Worst-Case Delay Analysis of Variable Bit-Rate Flows in Network-on-Chip with Aggregate Scheduling

Fahimeh Jafari, Axel Jantsch, and Zhonghai Lu
KTH Royal Institute of Technology, Sweden
Email: {fjafari, axel, zhonghai}@ kth.se

***Abstract*—Aggregate scheduling in routers merges several flows into one aggregate flow. We propose an approach for computing the end-to-end delay bound of individual flows in a FIFO multiplexer under aggregate scheduling. A synthetic case study exhibits that the end-to-end delay bound is up to 33.6% tighter than the case without considering the traffic peak behavior.**

I. Introduction

Real-time applications such as multimedia and gaming boxes etc., require stringent performance guarantees, usually enforced by a tight upper bound on the maximum end-to-end delay [1]. For the worst-case performance analysis, we derive the upper delay bound of a flow in a FIFO multiplexing and aggregate scheduling network. The behavior of a flow is determined by four parameters including the maximum transfer size (L), peak rate (p), burstiness (σ), and average sustainable rate (ρ). To calculate the tight delay bound per flow, the main problem is to obtain the *end-to-end Equivalent Service Curve (ESC)* which the tandem of routers provides to the flow. However, the required propositions for calculating performance metrics of *Variable Bit-Rate (VBR)* traffic characterized with (L, p, σ, ρ), transmitted in the FIFO order and scheduled as aggregate do not exist. Based on network calculus [2][3], we first present and prove the required propositions and then calculate the delay bound under the mentioned system model.

There are some works for deriving per-flow worst-case delay bound under different system models [4]-[6]. However, they investigate computing delay bounds only for average behavior of flows while we analyze both average and peak behavior. In [7], we presented a theorem for computing output traffic characterization. The aim of this paper is to represent and prove propositions for deriving end-to-end ESC and tighter upper bound on the end-to-end delay.

II. Network Calculus Background

In network calculus, traffic flows are modeled by arrival curves and network elements by service curves. Network calculus uses Traffic SPECification to model the average and peak characteristics of a flow. With TSPEC, f_j is characterized by an *arrival curve* $\alpha_j(t) = min(L_j + p_j t, \sigma_j + \rho_j t)$ in which L_j is the maximum transfer size, p_j the peak rate ($p_j \geq \rho_j$), σ_j the burstiness ($\sigma_j \geq L_j$), and ρ_j the average (sustainable) rate. We denote it as $f_j \propto (L_j, p_j, \sigma_j, \rho_j)$.

Network calculus also derives delay bound for lossless systems with service guarantees as the following theorem proves.

Theorem 1. *(Delay Bound [3]). Assume a flow, constrained by arrival curve α, traverses a system that offers a service curve of β, the virtual delay $d(t)$ for all t satisfies: $d(t) \leq h(\alpha, \beta)$.*

978-3-9810801-8-6/DATE12/

The theorem says that the delay is bounded by the maximum horizontal deviation between the arrival and service curves.

Now, we consider a node which guarantees a minimum service curve to an aggregate flow and also handles packets in order of arrival at the node.

Theorem 2. *(FIFO Minimum Service Curves [3]). Consider a lossless node serving two flows, 1 and 2, in FIFO order. Assume that packet arrivals are instantaneous. Assume that the node guarantees a minimum service curve β to the aggregate of the two flows. Assume that flow 2 has α_2 as an arrival curve. Define the family of functions $\beta^{eq}(t, \alpha_2, \tau) \equiv \beta_1^{eq}(t, \tau) = \beta_1^{eq}(t, \tau) = [\beta(t) - \alpha_2(t - \tau)]^+_{\{t>\tau\}}$. For any $\tau \geq 0$ such that $\beta_1^{eq}(t, \tau)$ is wide-sense increasing, then flow 1 is guaranteed the service curve $\beta_1^{eq}(t, \tau)$.*

III. System Model

We assume that flows are classified into a pre-specified number of aggregates. In addition, we assume that traffic of each aggregate is buffered and transmitted in the FIFO order, denoted as FIFO multiplexing. Different aggregates are buffered separately. The network is lossless, and packets traverse the network using a deterministic routing. We call the flow for which we shall derive its delay bound *tagged flow*, other flows that share resources with it *interfering flows*.

While building network calculus analysis models, we follow the notation conventions in the min-plus algebra [3]. $\otimes$ represents the min-plus convolution of two functions $f, g \in F$, the set of wide-sense increasing functions defined on $[0, t)$, $(f \otimes g)(t) = inf_{0 \leq s \leq t}\{f(t-s) + g(s)\}$; $\wedge$ represents the minimum operation, $f \wedge g = min(f, g)$. *Burst delay* function $\delta_T(t) = +\infty$, if $t > T$; $\delta_T(t) = 0$, otherwise. *Affine function* $\gamma_{b,r}(t) = b + rt$, if $t > 0$; $\gamma_{b,r}(t) = 0$, otherwise. Therefore, min-plus convolution of burst delay and affine function is given as $\delta_T \otimes \gamma_{b,r}(t) = b + r(t - T)$.

IV. Analysis

In this section, we propose and prove the propositions needed for analyzing performance of VBR flows in a FIFO multiplexing network. We consider a class of service curves, namely pseudoaffine curves [5], which is a multiple affine curve shifted to the right and given by $\beta = \delta_T \otimes [\otimes_{1 \leq x \leq n} \gamma_{\sigma_x, \rho_x}] = \delta_T \otimes [\wedge_{1 \leq x \leq n} \gamma_{\sigma_x, \rho_x}]$, where the non-negative term T is denoted as *offset*, and the affine curves between square brackets as leaky-bucket stages. In fact, a pseudoaffine curve represents the service received by single flows in tandems of FIFO multiplexing rate-latency nodes. It is clear that a rate-latency service curve is in fact pseudoaffine, since it can be expressed as $\beta = \delta_T \otimes \gamma_{0,R}$.

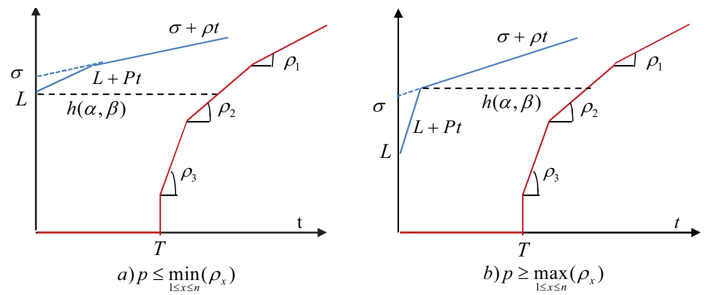

Fig. 1. Computation of delay bound for one VBR flow served by a pseudo affine curve

We now propose a proposition for computing delay bound as follows.

Proposition 1. *(Delay Bound) Let β be a pseudo affine curve, with offset T and n leaky-bucket stage $\gamma_{\sigma_x,\rho_x}, \quad 1 \leq x \leq n$, this means we have $\beta = \delta_T \otimes [\otimes_{1\leq x\leq n}\gamma_{\sigma_x,\rho_x}] = \delta_T \otimes [\wedge_{1\leq x\leq n}\gamma_{\sigma_x,\rho_x}]$ and let $\alpha = min(L+pt, \sigma+\rho t) = \gamma_{L,p} \wedge \gamma_{\sigma,\rho}$. If $\rho^*_\beta \geq \rho$ ($\rho^*_\beta = min_{1\leq x\leq n}\rho_x$), then the maximum delay for the flow is bounded by*

$$h(\alpha,\beta) = T + \left[\vee_{1\leq x\leq n}\frac{L-\sigma_x+\theta\left(p-\rho_x\right)^+}{\rho_x}\right]^+ \quad (1)$$

where $\theta_j = (\sigma_j - L_j)/(p_j - \rho_j)$.

Proof. *As stated before in Theorem 1, the delay is bounded by the maximum horizontal deviation between the arrival and service curves. Thus, due to Fig. 1, if $p \leq min_{1\leq x\leq n}(\rho_x)$, we have:*

$$\begin{cases} L = \sigma_1 + \rho_1\left(t_1 - T\right) \Rightarrow t_1 = T + \dfrac{L-\sigma_1}{\rho_1} \\ L = \sigma_2 + \rho_1\left(t_2 - T\right) \Rightarrow t_2 = T + \dfrac{L-\sigma_2}{\rho_2} \\ \vdots \quad \vdots \quad \vdots \\ L = \sigma_n + \rho_n\left(t_n - T\right) \Rightarrow t_n = T + \dfrac{L-\sigma_n}{\rho_n} \end{cases} \quad (2)$$

$$\Rightarrow h(\alpha,\beta) = max_{1\leq x\leq n}t_x = T + \left[\vee_{1\leq x\leq n}\frac{L-\sigma_x}{\rho_x}\right]^+ \quad (3)$$

If $p \geq max_{1\leq x\leq n}(\rho_x)$, due to Fig. 1, we have:

$$\begin{cases} L + p\theta = \sigma_1 + \rho_1\left(t_1 + \theta - T\right) \\ \qquad \Rightarrow t_1 = T + \dfrac{L+p\theta-\sigma_1}{\rho_1} - \theta \\ L + p\theta = \sigma_2 + \rho_2\left(t_2 + \theta - T\right) \\ \qquad \Rightarrow t_2 = T + \dfrac{L+p\theta-\sigma_2}{\rho_2} - \theta \\ \vdots \quad \vdots \quad \vdots \\ L + p\theta = \sigma_n + \rho_n\left(t_n + \theta - T\right) \\ \qquad \Rightarrow t_n = T + \dfrac{L+p\theta-\sigma_n}{\rho_n} - \theta \end{cases}$$

$$\Rightarrow h(\alpha,\beta) = max_{1\leq x\leq n}t_x = T + \left[\vee_{1\leq x\leq n}\frac{L+p\theta-\sigma_x}{\rho_x} - \theta\right]^+$$

$$= T + \left[\vee_{1\leq x\leq n}\frac{L-\sigma_x+\theta\left(p-\rho_x\right)}{\rho_x}\right]^+ \quad (4)$$

From Eq. 2 and 4, we can say:

$$h(\alpha,\beta) = T + \left[\vee_{1\leq x\leq n}\frac{L-\sigma_x+\theta\left(p-\rho_x\right)^+}{\rho_x}\right]^+ \quad (5)$$

In Propositions 2 and 3, we obtain ESC with FIFO multiplexing under different assumptions.

Proposition 2. *(Equivalent Service Curve) Let β be a pseudo affine curve as $\beta = \delta_T \otimes [\otimes_{1\leq x\leq n}\gamma_{\sigma_x,\rho_x}] = \delta_T \otimes [\wedge_{1\leq x\leq n}\gamma_{\sigma_x,\rho_x}]$ and let $\alpha = min(L+pt, \sigma+\rho t) = \gamma_{L,p} \wedge \gamma_{\sigma,\rho}$. If $\rho^*_\beta \geq \rho$ ($\rho^*_\beta = min_{1\leq x\leq n}\rho_x$) and $p \geq \rho^\circ_\beta$ ($\rho^\circ_\beta = max_{1\leq x\leq n}\rho_x$), then the equivalent service curve is obtained by subtracting arrival curve α, $\{\beta^{eq}(\alpha,\tau), \quad \tau = h(\alpha,\beta)\} \equiv \beta^{eq}(\alpha)$, with*

$$\beta^{eq}(\alpha) = \delta_{T+\vee_{1\leq i\leq n}\left[\frac{L-\sigma_i+\theta(p-\rho_i)^+}{\rho_i}\right]^+ +\theta} \otimes \left[\otimes_{1\leq x\leq n}\left[\gamma_{\rho_x\left\{\vee_{1\leq i\leq n}\left[\frac{L-\sigma_i+\theta(p-\rho_i)^+}{\rho_i}\right]^+ - \frac{\sigma-\sigma_x-(\rho_x-\rho)\theta}{\rho_x}\right\},\rho_x-\rho}\right]\right] \quad (6)$$

Proof. *Let us apply Theorem 2 to service curve β as follows.*

$$\begin{aligned} \beta^{eq}(\alpha,\tau) = & [\delta_T \otimes [\otimes_{1\leq x\leq n}\gamma_{\sigma_x,\rho_x}] \\ & -min\left(L+p\left(t-\tau\right), \sigma+\rho\left(t-\tau\right)\right)] \end{aligned} \quad (7)$$

Eq. (7) is wide-sense increasing for any $\tau \geq 0$. Since we assumed $\tau = h(\alpha,\beta)$, due to Proposition 1, we have:

$$\tau = T + \left[\vee_{1\leq x\leq n}\frac{L-\sigma_x+\theta\left(p-\rho_x\right)^+}{\rho_x}\right]^+ \quad (8)$$

Without losing generality, we follow proof for $n = 1$. Therefore, by Eq. (8) we have:

$$\tau - T = \left[\frac{L-\sigma_x+\theta\left(p-\rho_x\right)^+}{\rho_x}\right]^+ \quad (9)$$

We then apply Theorem 2 to service curve $\acute{\beta}$ ($\acute{\beta}$ is β when $n = 1$) as follows.

$$\begin{aligned} \acute{\beta^{eq}}(\alpha,\tau) &= \delta_T \otimes \gamma_{\sigma_x,\rho_x} - min\left(L+p\left(t-\tau\right), \sigma+\rho\left(t-\tau\right)\right) \\ &= \sigma_x + \rho_x\left(t-T\right) - min\left(L+p\left(t-\tau\right), \sigma+\rho\left(t-\tau\right)\right) \end{aligned} \quad (10)$$

We now consider two situations including $0 \leq t-\tau \leq \theta$ and $t-\tau > \theta$.

If $0 \leq t-\tau \leq \theta \Rightarrow min\left(L+p\left(t-\tau\right), \sigma+\rho\left(t-\tau\right)\right) = L+p\left(t-\tau\right)$. Let us assume $\acute{t} = t-\tau \Rightarrow t-T = \acute{t} + (\tau - T)$. From Eq. 9, we can say $t - T = \acute{t} + \left[\frac{L-\sigma_x+\theta(p-\rho_x)^+}{\rho_x}\right]^+$.

$$\begin{aligned} \acute{\beta^{eq}}(\alpha,\tau) &= \sigma_x + \rho_x\left(\acute{t} + \left[\frac{L-\sigma_x+\theta\left(p-\rho_x\right)^+}{\rho_x}\right]^+\right) \\ &\quad - \left(L + p\acute{t}\right) \\ &= \sigma_x + \rho_x\acute{t} + \left[L-\sigma_x+\theta\left(p-\rho_x\right)^+\right]^+ - L - p\acute{t} \\ &= -\left(p-\rho_x\right)\acute{t} + \theta\left(p-\rho_x\right)^+ \end{aligned}$$

Since $p \geq \rho^\circ_\beta$ and $\acute{t} \leq \theta$, we have:

$$\begin{aligned} \acute{\beta^{eq}}(\alpha,\tau) &= -\left(p-\rho_x\right)\acute{t} + \theta\left(p-\rho_x\right)^+ \\ &\leq -\left(p-\rho_x\right)\theta + \theta\left(p-\rho_x\right) \leq 0 \end{aligned}$$

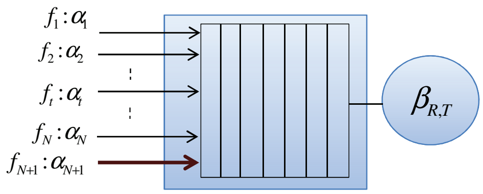

Fig. 2. Computation of ESC for flow $N+1$ in a rate-latency node

Therefore, $\acute{\beta^{eq}}(\alpha,\tau)=0$ where $0\le t-\tau\le\theta$. By definition of the service curve, we can say that if $0\le t\le\theta+\tau$ then $\acute{\beta^{eq}}(\alpha,\tau)=0$, and this means that the offset of $\acute{\beta^{eq}}(\alpha,\tau)$ is equal to $\tau+\theta$.

If $t-\tau>\theta \Rightarrow min\left(L+p\left(t-\tau\right),\sigma+\rho\left(t-\tau\right)\right)=\sigma+\rho\left(t-\tau\right)$. Therefore, $\acute{\beta^{eq}}(\alpha,\tau)=\sigma_x+\rho_x\left(t-T\right)-\left(\sigma+\rho\left(t-\tau\right)\right)$. If $\rho_x\tau$ is added to and subtracted from $\acute{\beta^{eq}}(\alpha,\tau)$, we have

$$\begin{aligned}\acute{\beta^{eq}}(\alpha,\tau)&=\sigma_x+\rho_x\left(t-T\right)-\left(\sigma+\rho\left(t-\tau\right)\right)+\rho_x\tau-\rho_x\tau\\&=\sigma_x-\sigma+\rho_x\left(\tau-T\right)+\left(\rho_x-\rho\right)\left(t-\tau\right)\\&=\delta_\tau\otimes\gamma_{\sigma_x-\sigma+\rho_x(\tau-T),\rho_x-\rho}\end{aligned}\tag{11}$$

Since we concluded that the offset of $\acute{\beta^{eq}}(\alpha,\tau)$ is $\tau+\theta$, we add $\left(\rho_x-\rho\right)\theta$ to Eq. 11 and then subtract it. We obtain:

$$\begin{aligned}\acute{\beta^{eq}}(\alpha,\tau)=&\ \sigma_x-\sigma+\rho_x\left(\tau-T\right)+\left(\rho_x-\rho\right)\left(t-\tau\right)\\&+\left(\rho_x-\rho\right)\theta-\left(\rho_x-\rho\right)\theta\\=&\ \sigma_x-\sigma-\rho\theta+\rho_x\left(\tau+\theta-T\right)+\left(\rho_x-\rho\right)\left(t-\tau-\theta\right)\\=&\ \delta_{\tau+\theta}\otimes\gamma_{\sigma_x-\sigma-\rho\theta+\rho_x(\tau+\theta-T),\rho_x-\rho}\end{aligned}\tag{12}$$

Thus, the offset of $\beta^{eq}(\alpha,\tau)$ is equal to $\tau+\theta$. Furthermore, each leaky bucket-stage in $\beta^{eq}(\alpha,\tau)$ can be computed as $\gamma_{\acute{\sigma_j},\acute{\rho_j}}$, with $\acute{\sigma_j}=\sigma_x-\sigma-\rho\theta+\rho_x\left(\tau+\theta-T\right)$ and $\acute{\rho_j}=\rho_j-\rho$. Therefore, we have $\beta^{eq}=\delta_{\tau+\theta}\otimes\left[\otimes_{1\le x\le n}\gamma_{\acute{\sigma_x},\acute{\rho_x}}\right]$ and by substituting (8) into β^{eq}, we prove the proposition.

We can specifically capitalize on Proposition 2 to obtain a parametric expression. We assume that the number of flows passing through a rate-latency node is $N+1$. Therefore, for computing ESC for the tagged flow, we should subtract the arrival curves of other N flows. It can be calculated by iteratively applying Proposition 2 for N times. Without any loss of generality, we presume that the tagged flow is flow $N+1$. We now present following proposition:

Proposition 3. *(Equivalent Service Curve for Rate-Latency Service Curve With $N+1$ Flows) Consider one node with a rate-latency service curve $\beta_{R,T}=\delta_T\otimes\gamma_{0,R}$. Let $\alpha_i=min(L_i+p_it,\sigma_i+\rho_it)=\gamma_{L_i,p_i}\wedge\gamma_{\sigma_i,\rho_i}$ be arrival curve of flow i and $p_i\ge R-\sum_{j=1}^{i-1}\rho_j$, where $1\le i\le N+1$ and $N+1$ is the number of flows passing through that node as shown in Fig. 2. The equivalent service curve for flow $N+1$ in the node, obtained by subtracting N arrival curves, is:*

$$\beta^{eq}_{N+1}=\delta_{T+\sum_{i=1}^{N}\left(\left[\frac{L_i+\theta_i(p_i-R+\sum_{j=1}^{i-1}\rho_j)^+}{R-\sum_{j=1}^{i-1}\rho_j}\right]^++\theta_i\right)}\otimes\gamma_{0,R-\sum_{j=1}^{N}\rho_j}\tag{13}$$

Proof. *We use the simplest form of mathematical inductive proof method. It proves that a statement involving a number N holds for all values of N. The proof consists of two steps:*

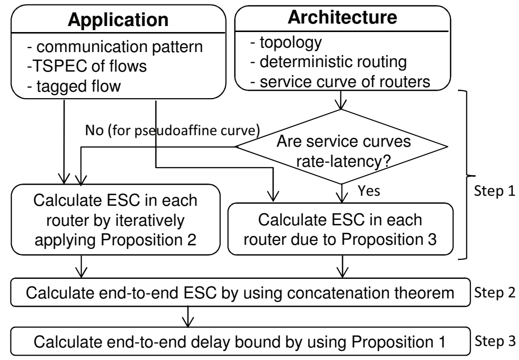

Fig. 3. End-to-end delay bound analysis flow

Base Step: *In this step, we show that the statement holds when $N=1$. In order to verify this, we compute the ESC obtained by subtracting one arrival curve ($N=1$), offered by Proposition 3:*

$$\beta^{eq}_2=\delta_{T+\left[\frac{L_1+\theta_1(p_1-R)^+}{R}\right]^++\theta_1}\otimes\gamma_{0,R-\rho_1}\tag{14}$$

If we apply Proposition 2 for a rate-latency service curve $\beta_{R,T}$ where $n=1$, $\sigma_x=0$ and $\rho_x=R$, Eq. 14 is easily obtained. Therefore, the statement holds when $N=1$.

Inductive Step: *In this step, we show if the statement holds for some N, then the statement also holds when $N+1$ is substituted for N. Assume that β^{eq}_{N+1} is an ESC for flow $N+1$, obtained by subtracting N arrival curves as represented in Eq. 13. We shall compute ESC β^{eq}_{N+2} for flow $N+2$. Therefore, in this case we should subtract $N+1$ arrival curves. After subtracting N arrival curves, the ESC for aggregated flow $\{N+1,N+2\}$ will be equal to β^{eq}_{N+1}. Therefore, for computing β^{eq}_{N+2}, it is enough to subtract flow $N+1$ from β^{eq}_{N+1} by applying Proposition 2.*

From β^{eq}_{N+1}, we can say n, ρ_x, σ_x and T_x in Proposition 2 are as $n=1$, $\rho_x=R-\sum_{j=1}^{N}\rho_j$, $\sigma_x=0$, and $T_x=T+\sum_{i=1}^{N}\left[\frac{L_i+\theta_i(p_i-R+\sum_{j=1}^{i-1}\rho_j)^+}{R-\sum_{j=1}^{i-1}\rho_j}\right]^++\sum_{j=1}^{N}\theta_j$. Also, α in Proposition 2 is equal to $\alpha_{N+1}=min(L_{N+1}+p_{N+1}t,\sigma_{N+1}+\rho_{N+1}t)$. After applying Proposition 2 and computing some straightforward algebraic manipulation, β^{eq}_{N+2} is given by:

$$\beta^{eq}_{N+2}=\delta_{T+\sum_{i=1}^{N+1}\left(\left[\frac{L_i+\theta_i(p_i-R+\sum_{j=1}^{i-1}\rho_j)^+}{R-\sum_{j=1}^{i-1}\rho_j}\right]^++\theta_t\right)}\otimes\gamma_{0,R-\sum_{j=1}^{N+1}\rho_j}\tag{15}$$

which proves the inductive step.

Fig. 3 shows the overall analysis flow for computing end-to-end delay bound of a tagged flow under the mentioned system model. We illustrate the steps with an example in section V.

V. Numerical Example

To show how the proposed propositions are used, we applied them to a simple example depicted in Fig. 4. The figure depicts a network with 4 flows and 3 routers which serve flows in the FIFO order. f_3 is the tagged flow and f_1, f_2 and f_4 are interfering flows. Flows follow TSPEC, $f_1\propto(1,1,2,0.128)$, $f_2\propto(1,1,2,0.032)$, $f_3\propto(1,1,4,0.256)$, and $f_4\propto(1,1,2,0.008)$. Each router guarantees the service curve of $\beta_{R,T}(t)=\delta_T\otimes\gamma_{0,R}=1(t-1)^+$, where the serving rate $R=1$ *flit/cycle* and the processing latency $T=1$ *cycle*.

A. Computation of the end-to-end equivalent service curve

Step 1: We first calculate the ESC for the tagged flow in each node. Then, we can model a flow passing through a series of routers as a series of concatenated pseudoaffine servers. Before that, θ_j is computed for each flow f_j as $\theta_1 = (\sigma_1 - L_1)/(p_1 - \rho_1) = (2-1)/(1-0.128) = 1.146$, $\theta_2 = 1.033$, $\theta_3 = 4.032$, and $\theta_4 = 1.008$.

We use sub-index "(j, r_i)" for notations to indicate that they are related to flow j in router i. For example, $\beta^{eq}_{(j,r_i)}$ denotes the ESC of flow j in router i.

From Proposition 3, we obtain the ESC for f_3 in node 1 by subtracting arrival curves of f_1 and f_2. The serving rate and latency for aggregate flow $f_{(1,2,3)}$ in node 1 is equal to $R_1 = 1$ and $T_1 = 1$, respectively. Therefore, we have $T^{eq}_{(3,r_1)} = T_1 + \left(\left[\frac{L_1+\theta_1(p_1-R_1)^+}{R_1}\right]^+ + \theta_1\right) + \left[\frac{L_2+\theta_2(p_2-R_1+\rho_1)^+}{R_1-\rho_1}\right]^+ + \theta_2 = 5.477$, $\rho^{eq}_{(3,r_1)} = R_1 - \rho_1 - \rho_2 = 0.84$, and $\sigma^{eq}_{(3,r_1)} = 0$.

$$\Rightarrow \beta^{eq}_{(3,r_1)} = \delta_{5.477} \otimes \gamma_{0,0.84} \tag{16}$$

This Proposition also allows computing the ESC for f_3 in node 2 by subtracting arrival curve of flow f_4, as well. $T^{eq}_{(3,r_2)} = T_2 + \left(\left[\frac{L_4+\theta_4(p_4-R_2)^+}{R_2}\right]^+ + \theta_4\right) = 3.008$, $\rho^{eq}_{(3,r_2)} = R_2 - \rho_4 = 0.992$, and $\sigma^{eq}_{(3,r_2)} = 0$.

$$\Rightarrow \beta^{eq}_{(3,r_2)} = \delta_{3.008} \otimes \gamma_{0,0.992} \tag{17}$$

Since there is no interfering flow in node 3, the ESC of flow 3 in this node is equal to

$$\beta^{eq}_{(3,r_3)} = \sigma_1 \otimes \gamma_{0,1} \tag{18}$$

Step 2: We use the theorem of concatenation of nodes [3] for obtaining the equivalent end-to-end service curve. Given is a flow traversing two nodes sequentially connected and each node is offering a service curve β_i, $i = 1, 2$ to the flow. Then the concatenation of the two nodes offers a service curve of $\beta_1 \otimes \beta_2$ to the flow. Thus, β^{eq}_3 is given by

$$\beta^{eq}_3 = \beta^{eq}_{(3,r_1)} \otimes \beta^{eq}_{(3,r_2)} \otimes \beta^{eq}_{(3,r_3)} \tag{19}$$
$$= \delta_{5.477+3.008+1} \otimes [\gamma_{0,0.84} \wedge \gamma_{0,0.992} \wedge \gamma_{0,1}] = \delta_{9.485} \otimes \gamma_{0,0.84}$$

B. Computation of the end-to-end delay bound

Step 3: According to Proposition 1 and Eq. 19, the maximum delay for flow 3 is bounded by

$$h(\alpha_3, \beta^{eq}_3) = 9.485 \vee \left[\left(\frac{1-0+4.032(1-0.84)^+}{0.84}\right)^+, \left(\frac{1-0+4.032(1-0.992)^+}{0.992}\right)^+, \left(\frac{1-0+4.032(1-1)^+}{1}\right)^+\right]$$
$$= 9.485 + max(1.958, 1.04, 1) = 11.443 \tag{20}$$

Here if we only use (σ, ρ) instead of TSPEC, each flow j would be constrained by arrival curve $\alpha_j = \sigma_j + \rho_j t = \gamma_{\sigma_j,\rho_j}$. Therefore, flows in the example are represented as $f_1 \propto (2, 0.128)$, $f_2 \propto (2, 0.032)$, $f_3 \propto (4, 0.256)$, and $f_4 \propto (2, 0.008)$. We then follow the stages of computing individual delay bound for a tagged flow as stated before. For this purpose, we can easily revise our proposed propositions for (σ, ρ) flows by substituting σ and ρ into L and p, respectively, in all formulas. We can also apply the method presented in [5]. With both approaches, the same value for $h(\alpha_3, \beta^{eq}_3)$ is achieved and equals to 17.241. Thus, we have about 33.6% improvement on the tightness of the delay bound.

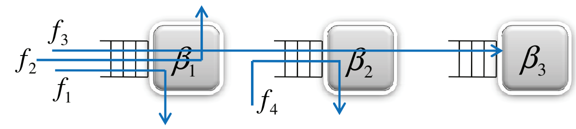

Fig. 4. An example

To analyze delay sensitivity, Table I depicts the end-to-end delay bound for tagged flow f_3 in a network with CBR (Constant Bit-Rate) flows ($Delay_{CBR}$) and also VBR flows ($Delay_{VBR}$) versus the different values of service rate R, along with values for the end-to-end equivalent service rate R^{eq}_3. From this table, it is clear that the end-to-end equivalent service rate, R^{eq}_3, is decreasing by reducing R, while the end-to-end delay bounds are increasing as well. Also, it is worth mentioning that the *improvement percentage (ImP)* decreases because of reduction of R^{eq}_3.

TABLE I
END-TO-END DELAY COMPARISON FOR f_3 UNDER DIFFERENT SERVICE RATES

	$R_1 = 1$	$R_2 = 0.7$	$R_3 = 0.5$
R^{eq}_3	0.84	0.54	0.34
$Delay_{CBR}$	17.241	22.804	31.327
$Delay_{VBR}$	11.443	17.773	27.541
Improvement Percentage	33.6%	22%	12%

VI. CONCLUSIONS

We have presented and proved the required propositions for computing delay bound of VBR flows in a FIFO multiplexing network. The propositions can be applied for an architecture based on aggregate scheduling. To exemplify the potential of our technique, derivation of formulas for computing equivalent service curve and the delay bound is detailed. In the future, we will apply our formal approach for performance analysis of concatenated routers with multiple virtual channels per inport.

ACKNOWLEDGMENT

The research is funded in part by Intel Corporation through a research gift.

REFERENCES

[1] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with h.264/avc: Tools, performance and complexity", *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7-28, 2004.

[2] C. Chang, *Performance Guarantees in Communication Networks*, Springer-Verlag, 2000.

[3] J. Y. L. Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*, (LNCS, vol. 2050). Berlin, Germany: Springer-Verlag, 2004.

[4] Y. Jiang, "Delay bounds for a network of guaranteed rate servers with FIFO aggregation", *Computer Networks*, vol. 40, no. 6, pp. 683-694, 2002.

[5] L. Lenzini, L. Martorini, E. Mingozzi, and G. Stea, "Tight end-to-end per-flow delay bounds in fifo multiplexing sink-tree networks", *Performance Evaluation*, vol. 63, no. 9, pp. 956-987, 2006.

[6] Y. Qian, Z. Lu, and Q. Dou, "QoS Scheduling for NoCs: Strict Priority Queueing versus Weighted Round Robin", *In the Proceedings of the 28th International Conference on Computer Design (ICCD'10)*, pp. 52-59, Amerstedam, The Netherlands, 2010.

[7] F. Jafari, A. Jantsch, Z. Lu, "Output process of variable bit-rate flows in on-chip networks based on aggregate scheduling", *In the Proceedings of the International Conference on Computer Design (ICCD)*, pp. 445-446, Amherst, USA, 2011.