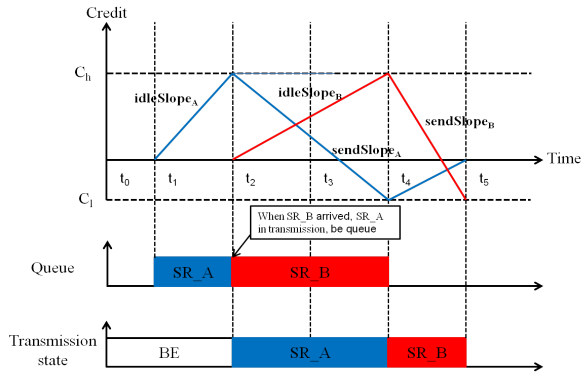

Fig. 2. Credit Based Shaper.

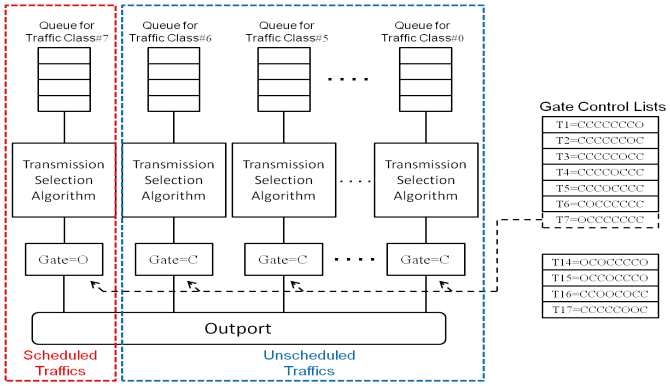

Fig. 3. Time-Aware Shaper.

AV and **BE** flows observe T-SPEC (traffic specification) [15] as follow:

$$\alpha^{CBS}_{CT,AV,BE}(t) = \min(Ct, b^{CBS}_{CT,AV,BE} + r^{CBS}_{CT,AV,BE}t). \quad (10)$$

Next, we derive the strict service curves of CBS provided to SR-Class flow.

Theorem 1*: The Strict Service Curve of CBS provided to SR-Class flow is given by*

$$\sigma^{CBS}_{SR}(t) = \frac{C \cdot idSl_{SR}}{idSl_{SR} - sdSl_{SR}} \left[t - \frac{C^{SR}_h}{idSl_{SR}} + \frac{C^{SR}_l}{idSl_{SR}} \right]^+. \quad (11)$$

Proof: The output cumulative function of SR-Class flow satisfies $R^*_{SR}(t+\Delta t) - R^*_{SR}(t) \le \sigma_{SR}$. We define duration $\Delta t = \Delta t^- + \Delta t^+$ where Δt^+ denotes transmission duration and Δt^- denotes waiting duration. From the definition of cumulative function, $R^*_{SR}(t+\Delta t) - R^*_{SR}(t) = C\Delta t^-_{SR}$. Moreover, $C^{SR}_h - C^{SR}_l \ge idSl_{SR}\Delta t^+_{SR} + sdSl_{SR}\Delta t^-_{SR}$. Thus, during duration Δt, the output cumulative function of SR-Class flow can be transformed to $R^*_{SR}(t+\Delta t) - R^*_{SR}(t) \le \frac{C \cdot idSl_{SR}}{idSl_{SR} - sdSl_{SR}} \left(\Delta t - \frac{C^{SR}_h}{idSl_{SR}} + \frac{C^{SR}_l}{sdSl_{SR}} \right)$. So the strict service curve of CBS provided to SR-Class flow is $\sigma^{CBS}_{SR}(t) = \frac{C \cdot idSl_{SR}}{idSl_{SR} - sdSl_{SR}} \left[t - \frac{C^{SR}_h}{idSl_{SR}} + \frac{C^{SR}_l}{idSl_{SR}} \right]^+$. ■

Note that the number of SR-Classes is variable, and the upper bound of low priority SR-Class is affected by high priority SR-Class, as shown in Fig. 2.

V. Network Calculus-based Modeling for TAS

A. Flows in TAS

TAS introduces time-based gates at the output ports that binds the transmission of frames from the queues to an off-line configured periodic schedule list called gate control lists (GCLs). At a given time, a gate can be either in open or closed state. When the gate is open, frames from the respective queue selected by through transmission selection mechanism are forwarded in first-in-first-out (FIFO) order to the connected physical link [2]. A time-critical communication is configured by so-called time-triggered (TT) flow. The queues are assigned to be either scheduled (i.e., TT queues) or unscheduled. Time-critical flow is isolated from non-time-critical flow and loaded into the scheduled queues which have separate transmission time-slot. Unscheduled queues follow a priority arbitration to compete transmission time-slot, as shown in Fig. 3.

B. Arrival and Service Curves in TAS

CT flow is configured as TT flow and loaded into scheduled queues. **AV** and **BE** flows are configured as priority class and loaded into unscheduled queues. A separate transmission time-slot allocated to **CT** flow makes **CT** flow can employ all network bandwidth, so the arrival curve of CT flow is

$$\alpha^{TAS}_{CT}(t) = Ct. \quad (12)$$

The open periods of time-slots of **AV** and **BE** flows may be overlapping, and **AV** and **BE** flows compete for network bandwidth. The arrival curves of **AV** and **BE** flows are also T-SPEC, given by

$$\alpha^{TAS}_{AV,BE}(t) = \min(Ct, b^{TAS}_{AV,BE} + r^{TAS}_{AV,BE}t). \quad (13)$$

In GCLs, overlapping among the open period of time-slots of **AV** and **BE** flows could affect their communication behavior. However, the complete isolation of **CT** flow protects **CT** from overlapping. In addition, guard band is a mechanism to limit the bursts to protect adjacent transmission time-slot of other flows.

- *Overlapping among different priorities*

AV flow has higher priority and stricter real-time requirements than **BE** flow. The overlapping among the open period of time-slots of **AV** and **BE** flows are shown in Fig. 4.

In Fig. 4(a), **BE** flow starts transmission earlier than **AV** flow, i.e., $t^{start}_{BE} < t^{start}_{AV}$, and **BE** flow stops transmission earlier than **AV** flow, i.e., $t^{stop}_{BE} < t^{stop}_{AV}$. No consideration of preemption, **BE** flow that has been transmitted will not be interrupted. **AV** flow will be delayed $d_{AV} = \min\left\{ \frac{l^{max}_{BE}}{C}, \left(t^{stop}_{BE} - t^{start}_{AV}\right) \right\}$. Fig.4 (b) is contrary to Fig.4 (a), **BE** flow starts transmission later than **AV** flow, i.e., $t^{start}_{BE} > t^{start}_{AV}$, and **BE** flow stops transmission later than **AV** flow, i.e., $t^{stop}_{BE} > t^{stop}_{AV}$. **AV** flow will not be delayed while **BE** flow will be delayed. Fig.4 (c) is a special case of Fig.4 (b), the transmission period of **BE** flow is completely overlapped by that of **AV** flow, i.e., $t^{start}_{BE} > t^{start}_{AV}$ and $t^{stop}_{BE} < t^{stop}_{AV}$. **AV**

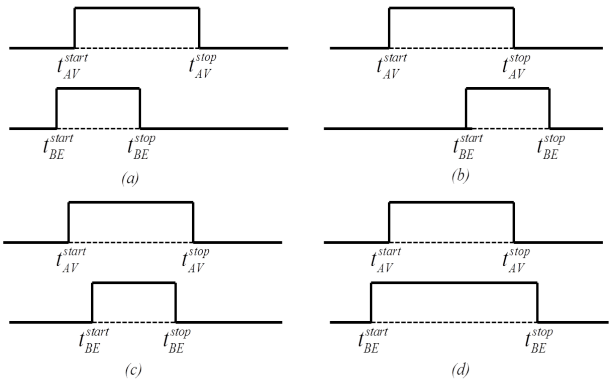

Fig. 4. Overlapping among AV and BE Flow.

flow will not be delayed while **BE** flow will be lost. Fig.4 (d) is a special case of Fig.4 (a), **AV** flow will be delayed $d_{AV} = \frac{l_{BE}^{max}}{C}$.

Therefore, only $t_{BE}^{start} < t_{AV}^{start} < t_{BE}^{stop}$, overlapping among the open period of gate of **AV** and **BE** flows will delay the transmission of **AV** flow.

Considering the worst case of multi-priority **BE** flow coexistence,

$$D_{AV} = \max_{mul-prio\ BE}(d_{AV}). \tag{14}$$

- *Guard band*

Guard band is a cycle before the cut-off time of former transmission time-slot to limit burst interference to later transmission time-slot [3], shown in Fig. 5. The duration of guard band is defined as follows:

$$L_{CT,AV,BE}^{gb} = \frac{l_{CT,AV,BE}^{max}}{C}. \tag{15}$$

Separate time-slot for **CT** flow result in that the strict service curve of TAS provided to **CT** flow is influenced by guard band and FIFO order, so the strict service curve of TAS provided to **CT** flow is defined in Theorem 2.

Theorem 2*: The Strict Service Curve of TAS provided to CT flow is given by*

$$\sigma_{CT}^{TAS} = C\left(t - L_{CT}^{gb} - \frac{l_{CT}^{max}}{C}\right). \tag{16}$$

Proof: Assuming that when a **CT** flow arrives, an **AV** or **BE** flow is already in transmission. The guard band allows the transmission of an unfinished frame of **AV** or **BE** flow before the gate of **CT** open to ensure the integrity of the **CT** flow. When the **CT** flow starts to transmit, it is delayed $\frac{l_{CT}^{max}}{C}$ due to FIFO. ■

AV flow competes the same time-slot with **BE** flow, thus the strict service curve of TAS provided to **AV** flow is defined in Theorem 3.

Theorem 3*: The Strict Service Curve of TAS provided to AV flow is given by*:

$$\sigma_{AV}^{TAS} = C\left(t - L_{AV}^{gb} - \frac{l_{AV}^{max}}{C} - D_{AV} - P_{CT}\right). \tag{17}$$

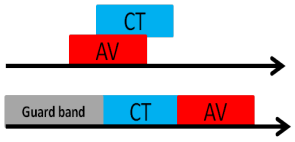

Fig. 5. Guard Band.

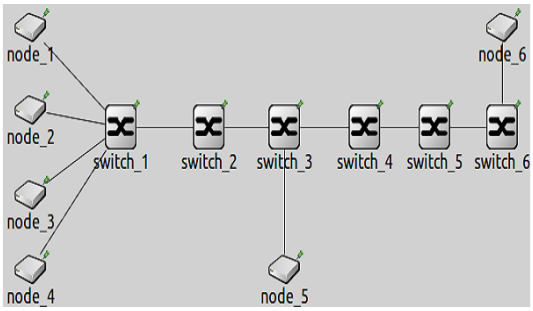

Fig. 6. System Model in Simulation.

Proof: The main factors interfering an **AV** flow include: guard band and FIFO of **AV** flow, overlapping among **AV** and **BE** flows and transmission time-slot of **CT** flow. The interference from guard band and FIFO of **AV** flow use the same methodology as the proof of (16). The interference from overlapping between **AV** and **BE** flow is presented in section V-B. The interference from transmission time-slot of **CT** flow is result from the cycle period. If the **AV** flow just catch up with the transmission time-slot of the **CT** flow, the **AV** flow will be delayed P_{CT}. ■

VI. QoS Analysis

A. Network Calculus-Based QoS of CT flow

Based on (15)(16), the total service latency of **CT** flow in TAS is $T_{CT}^{TAS} = L_{CT}^{gb} + \frac{l_{CT}^{max}}{C} = \frac{2l_{CT}^{max}}{C}$. Assume (i) **CT** flow is transmitted by SR-A, (ii) there is no packet of **CT** flow in the queue and the initial credit of **CT** is 0. Based on (11), the total service latency of **CT** flow in CBS is $T_{CT}^{CBS} = \frac{C_h^A - C_l^A}{idSl_A}$. When a frame of **CT** flow is in transmission, the credit of **CT** flow is decreased. The minimal value of the credit is reached when the size of the transmitted frame is the maximal size l_{CT}^{max}, i.e., $C_l^A = \frac{l_{CT}^{max}}{C} \cdot sdSl_A$. When a frame of **CT** flow is waiting for transmission in the queue, the credit of **CT** flow is increased. The maximal value of the credit is reached at the end of the transmission of **AV** or **BE** frame with the maximal size $l_{AV,BE}^{max}$, i.e., $C_h^A = \frac{l_{AV,BE}^{max}}{C} \cdot idSl_A$. Thus, $T_{CT}^{CBS} = \frac{l_{BE}^{max}}{C} + \frac{l_{CT}^{max}}{C} \cdot \frac{C - idSl_A}{idSl_A}$. Obviously, $T_{CT}^{TAS} < T_{CT}^{CBS}$, i.e., the upper delay bound of **CT** flow in TAS is smaller than it in CBS.

B. Network Calculus-Based QoS of AV flow

Based on (14)(15)(17), the total service latency of **AV** flow in TAS is $T_{AV}^{TAS} = L_{AV}^{gb} + \frac{l_{AV}^{max}}{C} + D_{AV} + P_{CT}$. Assume (i) **AV** flow is transmitted by SR-B, (ii) there is no packet of **AV** flow in the queue and the initial credit of **AV** is

0. Based on (11), the total service latency of **AV** flow in CBS is $T_{AV}^{CBS} = \frac{C_h^B - C_l^B}{idSl_B}$. When a frame of **AV** flow is in transmission, the credit of **AV** flow is decreased. The minimal value of the credit is reached when the size of the transmitted frame is the maximal size l_{AV}^{max}, i.e., $C_l^B = \frac{l_{AV}^{max}}{C} \cdot sdSl_B$. When a frame of **AV** flow is waiting for transmission in the queue, the credit of **AV** flow is increased. The maximal value of the credit is reached at the end of the transmission of maximum frames of **BE** and **CT** flows. We define the maximum transmission duration of **CT** flow as $\Delta t_{CT}^{max} = \frac{C_l^A - C_h^A}{sdSl_A}$. Then, $C_h^B = \left(\frac{l_{BE}^{max}}{C} + \Delta t_{CT}^{max}\right) \cdot idSl_B$. Thus, $T_{AV}^{CBS} = \frac{l_{BE}^{max}}{C}\left(1 + \frac{idSl_A}{C - idSl_A}\right) + \frac{l_{CT}^{max}}{C} + \frac{l_{AV}^{max}}{C}\left(\frac{C}{idSl_B} - 1\right)$. So, the relation of T_{AV}^{TAS} and T_{AV}^{CBS} is uncertain.

VII. Simulation Results

A. Simulation setup

We employ OMNeT++ [16] to simulate and verify our previous analysis. Without loss of generality, we formulate a TSN-enabled industrial automation network by 6 ENs (i.e., $node\_1$ to $node\_6$), and 6 SWs (i.e., $switch\_1$ to $switch\_6$), as shown in Fig. 6. The physical links are full duplex and the rate is $C = 100Mbps$. **CT** flow is usually with short packet, and the size of **AV** frame is usually larger than **CT**. **BE** flow is used as interference flow with the maximum size of frame.

For CBS, the applications running in ENs are set as follows: $node\_1$ and $node\_2$ generate **CT** flow with $l_{CT} = 100byte$ and $T_{CT} = 125us$ by SR-A. $node\_3$ and $node\_4$ generate **AV** flow with $l_{AV} = 400byte$ and $T_{AV} \sim U(0, 250us)$ by SR-B, where $U(\cdot)$ denotes uniform distribution. $node\_5$ broadcasts **BE** flow with $l_{AV} = 1500byte$ and $T_{AV} \sim U(0ms, 2ms)$ by best-effort to all nodes. $node\_6$ is a listener, receiving all flow from $node\_1$ to $node\_5$.

Similarly, for TAS, the applications running in ENs are set as follows: $node\_1$ and $node\_2$ generate **CT** flow with $l_{CT} = 100byte$ and cycle period of time slot is $T_{CT} = 125us$. $node\_3$ and $node\_4$ generate **AV** flow with $l_{AV} = 400byte$ and $T_{AV} \sim U(0, 250us)$ by higher priority. $node\_5$ broadcasts **BE** flow with $l_{AV} = 1500byte$ and $T_{AV} \sim U(0ms, 2ms)$ by lower priority to all nodes. $node\_6$ is a listener, receiving all flow from $node\_1$ to $node\_5$.

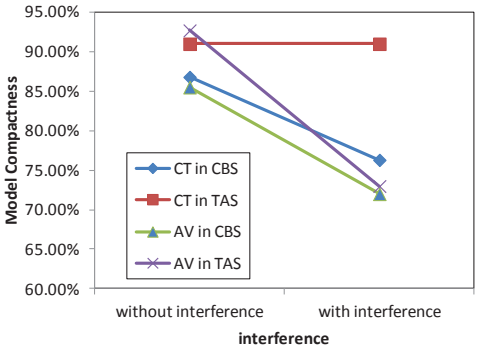

Fig. 7. Model Compactness.

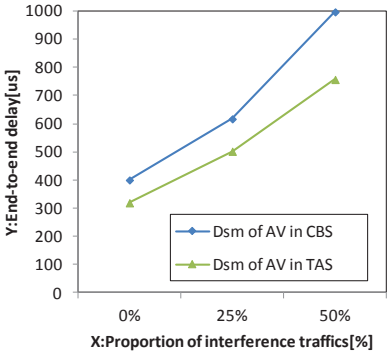

Fig. 8. Delay of AV Flow in Shapers.

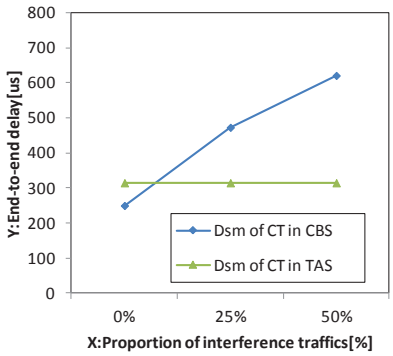

Fig. 9. Delay of CT Flow in Shapers.

Moreover, parameters in arrival and service curves are given as follows: b_{CT}^{CBS} is set as $100byte$, and r_{CT}^{CBS} is set as $100byte/125us = 6.4Mbps$. b_{AV}^{CBS} is set as $400byte$, and r_{AV}^{CBS} is set as $400byte/250us = 12.8Mbps$. b_{BE}^{CBS} is set as $1500byte$, and r_{BE}^{CBS} varies due to its sending-interval follows a uniform distribution. Similarly, b_{AV}^{TAS} is set as $400byte$, and r_{AV}^{TAS} is set as $400byte/250us = 12.8Mbps$. b_{BE}^{TAS} is set as $1500byte$, and r_{BE}^{TAS} varies due to its sending-interval follows a uniform distribution.

B. Results and Discussion

To guarantee the most stringent real-time capabilities of the traffics in industrial automation networks, isochronous real-time traffic-class is defined depending on the time-criticality [17][18]: cycle time less than $1ms$, jitter less than $1us$.

To verify the accuracy of the models, the compactness function is defined as

$$\lambda = \frac{D_{sm}}{D_{nc}} \times 100\%, \tag{18}$$

where D_{sm} denotes the delay in simulation, and D_{nc} denotes the worst-case delay in network calculus-based models.

Simulations have been done with different proportion of **BE** flow (as interference flow) where the proportion of **BE** flow is set to 0%, 25% and 50%. Limited by the length of the paper, we only verify the compactness with or without interference flow, but it also adequately reflects the accuracy of the proposed network calculus-based models of CBS and

TAS with variable loads and different heterogeneous traffics, shown in Fig. 7. The delay variation trends of **AV** and **CT** flows are synthesized in Figs. 8 and 9.

The delay in network calculus-based model is more pessimistic than that in simulation. With the increase of heterogeneous traffics in industrial automation network, the compactness will decrease. The maximum frames of heterogeneous traffics in transmission and the worst-case overlapping are considered, and the delay is enlarged through min-plus convolution in network calculus-based model. However, in real network, the size of frame varies, and overlapping among **AV** flow and all priorities of **BE** flows will not happen simultaneously.

For **CT** flow, with the increase of the proportion of interference flow, the delay in CBS increases significantly due to priority-based arbitration essentially, while the delay in TAS hardly changes because the complete isolation from **AV** and **BE** flows. There is a phenomenon that when the proportion of interference flow is small, the delay in CBS is less than it in TAS. Because as long as no **BE** flow is in transmission, **CT** flow is served as soon as it arrived in CBS while **CT** flow need to wait the periodic time-slot. When the proportion of interference flow fluctuates, the queue congestion also cause the more jitter in CBS, while the **CT** flow in TAS is not affected by the change of interference flow because of the separate transmission time-slot.

For **AV** flow, with the increase of the proportion of interference flow, the delay in CBS increases due to competing for channel resources. In TAS, **AV** flow and interference flow compete network resources based on priority arbitration due to the existence of shared time-slot. Large amount of interfrernce flow will inevitably cause large delay of **AV** flow.

In general, when the proportion of interference flow is small, i.e., low-load, both CBS and TAS can guarantee the QoS while CBS may perform better. As the proportion of interference flow increasing, TAS has better QoS. In other words, CBS is well-suited for the prioritized transmission in the coexistence of less traffic-classes, and the maximum delay could not be met with the further increasment of interference flow. In contrast, scheduled time-slot in TAS is more suitable for high-load and the coexistence of multiple traffic-classes in the industrial automation networks, where the maximum end-to-end latency is far less than $1ms$.

VIII. Conclusion

TSN shapers bounded the QoS of traffics with different time-criticality in a single network infrastructure. By deriving the arrival and service curves, we for the first time described the input and output characteristics of different time-critical industrial traffics across shapers. Then, we modeled two basic shapers by network calculus. In order to evaluate the accuracy of proposed models, we defined a compactness function which compares the theoretical values obtained from the network calculus-based models with the simulation results. The compactness showed that the proposed network calculus-based models of TSN shapers for industrial traffics is feasible. Finally, we validated both CBS and TAS can satisfy the QoS of industrial traffics for different industrial automation scenarios.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61803368, 61533015, Liaoning Provincial Natural Science Foundation of China under Grant No. 20180540114, and the Youth Innovation Promotion Association CAS under Grant 2019202.

References

[1] IEEE 802.1 TSN Task Group, "IEEE 802.1 Time-Sensitive Networking," Available: http://www.ieee802.org/1/pages/tsn.html.

[2] IEEE 802.1 TSN Task Group, "IEEE 802.1Qbv-Enhancements for Scheduled Traffic," Available: http://www.ieee802.org/1/pages/802.1bv.html.

[3] IEEE 802.1 AVB Task Group, "IEEE 802.1Qav- Forwarding and Queuing Enhancements for Time-Sensitive Streams," Available:http://www.ieee802.org/1/pages/802.1av.html.

[4] F. J. Gotz, "Traffic Shaper for Control Data Traffic (CDT)," IEEE 802 AVB Meeting, 2012.

[5] M. J. Teener, "Peristaltic Shaper: updates, multiple speeds," IEEE 802 January Plenary Meeting, 2014.

[6] D. Azua, J. A. Ruiz and M. Boyer, "Complete modelling of AVB in Network Calculus Framework," Proceedings of the 22nd International Conference on Real-Time Networks and Systems(RTNS) , 2014, pp.55-64.

[7] L. Zhao, F. He and J. Lu, "Comparison of AFDX and audio video bridging forwarding methods using network calculus approach," 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), IEEE, 2017, pp.1-7.

[8] D. Maxim, and Y. Q. Song, "Delay analysis of AVB traffic in time-sensitive networks (TSN)," Proceedings of the 25th International Conference on Real-Time Networks and Systems, ACM, 2017, pp. 18-27.

[9] L. X. Zhao et al, "Timing analysis of AVB traffic in TSN networks using network calculus," 2018 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS), 2018, pp.25-36.

[10] D. Thiele, and R. Ernst, "Formal worst-case timing analysis of Ethernet TSN's burst-limiting shaper," Proceedings of the 2016 Conference on Design, Automation and Test in Europe, EDA Consortium, 2016, pp.187-192.

[11] D. Thiele, R. Ernst and J. Diemer, "Formal worst-case timing analysis of Ethernet TSN's time-aware and peristaltic shapers," 2015 IEEE Vehicular Networking Conference (VNC), 2015, pp.251-258.

[12] J. H. Jiang et al, "A Time-sensitive Networking (TSN) Simulation Model Based on OMNET++," 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, pp.643-648.

[13] P. Heise, F. Geyer and R. Obermaisser, "TSimNet: An industrial time sensitive networking simulation framework based on OMNeT++," 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2016, pp.1-5.

[14] M. Pahlevan, and R. Obermaisser, "Evaluation of time-triggered traffic in time-sensitive networks using the opnet simulation framework," 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 2018, pp.283-287.

[15] J.-Y. L. Boudec and P. Thiran, Network calculus: a theory of deterministic queuing systems for the internet, Springer Science & Business Media, 2001.

[16] OMNeT++ Community, OMNeT++ 5.0, Available: https://www.omnetpp.org/.

[17] S. Y. Nof , Springer handbook of automation, Springer Science & Business Media, 2009.

[18] P. Neumann , "Communication in industrial automation: What is going on?" Control Engineering Practice 15.11 2007, pp.1332-1347.