A New Deterministic Traffic Model for Core-Stateless Scheduling

Gang Cheng, Li Zhu, and Nirwan Ansari

***Abstract*—Core-stateless scheduling algorithms have been proposed in the literature to overcome the scalability problem of the stateful approach. Instead of maintaining per-flow information or performing per-packet flow classification at core routers, packets are scheduled according to the information (time stamps) carried in their headers. They can hence provision quality of service (QoS) and achieve high scalability. In this paper, which came from our observation that it is more convenient to evaluate a packet's delay in a core-stateless network with reference to its time stamp than to the real time, we propose a new traffic model and derive its properties. Based on this model, a novel time-stamp encoding scheme, which is theoretically proven to be able to minimize the end-to-end worst case delay in a core-stateless network, is presented. With our proposed traffic model, performance analysis in core-stateless networks becomes straightforward.**

***Index Terms*—Core-stateless network, quality of service (QoS), traffic model, traffic scheduling.**

I. Introduction

THE Internet is expected to accommodate a variety of applications with different quality-of-service (QoS) requirements, such as video conferencing, interactive TV, and Internet telephony, as it evolves into a globe commercial infrastructure. However, today's Internet only provides one simple service: best-effort datagram delivery, in which data packets may experience unpredictable delay and packet loss rate and arrive at the destination out of order. Hence, more sophisticated mechanisms are urgently needed to provide less oscillatory and more predictable services for various applications.

Two fundamental frameworks, namely, Integrated Services (Intserv) and Differentiated Services (Diffserv), have been proposed for this purpose. The Intserv approach [3]–[8], which aims to provide "hard" end-to-end QoS guarantees to each individual data flow, requires per-flow-based resource allocation and service provisioning and, thus, suffers from the scalability problem due to the huge number of data flows that may coexist in today's high-speed core routers. The proposed Diffserv model simplifies the design of core routers by aggregating individual flows at edge routers and provisioning only a number of services to the aggregated data flows at each core router. However, in this model, it is difficult to identify each individual flow's QoS requirements at core routers and to contrive efficient resource allocation mechanisms to guarantee the end-to-end QoS of each individual data flow. In addition, it has been shown [9] that, if all packets are served in a first-in-first-out fashion, the worst case delay bound is a function of the hop count and explodes at a certain utilization level. Thus, the overall utilization in such networks may be limited to a small fraction of its link capacities in order to provide guaranteed service delay. Various alternatives have been proposed in order to exploit the benefits of both Intserv and Diffserv and, at the same time, to mitigate their drawbacks. The operation of Intserv over the Diffserv model was introduced in [10]. In this model, the admission control and resource allocation procedures are adopted from those in the Intserv model so that sufficient resources can be reserved to satisfy the data flows' QoS requirements, while the data flows are served in the network domain in a Diffserv fashion, i.e., data flows are aggregated and provided only with a limited number of services. Along with two new classes of aggregated packet scheduling algorithms, the static earliest time first (SETF) and dynamic earliest time first (DETF), Zhang *et al.* [11] showed that the maximum allowable network utilization level can be greatly increased while the worst case delay bound is decreased if additional time-stamp information is encoded in the packet header. In [12], a core-stateless version of jitter virtual clock (CJVC), which achieves the same worst case delay bound as jitter virtual clock (JVC), has been proposed. Like JVC, CJVC is nonwork-conserving, i.e., the server may be free even if there are packets in the buffer and, therefore, the network resource may be underutilized. Capable of providing the same delay bound as the corresponding stateful Guaranteed Rate (GR) server, a methodology to transform stateful GR per-flow scheduling algorithms into core-stateless version ones was proposed [13]. Based on the methodology [13], the authors also proposed the core-stateless guaranteed throughput (CSGT) network architecture in [14], which is a work-conserving network architecture that provides throughput guarantees to flows over finite timescales without maintaining a per-flow state in core routers. In [15], a distributed admission control to support guaranteed services in core-stateless networks has been proposed. Based on the virtual time reference system [16], admission control under the bandwidth broker architecture has been studied in [17].

Paper approved by A. Pattavina, the Editor for Switching Architecture Performance of the IEEE Communications Society. Manuscript received August 5, 2003; revised October 8, 2004, and September 25, 2005. This work was supported in part by the New Jersey Commission on Higher Education via the NJ I-TOWER Project and the New Jersey Commission on Science and Technology via the NJ Center for Wireless Networks and Internet Security. This paper was presented in part at the IEEE GLOBECOM, San Francisco, CA, 2003.

G. Cheng was with the Advanced Networking Laboratory, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07012 USA. He is now with VPIsystems Corporation, Holmdel, NJ 07733 USA.

L. Zhu and N. Ansari are with the Advanced Networking Laboratory, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07012 USA (e-mail: nirwan.ansari@njit.edu).

Digital Object Identifier 10.1109/TCOMM.2006.873091

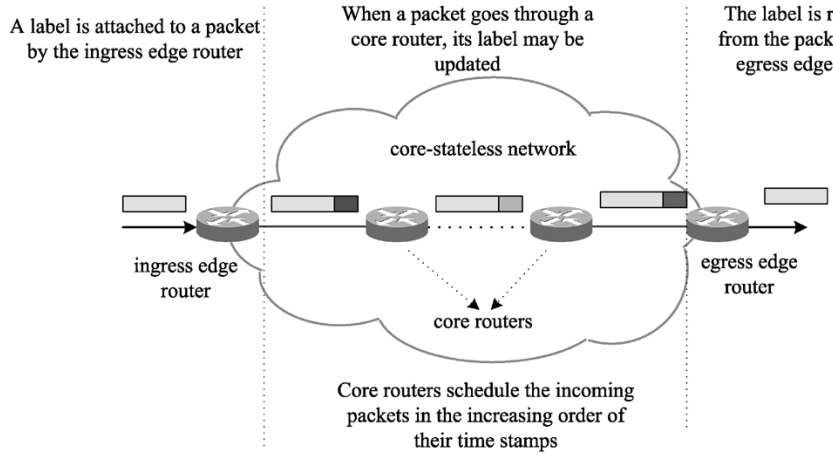

Fig. 1. Core-stateless network model.

Meanwhile, many core-stateless queueing schemes [18]–[23] have been proposed in literature, which are, however, beyond the scope of this paper.

In the literature, many traffic models have been proposed to characterize network traffic. Among them, the (σ, ρ) traffic model [24], owing to its simplicity and efficiency, has been widely adopted for the network performance analysis; here, the network performance analysis is referred to as the analysis of the worst case delay, worst case jitter, packet loss ratio, and so forth. In this paper, we show that the (σ, ρ) traffic model is not efficient for characterizing traffic in a core-stateless network. Instead, we introduce a new traffic model, which will be referred to as the (β, α) traffic model, which can better describe flows in a core-stateless network. Based on this model, three important issues are addressed: time-stamp encoding at the network edge, traffic pattern distortion in a core network, and worst case delay analysis.

The remainder of this paper is organized as the follows. We introduce the (β, α) traffic model and derive its properties in Sections II and III, respectively. Based on the model, a novel time stamp is presented in Section IV. Finally, the conclusion is provided in Section V.

II. $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ Traffic Model

A. Network Model

We first introduce the core-stateless network model adopted in this paper.

As shown in Fig. 1, routers in a core-stateless network are classified into two categories: edge routers and core routers. Namely, edge nodes are located at the network boundary, through which connections join and leave the network. Core nodes are located inside the network. When a packet arrives at the network boundary, the edge router will attach a label to the packet. The label includes some per-flow information such as the reserved bandwidth of the flow and a time stamp, which could be a function of the arrival time of the packet, the packet length, and the reserved bandwidth. The time stamp may be updated at each core router. The label will be removed from each packet after it traverses the network. At all routers, packets are served by the increasing order of their time stamps. Here, we adopt the Dynamic Earliest Time First (DETF) [11] scheduler[1] as an example and consider a special case: all flows injected to the core-stateless network are constant bit rate (CBR),[2] and the propagation delay and link capacity of any link are 0 and c, respectively. Here, the sequence of packets transmitted by a source to a destination is referred to as a flow [25], and we assume that the path is predetermined and fixed throughout its duration. Using the DETF scheduler, the worst case delay of flow i at any router is no larger than $(L_{\max}/c) + (L_i/\alpha_i)$ if the following conditions exist.

1) At the ingress edge router, packet k flow i is attached with a time stamp of $A_i^k + (L_i/\alpha_i)$, where α_j, L_i, and A_i^k are the input rate, packet length, and the arrival time of packet k at the ingress edge router of flow i respectively.
2) At a core router, the time stamp of packet k of flow i is updated with an increment of $(L_{\max}/c) + (L_i/\alpha_i)$, and packets are served at the increasing order of their time stamps, where $L_{\max}$ is the maximum packet length of all flows.

It should be noted that, in order to update time stamps by core routers, the per-flow information α_j should also be carried by the packets of flow i throughout the core-stateless network. On the other hand, since each router has the per-flow information in a stateful network, it is not necessary to attach packets with labels to provide guaranteed services. Even though this example considers an extreme case, it provides us an insight as to how a core-stateless network operates.

B. (σ, ρ) Traffic Model and Assumptions

In literature, the (σ, ρ) traffic model [24] has been widely adopted for characterizing traffic in a network, i.e., if the total traffic of a flow $F(t_1, t_2)$ arriving in the time interval $(t_1, t_2]$ is bounded by

$$F(t_1, t_2) \leq \sigma + \rho(t_2 - t_1) \tag{1}$$

[1]DETF is an output queuing scheduler that does not perform traffic shaping and reshaping inside the network.

[2]We assume that the total arrival rate (at the network edge) of the flows that share a same link is less than the corresponding link's capacity.

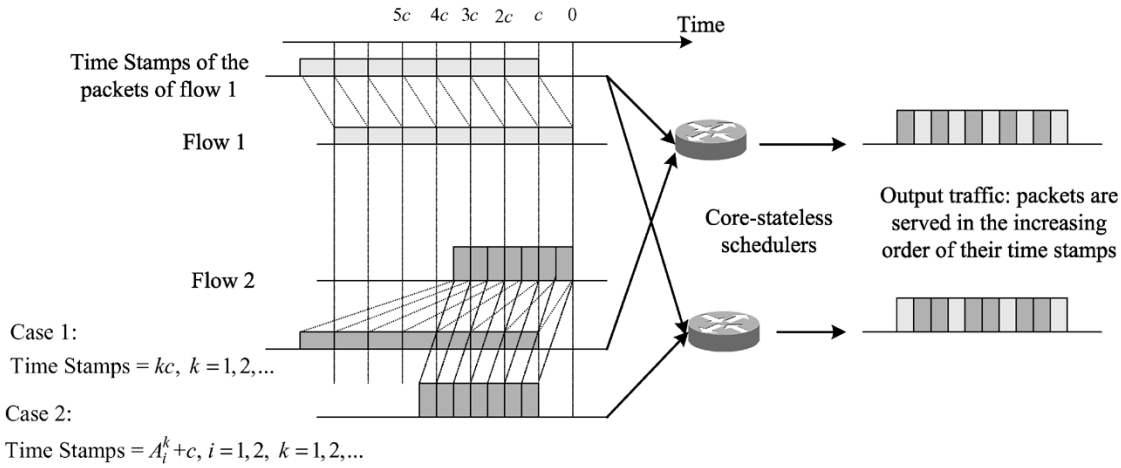

Fig. 2. Illustrative example.

then this flow is referred to as conforming to the traffic parameter (σ, ρ). In other words, we can claim that a flow conforms to the traffic parameter (σ, ρ) if no overflow occurs when this flow is injected to a leaky bucket with parameters of σ (buffer size) and ρ (output rate). We can also view ρ as the long-term average traffic rate bound of the flow and σ as its burst bound.

We adopt the following assumptions in this paper.

1) We only consider an arbitrary network topology with links and switches where each link is associated with a delay bound (propagation delay) and each switch is nonblocking.
2) A packet is considered "arrived" only after its last bit has arrived, and the delivery time of a packet at a node is the time when the last bit of the packet leaves the node.
3) Since a packet will only be delayed at a node if there is a packet being served or there are packets waiting in the buffer with earlier time stamps, we assume that the start time of each busy period is initialized at 0. Here, a busy period is an interval of time during which the transmission queue of the output link is continuously backlogged.
4) We assume that the time stamp of each packet lags behind its arrival time at any given node. This assumption may not hold for some core-stateless scheduling algorithms. However, note that packets are served by the order of their time stamps; the delivery order of packets will not change (thus, the delay for each packet to traverse the network remains the same) if the time stamps of all packets are increased by a constant D at the network boundary. Therefore, if D is large enough (for example, let D be the worst case delay of any packet through a given network, if such a bound exists), our assumption can be satisfied. We assume that, if the burst of each flow is bounded and the capacity of any link is no less than the average rate of the flows traversing the link, there exists a worst case delay bound in the network, i.e., the worst case delay of a flow to traverse any pair of nodes in the network with a limited number of hops is bounded. The framework proposed in this paper is only applicable to a work-conserving core-stateless network with bounded delay.

C. (β, α) Traffic Model

In a stateful network, packets are served by the order of their time stamps. Note that per-flow information is maintained at core nodes in the stateful network, and the performance parameters of each flow are static. Therefore, only one time parameter (arrival time) associated with each packet is enough for performance analysis in the stateful network, i.e., given the arrival times and sizes of all packets, the delivery time of each packet can be derived, and thus the worst-case delay and jitter of each flow can be computed. However, in a core-stateless network, per-flow information is not maintained in core nodes, and packets in the buffer are served by the order of their time stamps, not their arrival times. There is also no distinct relation between the time stamp of a packet and its arrival time. Consider the following example.

As shown in Fig. 2, two CBR flows 1 and 2 are contending for the bandwidth of a link with a capacity of $2L/c$. The reserved bandwidths of the two flows are both L/c, and all packets are of size L. However, the inter-arrival times of two consecutive packets of flows 1 and 2 are c and $c/2$, respectively. Assume that the first packets of both flows arrive at time 0, and the arrival time of the kth packet of flow i, $i = 1, 2$, is A_i^k, where $A_i^k = (k-1)c$ if $i = 1$, and $A_i^k = (k-1)c/2$ if $i = 2$. The time stamp attached to the kth packet of flow i is, however, kc, which is independent of i and will make each flow attain its reserved bandwidth. Therefore, it can be observed that the worst case delay of flow 1 is c, and it is infinity for flow 2. However, if the time stamp of the kth packet of flow i, $i = 1, 2$, is set to $A_i^k + c$, then the worst case delays of both flows become infinity. With (σ, ρ), the worst case delay of the aggregated traffic (the aggregation of flows 1 and 2), which is infinity, can be computed. However, we cannot tell which flow will experience such delay. Therefore, instead of using the (σ, ρ) traffic model, we will develop another traffic model to characterize traffic in a core-stateless network that could enable us to easily compute the worst case delay of all packets with respect to their time stamps. Moreover, from the point of view of a node, packets are served only by the order of their time stamps, and their arrival times seem irrelevant. Thus, a packet with an earlier time stamp than another packet, though it arrives later, may be served first.

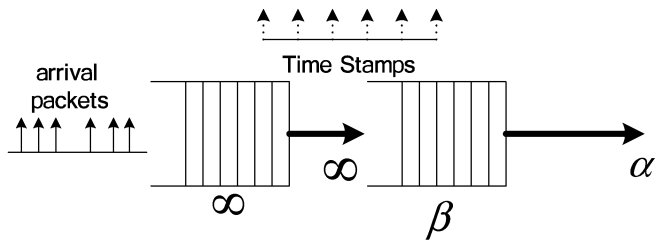

Fig. 3. Intuition of the (β, α) traffic model.

Thus, it is more reasonable to evaluate a packet's delay with reference to its time stamp, which is referred to as the *virtual delay* of a packet, rather than merely its arrival time. Therefore, a new mechanism to characterize traffic in the core-stateless network is necessary.

Since we evaluate the delay of a packet with reference to its time stamp, an intuitive idea to characterize a flow in the core-stateless network is to define a parameter (β, α) such that the total traffic of the flow of packets, whose time stamps are in the range of $(t_1, t_2]$, is no larger than $\beta + \alpha(t_2 - t_1)$, which is similar to the (σ, ρ) traffic model. Assume that packets are ordered by their time stamps as $P_1, P_2, \ldots, P_k, \ldots$ ($R_i \geq R_j$, if $i > j$, where R_i is the time stamp of packet P_i). Equivalently, for any two packets P_m and $P_k (k \geq m)$, $\beta + \alpha(R_k - R_m) \geq \sum_{i=m}^{k} L_i$, where L_i is the size of P_i. In this case, the parameter for the aggregated traffic of flows 1 and 2 in the above example is (L, c). However, note that the virtual delay of each packet is 0, and the intuitive implication of the virtual traffic parameter (L, c) is that the worst case virtual delay of a packet (i.e., the worst case delay with reference to its time stamp) is L/c. A packet may receive service as long as there is no packet in the buffer when it arrives. Thus, it is necessary to take into account the arrival time of a packet to characterize traffic in the core-stateless network. Therefore, we define the virtual traffic parameter (β, α) $(\alpha > 0, \beta \geq 0)$ of a flow as follows: for any two packets P_k and P_m of this flow $(k \geq m \geq 1)$, $\beta + \alpha(R_k - \max\{R_{m-1}, \min\{A_m, A_{m+1}, \ldots, A_k\}\}) \geq \sum_{i=m}^{k} L_i$, where A_i is the arrival time of packet P_i, $i = 1, 2, \ldots$; we refer to $F(t_1, t_2) = \beta + \alpha(t_2 - t_1)$ in the time interval $(t_1, t_2]$ as the virtual traffic function of this flow with the virtual traffic parameter (β, α), and the traffic model for characterizing traffic in the core-stateless network with the virtual traffic parameter is referred to as the (β, α) traffic model. The intuition of our traffic model is illustrated in Fig. 3. As shown in Fig. 3, imagine that there exist two virtual concatenated buffers, whose sizes are infinity and β, respectively. The bandwidth between the two buffers is infinity, and the packets in the second buffer are served sequentially at a rate of α. When packets arrive, they are stored in the first buffer. At the times equal to their time stamps, they are moved to the second buffer. If the second buffer never overflows, we claim that the arriving traffic conforms to the (β, α) traffic model.

Our proposed traffic model, which is the (β, α) traffic model, is different from those proposed in the literature. A virtual reference system that has the virtual space property $R_{k+1} - R_k \geq L_{k+1}/\alpha$ is introduced in [16]. It can be observed that, only when $\beta = 0$, the (β, α) traffic model possesses the virtual space property. A scheduler is said to possess the coordinated multihop scheduling (CMS) property [26] if the following is true:

- $R_k = A_k + \delta_k$ at the entrance node;
- $R_k = R_{k-1} + \delta_{k-1}$ at a core node.

For these conditions, $\delta_k \in [\delta - \eta, \delta + \eta]$, and δ and η are two constants that may vary with different nodes and flows. Since we do not place any constraint on the difference of the time stamps of two consecutive packets ($|R_k - R_{k-1}|$ and $|R_k - A_k|$ could be infinity in our traffic model), the (β, α) traffic model does not possess the CMS property. Note that the time stamp is referred to as the priority index in [26].

III. Properties of the (β, α) Traffic Model

Since packets are served by the order of their time stamps and no per-flow information is maintained at core nodes, all packets are treated as if they belong to a single flow. Therefore, the performance analysis of an individual flow at a node can be achieved by analyzing the performance of the aggregated flow at this node; this can be facilitated with the knowledge of the aggregated flow's traffic parameter. It is well known that the aggregated traffic of two flows with traffic parameters of (σ_1, ρ_1) and (σ_2, ρ_2) in the (σ, ρ) traffic model, respectively, has the traffic parameter $(\sigma_1 + \sigma_2, \rho_1 + \rho_2)$. Here, we show that the aggregated traffic in a core-stateless network also possesses the same additive property by Theorem 1 with respect to the virtual traffic parameter.

Theorem 1: Given two flows with virtual traffic parameters (β_1, α_1) and (β_2, α_2), the virtual traffic parameter of the aggregated traffic of the two flows is $(\beta_1 + \beta_2, \alpha_1 + \alpha_2)$.

Proof: Assume that packets are ordered by their time stamps. Given any two packets P_k and $P_m (k \geq m)$ of the aggregated flow, assume packets $P_{i_1}, P_{i_2}, \ldots,$ and P_{i_n} ($i_1 < i_2 < \ldots < i_n$ and $n \leq (k - m + 1)$) belong to flow 1, and the rest of the packets $P_{j_1}, P_{j_2}, \ldots, P_{j_p}$ ($j_1 < j_2 < \ldots < j_p$ and $p \leq k - m + 1$) belong to the other flow. Thus, by definition of the virtual traffic parameter, we have

$$\beta_1 + \alpha_1 (R_{i_n} - \max\{R_{i_1 - 1}, \min\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}\}) \geq \sum_{s=i_1}^{i_n} L_s \tag{2}$$

$$\beta_2 + \alpha_2 (R_{j_p} - \max\{R_{j_1 - 1}, \min\{A_{j_1}, A_{j_2}, \ldots, A_{j_p}\}\}) \geq \sum_{s=j_1}^{j_p} L_s. \tag{3}$$

Since $\max\{R_{i_n}, R_{j_p}\} = R_k$, $\min\{\min\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}, \min\{A_{j_1}, A_{j_2}, \ldots, A_{j_p}\}\} = \min\{A_m, A_{m+1}, \ldots, A_k\}$ and $\min\{R_{i_1 - 1}, R_{j_1 - 1}\} = R_{m-1}\}$, we have

$$\begin{aligned}
&(\beta_1 + \beta_2) + (\alpha_1 + \alpha_2) \\
&\quad \times (R_k - \max\{R_{m-1}, \min\{A_m A_{m+1}, \ldots, A_k\}\}) \\
&\quad \geq [\beta_1 + \alpha_1 \\
&\qquad \times (R_{i_n} - \max\{R_{i_1 - 1}, \min\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}\})] \\
&\qquad + [\beta_2 + \alpha_2 \\
&\qquad\quad \times (R_{j_p} - \max\{R_{j_1 - 1}, \min\{A_{j_1}, A_{j_2}, \ldots, A_{j_p}\}\})] \\
&\quad \geq \sum_{s=m}^{k} L_s.
\end{aligned} \tag{4}$$

■

By Theorem 1, the virtual traffic function of an aggregated traffic can be derived provided that all of the virtual traffic functions of individual flows are known.

In Theorem 1, in order to derive the virtual traffic parameter of the aggregated flow, we assume that the traffic parameters of all individual flows are known. A connection's traffic can be characterized at the entrance to the network, but the traffic pattern may be distorted inside the network, and thus the source characterization is not applicable at a core node traversed by the connection. Moreover, as one of the major components for some core-stateless scheduling algorithms, such as DETF and GR, packets' time stamps are updated at core nodes, and that may also contribute to the traffic pattern distortion. From the viewpoint of the virtual traffic function, we provide Theorem 2 to analyze the variation of the traffic parameter of a flow in a core-stateless network.

Theorem 2: Assume that the traffic parameter of the input traffic of a flow at a node is (β, α) and the worst case virtual delay to traverse this node is D. The virtual traffic parameter of the output traffic of this flow is (β', α) if all of its packets are updated by an increment d at this node, where $\beta' = \max\{0, \alpha(D-d) + L_{\max}\} + \beta$.

Proof: Assume that packets are ordered by their delivery times, i.e., for packets P_k and P_m, $k \geq m$, $T_k \geq T_m$, where T_i, which is the delivery time of packet P_i, $i = 1, 2, \ldots$, is also the arrival time of P_i of the output traffic. Since the worst case virtual delay is D, for any packet P_i, $i = 1, 2, \ldots$, we have

$$T_i \leq R_i + D. \tag{5}$$

Furthermore, since the time stamp of each packet that has been delivered by node j is updated by d, and $\beta' = \max\{0, \alpha(D-d) + L_{\max}\} + \beta$, for any two packets k and m $(k \geq m \geq 1)$, we have

$$\begin{aligned}
&\beta' + \alpha\left[R_k + d - \max\{R_{m-1} + d, T_m\}\right] \\
&\quad \geq \beta' + \alpha\left[R_k + d - \max\{R_{m-1} + d, R_m + D\}\right] \\
&\quad \geq \min\left\{\beta + \alpha(R_k - R_{m-1}),\right. \\
&\qquad \left.\beta + L_{\max} + \alpha(R_k - R_m)\right\}.
\end{aligned} \tag{6}$$

By the definition of the virtual traffic function, we have

$$\begin{aligned}
&\beta + \alpha\left[R_k - \max\left\{R_{m-1}, \min[A_m, A_{m+1}, \ldots, A_k]\right\}\right] \\
&\qquad \geq \sum_{i=m}^{k} L_i \Rightarrow \alpha(R_k - R_{m-1}) \geq \sum_{i=m}^{k} L_i.
\end{aligned} \tag{7}$$

Thus, define $R'_i = R_i + D + (L_{\max}/\alpha)$ as the time stamp of packet P_i in the output traffic, $i = 1, 2, \ldots$, by (6) and (7) to yield

$$\begin{aligned}
&\beta + \alpha\left[R'_k - \max\left\{R'_{m-1}, \min[T_m, T_{m+1}, \ldots, T_k]\right\}\right] \\
&\quad \geq \min\left\{\beta + \alpha(R_k - R_{m-1}), \beta + L_{\max} + \alpha(R_k - R_m)\right\} \\
&\quad \geq \min\left\{\sum_{i=m}^{k} L_i, \sum_{i=m+1}^{k} L_i + L_{\max}\right\} \geq \sum_{i=m}^{k} L_i.
\end{aligned} \tag{8}$$

Therefore, the virtual traffic parameter of the output traffic of this flow is (β', α). ■

Lemma 1: Assume that the traffic parameter of the input traffic of a flow at a node is (β, α), and the worst case virtual delay to traverse this node is D. Assume that the propagation delay for a packet of this flow to transmit from node j to its next node is δ. The virtual traffic parameter of the input traffic of this flow at the next node is (β', α) if all of its packets are updated by an increment d at this node, where $\beta' = \max\{0, \alpha(D + \delta - d) + L_{\max}\} + \beta$.

Proof: From the perspective of a node, the worst case virtual delay of a flow is $D + \delta$ if there is no time stamp update at the previous node of this flow. Thus, by Theorem 2, the virtual traffic parameter of the input traffic of this flow at this node is (β', α) if all of its packets are updated by an increment d at the previous node, where $\beta' = \max\{0, \alpha(D+\delta-d)+L_{\max}\}+\beta$. ■

Based on the concept of the virtual traffic function and parameter and their properties, we shall next analyze and derive the worst case delay of a flow to traverse a node in a work-conserving core-stateless network, with the assumption that the virtual traffic function of the aggregated flow or all individual flows is known.

Theorem 3: Assume that the input traffic of a node consists of flows $1, 2, \ldots, v$, whose virtual traffic parameters are (β_i, α_i), respectively, and the capacity of the output link of this node is c, $c \geq \sum_{i=1}^{v} \alpha_i$. Let packets of each flow be ordered by their delivery times, and P_k^i represents the kth packet of flow i. Define $\theta_i = \max\{\min_{k \geq m > 1}\{R_{m-1}^i - \min\{A_m^i, A_{m+1}^i, \ldots, A_k^i\}\}, 0\}$, where R_m^i and A_m^i are the time stamp and arrival time of P_m^i. Thus, the worst case virtual delay at this node is bounded by

$$\frac{\sum_{i=1}^{v}(\beta_i - \alpha_i\theta_i) + L_{\max}}{c} \tag{9}$$

where $L_{\max}$ is the maximum size of a packet.

Proof: Let P_k, $k = 1, 2, \ldots$, represent the kth packet of an aggregated flow in which packets are ordered by their time stamps. For any packet P_k, assume m to be the largest integer $k > m > 0$ such that $R_k < R_m$ and $T_k > T_m$, where R_i and T_i are the time stamp and delivery time of P_i. Thus

$$R_m > R_k \geq R_i, \qquad \text{for all } m < i < k \tag{10}$$

$$T_k > T_i > T_m, \qquad \text{for all } m < i < k \tag{11}$$

i.e., packet P_m is transmitted before packets $P_{m+1}, \ldots, P_k$; however, its time stamp is larger than that of packets $P_{m+1}, \ldots, P_k$. Thus

$$\min\{A_{m+1}, \ldots, A_k\} > T_m - \frac{L_m}{c}. \tag{12}$$

Since $P_{m+1}, \ldots, P_k$ arrive after $T_m - (L_m/c)$ and depart before P_k, we have

$$T_k = T_m + \frac{\sum_{i=m+1}^{k} L_i}{c}. \tag{13}$$

Note that $R_i \geq A_i$ (Assumption 4) for all $i = 1, 2, \ldots$, and thus $R_k \geq R_i \geq A_i \geq T_m - (L_m/c)$ for $i = m+1, \ldots, k-$

1. Furthermore, according to the definition of the virtual traffic function, we have

$$\begin{aligned}\theta_i &= \max\left\{\min_{k\geq m>1}\left\{R^i_{m-1}-\min\left\{A^i_m,A^i_{m+1},\ldots,A^i_k\right\}\right\},0\right\}\\ &\Rightarrow \min\left\{A^i_m,A^i_{m+1},\ldots,A^i_k\right\}+\theta\\ &\leq \max\left\{R^i_{m-1},\min\left\{A^i_m,A^i_{m+1},\ldots,A^i_k\right\}\right\}\\ &\Rightarrow \beta_i+\alpha_i\left[R^i_k-\left(\min\left\{A^i_m,A^i_{m+1},\ldots,A^i_k\right\}+\theta_i\right)\right]\\ &\geq \sum_{j=1}^{k}L^i_j. \end{aligned} \tag{14}$$

Since packets $P_{m+1},\ldots,P_k$ comprise the packets of flows $1,2,\ldots,v$, then

$$\begin{aligned}&\sum_{i=m+1}^{k}L_i\\ &\quad\leq\sum_{i=1}^{v}\left\{\beta_i+\alpha_i\left[R_k-(\min\{A_m,A_{m+1},\ldots,A_k\}+\theta_i)\right]\right\}.\\ &\quad\leq\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)+\left(\sum_{i=1}^{v}\alpha_i\right)\left[R_k-\left(T_m-\frac{L_m}{c}\right)\right].\end{aligned} \tag{15}$$

From (13) and (15), we have

$$\begin{aligned}T_k&=T_m+\frac{\sum_{i=m+1}^{k}L_m}{c}\\ &\leq T_m\\ &\quad+\frac{\left(\sum_{i=1}^{v}\alpha_i\right)\left[R_k-\left(T_m-\frac{L_m}{c}\right)\right]+\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}\\ &\leq R_k+\frac{L_{\max}}{c}+\frac{\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}\\ &\Rightarrow T_k-R_k\leq\frac{L_{\max}}{c}+\frac{\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}.\end{aligned} \tag{16}$$

If there does not exist such m, then $P_1,\ldots,P_{k-1}$ all leave the node before P_k and thus we have

$$\begin{aligned}T_k=\frac{\sum_{i=1}^{k}L_i}{c}&\leq\frac{\left(\sum_{i=1}^{v}\alpha_i\right)R_k+\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}\\ \Rightarrow T_k-R_k&\leq\frac{\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}.\end{aligned} \tag{17}$$

Thus, the virtual delay is bounded by (9). ■

It is possible to tighten the virtual delay bound provided in Theorem 3. For example, if $(\sum_{i=1}^{v}\alpha_i/c)\to 0$, then the worst case delay bound in Theorem 3 would be $(\sum_{i=1}^{v}\beta_i+L_{\max})/c$. However, if $\theta=\min_i\{\theta_i\}$, and the time stamps of all packets are decreased by θ, then the virtual traffic functions of all flows remain the same. In this case, by Theorem 3, the worst case virtual delay bound also remains the same as $(\sum_{i=1}^{v}\beta_i+L_{\max})/c$. Note that the time stamps of all packets are decreased by θ, and the actual worst case virtual delay bound becomes $((\sum_{i=1}^{v}\beta_i+L_{\max})/c)-\theta$. Thus, the worst case virtual delay bound can be tightened by the following lemma.

Lemma 2: Let P_k $(k=1,2,\ldots)$ be the kth packet of the aggregated flow ordered by packets' time stamps. Define

$$\theta=\max\left\{\min_{k\geq m>1}\left\{R_{m-1}-\min\{A_m,A_{m+1},\ldots,A_k\}\right\},0\right\} \tag{18}$$

where R_i and A_i are the time stamp and arrival time of packet P_i, respectively. Thus, for any packet, its worst case virtual delay is bounded by

$$\frac{\sum_{i=1}^{v}\left[\beta_i-\alpha_i(\theta_i-\theta)\right]+L_{\max}}{c}-\theta \tag{19}$$

where θ_i, (β_i,α_i), v, and $L_{\max}$ have been defined in Theorem 3.

Proof: Note that, if all packets' time stamps are decreased or increased by a constant at the entrance to a node, their delivery times remain unchanged. Furthermore, if all packets' time stamps are decreased by $\theta=\max\{\min_{k\geq m>1}\{R_{m-1}-\min\{A_m,A_{m+1},\ldots,A_k\}\},0\}$, the virtual traffic function parameter also remains the same, while θ_i would be decreased by θ. Thus, in this case, by Theorem 3, for any packet P_k, we have

$$\begin{aligned}T_k-(R_k-\theta)&\leq\frac{\sum_{i=1}^{v}\left[\beta_i-\alpha_i(\theta_i-\theta)\right]+L_{\max}}{c}\\ \Rightarrow T_k-R_k&\leq\frac{\sum_{i=1}^{v}\left[\beta_i-\alpha_i(\theta_i-\theta)\right]+L_{\max}}{c}-\theta\end{aligned} \tag{20}$$

i.e., the worst case virtual delay is bounded by (19). ■

Therefore, if all packets' time stamps at node i are increased by $((\sum_{i=1}^{v}[\beta_i-\alpha_i(\theta_i-\theta)]+L_{\max})/c)-\theta$, then assumption 4) can be satisfied for all flows at the output port of this node. However, when the propagation delay is taken into account, the increment for flow $n, 1\leq n\leq v$, should be $((\sum_{i=1}^{v}[\beta_i-\alpha_i(\theta_i-\theta)]+L_{\max})/c)-\theta+\delta_{n,i}$, where $\delta_{n,i}$ is the propagation delay of flow n to traverse the link between node i and its next node.

For illustrative purposes, we shall demonstrate how to use our proposed traffic model and its properties to compute the worst case delays of the two flows in the earlier example in Section II. Two cases are considered in the example: the time stamps attached to the kth packet of flow 2 are kc and A^k_i+c, respectively. In the former case, the virtual traffic parameters of both flows are $(0, L/c)$. By Theorem 1, the virtual traffic parameter of the aggregated flow is $(0, 2L/c)$, and the virtual delay bound of any packet is $c/2$ $(\theta_1=\theta_2=0)$ by Theorem 3. Hence, since the time stamp of a packet lags behind its arrival time, bounded by c and infinity in flows 1 and 2, respectively, then the worst case delay of the two flows are $1.5c$ and infinity, respectively. In the latter case, the virtual traffic parameter of flow 2 is $(\infty, L/c)$. Hence, the virtual traffic parameter of the aggregated flow is $(\infty, 2L/c)$ and the worst case virtual delay is infinity. Therefore, the worst case delays of both flows become infinity.

IV. Time-Stamp Encoding Scheme

As the first step for a core-stateless network to deliver packets, time stamps are encoded in packets at network boundaries. Moreover, since we propose the (β,α) traffic model to

describe the traffic in a core-stateless network, it is necessary to investigate how to encode time stamps of packets of a flow that conforms to a given virtual traffic parameter. We will show later that a good time-stamp encoding mechanism can reduce the worst case delay of a flow traversing a core-stateless network.

Proposition: Denote A_i and L_i as the arrival time and the size of packet P_i, respectively. Assume that packets are ordered by their arrival times, i.e., $A_{i+1} > A_i$, $i = 1, 2, 3, \ldots$. Define $\beta_0 = 0$ and

$$\beta_i = \max\{0, \beta_{i-1} - \alpha(A_i - A_{i-1})\} + L_i. \tag{21}$$

If P_i's time stamp R_i is encoded by

$$R_i = \frac{\max\{\beta_i - \beta, 0\}}{\alpha} + A_i \tag{22}$$

then the virtual traffic parameter of this flow is (β, α).

Proof: Note that

$$\begin{aligned} R_i &= \frac{\max\{\beta_i - \beta, 0\}}{\alpha} + A_i \\ &= \frac{\max\{\max\{0, \beta_{i-1} - \alpha(A_i - A_{i-1})\} + L_i - \beta, 0\}}{\alpha} \\ &\quad + A_i \\ &\geq \frac{\max\{\beta_{i-1} - \alpha(A_i - A_{i-1}) + L_i - \beta, 0\}}{\alpha} + A_i \\ &= \frac{\max\{\beta_{i-1} - \beta + L_i, \alpha(A_i - A_{i-1})\}}{\alpha} + A_{i-1} \\ &> \frac{\max\{\beta_{i-1} - \beta, 0\}}{\alpha} + A_{i-1} = R_{i-1}. \end{aligned} \tag{23}$$

This means that packets are also ordered by their time stamps by this encoding, i.e., using this encoding, for two packets P_k and P_m $(k > m) \Rightarrow A_k > A_m \Rightarrow R_k > R_m$. Thus, in order to prove that (β, α) is the virtual traffic parameter of this flow, we only need to prove that, for any two packets P_k and P_m, $k \geq m > 0$, that $\beta + \alpha[R_k - \max\{R_{m-1}, A_m\}] \geq \sum_{i=m}^{k} L_i$ (since $\min\{A_m, A_{m+1}, \ldots A_k\} = A_m$) as follows:

$$\begin{aligned} \sum_{i=m+1}^{k} L_i &= \sum_{i=m+1}^{k} \{\beta_i - \max\{0, \beta_{i-1} - \alpha(A_i - A_{i-1})\}\} \\ &= \sum_{i=m+1}^{k} \min\{\beta_i, \beta_i - \beta_{i-1} + \alpha(A_i - A_{i-1})\} \\ &= \min\left\{ \sum_{i=m+1}^{k} \beta_i, \right. \\ &\qquad \left. \sum_{i=m+1}^{k} \beta_i - \beta_{i-1} + \alpha(A_i - A_{i-1}) \right\} \\ &= \min\left\{ \sum_{i=m+1}^{k} \beta_i, \beta_k - \beta_m + \alpha(A_k - A_m) \right\} \\ &\leq \beta_k - \beta_m + \alpha(A_k - A_m). \end{aligned} \tag{24}$$

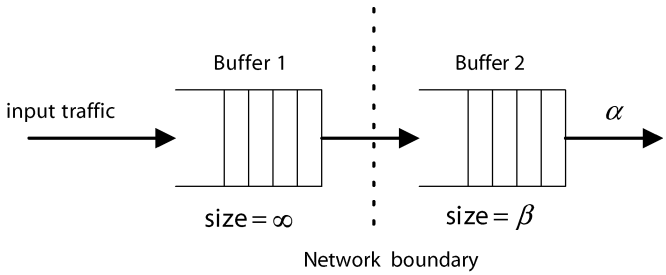

Fig. 4. Model of the proposed time-stamp encoding scheme. In reality, both buffers 1 and 2 do not exist; they are used here for illustrative purposes.

Furthermore

$$\begin{aligned} R_i &= \frac{\max\{\beta_i - \beta, 0\}}{\alpha} + A_i \Rightarrow \alpha R_i + \beta \\ &\geq \alpha A_i + \beta_i \end{aligned} \tag{25}$$

$$\begin{aligned} \beta_i &= \max\{0, \beta_{i-1} - \alpha(A_i - A_{i-1})\} + L_i \Rightarrow \beta_i \\ &\geq L_i \end{aligned} \tag{26}$$

$$\begin{aligned} R_i &= \frac{\max\{\beta_i - \beta, 0\}}{\alpha} + A_i \Rightarrow \alpha R_i \\ &\leq \alpha A_i + \beta_i. \end{aligned} \tag{27}$$

Therefore

$$\begin{aligned} &\alpha[R_k - \max\{R_{m-1}, A_m\}] + \beta \\ &\quad = \min\{\alpha R_k - \alpha R_{m-1} + \beta, \alpha R_k - \alpha A_m + \beta\} \\ &\quad \geq \min\{\alpha A_k + \beta_k - \alpha R_{m-1}, \alpha(A_k - A_m) + \beta_k\} \\ &\quad \geq \min\left\{\alpha A_k + \beta_{m-1} - \alpha A_{m-1} - \beta_{m-1}, \right. \\ &\qquad \left. \sum_{i=m+1}^{k} L_i + \beta_m \right\} \\ &\quad \geq \sum_{i=m}^{k} L_i. \end{aligned} \tag{28}$$

Thus, the virtual traffic parameter of this flow is (β, α). ■

The mechanism of the proposed time-stamp encoding scheme is illustrated in Fig. 4. We propose to assign the time stamp of each packet as the earliest time that it can reach buffer 2 while buffer 2 is not overflowed. Conceptually, Fig. 4 illustrates the (σ, ρ) regulator proposed in [24], and, without buffer 1, it is just the so-called leaky bucket. In the (σ, ρ) regulator, a counter is maintained for each flow entering the network, which is decremented at a prespecified rate (ρ) as long as it is positive. When a packet arrives, the value of the counter is compared to the prespecified threshold (σ). If it is less than the threshold, the packet is admitted into the network (buffer 2). Otherwise, it is stored in buffer 1. However, buffers 1 and 2 do not really exist in our time-stamp encoding scheme; they are just used to illustrate our proposed time-stamp encoding. When packets arrive at the network boundary, they directly go into the core network with the corresponding time stamps inserted in the headers.

Next, under the assumptions that our time-stamp encoding scheme is deployed and a leaky bucket is used for traffic shaping at network boundaries, we will compute the worst case

end-to-end delay bound of a flow in a core-stateless network. Assume packet P traverses nodes $1, 2, \ldots, n$; its arrival time, delivery time, time stamp (before transmission), and the worst case virtual delay at node i are a_i, t_i, r_i, and D_i, respectively. Thus, the delay D of this packet through nodes $1, 2, \ldots, n$ of the network can be expressed as

$$D = t_n - a_1 \le r_n + D_n - a_1 = \sum_{j=1}^{n-1}(r_{j+1} - r_j) + D_n + r_1 - a_1. \tag{29}$$

Note that $d_i = r_{i+1} - r_i$ is just the time-stamp increment of this packet updated by node i. Thus

$$D \le \sum_{i=1}^{n-1} d_i + D_n + r_1 - a_1. \tag{30}$$

Therefore, the delay of a packet in a core-stateless network is bounded by the sum of the increments updated by the nodes it traverses (except for the last node), its worst case virtual delay at the outgoing (last) node, and the amount of time its time stamp lags behind its arrival time at the first node. Equation (30) also reveals three important properties of the worst case delay bound of a packet to traverse a core-stateless network.

1) The smaller the sum of the time-stamp increments of a packet updated by the core-nodes it traverses, the smaller the worst case delay bound of this packet to traverse the core-stateless network is.
2) The smaller the amount of time its time stamp lags behind its arrival time at the first core node it traverses, the smaller the worst case delay bound of this packet is.
3) The smaller the worst case virtual delay of this packet at the last core node it traverses, the smaller the worst case delay bound of this packet is.

Hence, minimizing the amount of time that the time stamp of a packet lags behind its arrival time at the first core node is preferred. In this paper, under the assumption that the traffic is shaped by the leaky bucket at the edge of the network, we show that the time-stamp encoding scheme proposed in our proposition is optimal in terms of minimizing the amount of time a packet's time stamp lags behind its arrival time at the first core node.

Theorem 4: Assume that a flow is shaped by a leaky bucket with parameter (σ, ρ) before it enters the network and its arriving traffic in any interval $(t_1, t_2]$ is bounded by $\min\{c(t_2 - t_1), \rho(t_2 - t_1) + \sigma\}$, where c is the bandwidth of the input link of the edge node. Its virtual traffic parameter is (β, α) $(\alpha \ge \rho)$ at the first node. Thus, for any packet P_k, we have

$$R_k - A_k \le \max\left\{\frac{L_{\max} - \beta}{\alpha}, \frac{(c-\alpha)\sigma - \beta(c-\rho)}{\alpha(c-\rho)}, 0\right\}. \tag{31}$$

Proof: Let m, $0 \le m \le k$, be the largest integer such that $\beta_m - L_m = 0$. Consider the case that $m = k$, and then $\beta_k = L_l$. From (21) and (22), we have

$$R_k - A_k = \max\left\{0, \frac{L_k - \beta}{\alpha}\right\} \le \max\left\{0, \frac{L_{\max} - \beta}{\alpha}\right\}. \tag{32}$$

Consider the other case that $m \ne k$, which is

$$\begin{aligned} \beta_k &= \beta_{k-1} - \alpha(A_k - A_{k-1}) + L_l \\ &\Rightarrow \sum_{i=m+1}^{k} L_i = \beta_k - \beta_m + \alpha(A_k - A_m) \\ &\Rightarrow \sum_{i=m+1}^{k} L_i + \beta_m \\ &= \beta_k + \alpha(A_k - A_m). \end{aligned} \tag{33}$$

Note that

$$\sum_{i=m}^{k} L_i \le \min\left\{c(A_k - A_m), \rho(A_k - A_m) + \sigma\right\}. \tag{34}$$

Since $\beta_m = L_m$, from (33), we have

$$\begin{aligned} &\beta_k + \alpha(A_k - A_m) \\ &\quad = \sum_{i=m}^{k} L_i \le \min\left\{c(A_k - A_m), \rho(A_k - A_m) + \sigma\right\} \\ &\quad \Rightarrow \beta_k \le \frac{(c-\alpha)\sigma}{c-\rho}. \end{aligned} \tag{35}$$

Thus

$$\begin{aligned} R_k - A_k &= \frac{\max\{\beta_k - \beta, 0\}}{\alpha} \\ &\le \max\left\{0, \frac{(c-\alpha)\sigma - \beta(c-\rho)}{\alpha(c-\rho)}\right\}. \end{aligned} \tag{36}$$

Thus, for any packet P_k, $R_k - A_k \le \max\{(L_{\max} - \beta/\alpha), ((c - \alpha)\sigma - \beta(c-\rho)/\alpha(c-\rho)), 0\}$. ■

By Theorems 3 and 4, the worst case delay of a packet in the work-conserving core-stateless network is bounded by

$$\max\left\{\frac{L_{\max} - \beta}{\alpha}, \frac{(c-\alpha)\sigma - \beta(c-\rho)}{\alpha(c-\rho)}, 0\right\} + \sum_{i=1}^{n-1} d_i + D_n \tag{37}$$

which is derived under the assumptions that the time-stamp encoding scheme proposed in our proposition is deployed and the leaky bucket for admission control at the edge of the network is adopted. In fact, if the arriving traffic of a flow in $(0, t]$ is ***equal*** to $\min\{ct, \sigma + \rho t\}$, it can be proved that

$$\begin{aligned} &\max_{k=1,2,\ldots}\{R_k - A_k\} \\ &\quad \ge \max\left\{\frac{L_{\max} - \beta}{\alpha}, \frac{(c-\alpha)\sigma - \beta(c-\rho)}{\alpha(c-\rho)}, 0\right\} \end{aligned} \tag{38}$$

thus implying that our proposed time-stamp encoding scheme is optimal in terms of minimizing the worst case delay bound of a flow by minimizing the maximum time that the time stamp of a packet may lag behind its arrival time.

It should be noted that, by deploying our proposed (β, α) traffic model, the design and performance analysis of core-stateless algorithms such as the one proposed in [13] and DETF [11] can be easily achieved. For example, in order to make the virtual traffic parameter of each flow at the input port of any node $(0, \alpha)$, where α is the requested rate of this flow, by Lemma 1, all of its packets' time stamp of this flow should be updated by an increment $d = D + \delta + (L_{\max}/\alpha)$ at a node, where D is its worst case virtual delay to traverse this node and δ is the propagation delay from its previous node to this node. Moreover, by Theorem 3, it can be derived that $D = L_{\max}/c$, because the virtual traffic parameters of all flows are in the form of $(0, \alpha)$. Therefore, the time-stamp increment of a flow at a node is $d = (L_{\max}/c) + \delta + (L_{\max}/\alpha)$, which is the same as that in [13].

V. Conclusion

In this paper, a new framework for a bounded-delay work-conserving core-stateless network has been presented, covering three important issues in the core-stateless network: time-stamp encoding, traffic distortion, and worst case delay analysis. All of these are achieved based on a new and efficient traffic model, which is the (β, α) traffic model, for characterizing traffic in a core-stateless network. Based on this model, a time-stamp encoding scheme has been proposed and proven to be effective in minimizing the end-to-end worst case delay bound.

References

[1] R. Braden, D. Clark, and S. Shenker, Integrated Services in the Internet Architecture: An Overview RFC1633, Jun. 1994.

[2] S. Blake, D. Black, M. Calson, E. Davies, Z. Wang, and W. Weiss, An Architecture for Differentiated Services RFC2475, Dec. 1998.

[3] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control—the single node case," in *Proc. INFOCOM'92*, May 1992, pp. 915–924.

[4] L. Zhang, "Virtual clock: A new traffic control algorithm for packet switching networks," in *Proc. ACM SIGCOMM'90*, Sep. 1990, pp. 19–29.

[5] J. C. R. Bennett and H. Zhang, "WF ^{2}Q: Worst-case fair weighted fair queueing," in *Proc. IEEE INFOCOM'96*, Mar. 1996, pp. 120–128.

[6] S. Golestani, "A self-clocked fair queueing scheme for broadband applications," in *Proc. IEEE INFOCOM'94*, Jun. 1994, pp. 636–646.

[7] D. Verma, H. Zhang, and D. Ferrari, "Guaranteeing delay jitter bounders in packet switching networks," in *Proc. Tricomm'91*, Apr. 1991, pp. 35–46.

[8] S. Golestani, "Congestion-free transmission of real-time traffic in packet networks," in *Proc. IEEE INFOCOM'90*, Jun. 1990, pp. 527–542.

[9] A. Charny and J.-Y. LeBoudec, "Delay bounds in a network with aggregate scheduling," in *Proc. QoFIS*, Oct. 2000, pp. 1–13.

[10] Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine, A Framework for Integrated Services Operation Over Diffserv Networks RFC2998, Nov. 2000.

[11] Z. Zhang, Z. Duan, and Y. Hou, "Fundamental trade-offs in aggregate packet scheduling," in *Proc. IEEE ICNP'01*, 2001, pp. 129–137.

[12] I. Stoica and H. Zhang, "Providing guaranteed services without per-flow management," in *Proc. ACM SIGCOMM'99*, Sep. 1999, pp. 81–94.

[13] J. Kaur and H. M. Vin, "Core-stateless guaranteed rate scheduling algorithms," in *Proc. IEEE INFOCOM'01*, Apr. 2001, pp. 1484–1492.

[14] ——, "Core-stateless guaranteed throughput networks," in *Proc. IEEE INFOCOM*, 2003, vol. 3, pp. 2155–2165.

[15] S. Bhatnagar and B. Nath, "Distributed admission control to support guaranteed services in core-stateless networks," in *Proc. IEEE INFOCOM*, 2003, vol. 3, pp. 1659–1669.

[16] Z. Zhang, Z. Duan, and Y. Hou, "Virtual time reference system: a unifying scheduling framework for scalable support of guaranteed services," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 12, pp. 2684–2695, Dec. 2000.

[17] Z. Duan, Z. Zhang, Y. Hou, and L. Gao, "A core stateless bandwidth broker architecture for scalable support of guaranteed services," *IEEE Trans. Parallel Distrib. Syst.*, vol. 15, no. 2, pp. 167–182, Feb. 2004.

[18] J. Li, Y. Lin, and C. Yang, "Core-stateless fair rate estimation fair queuing," in *Proc. MILCOM*, 2002, vol. 2, pp. 1165–1170.

[19] I. Stoica, S. Shenker, and H. Zhang, "Core-stateless fair queueing: A scalable architecture to approximate fair bandwidth allocations in high-speed networks," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 33–46, Feb. 2003.

[20] I. Stoica, H. Zhang, and S. Shenker, "Self-verifying CSFQ," in *Proc. IEEE INFOCOM*, 2002, vol. 1, pp. 2–30.

[21] C. Albuquerque, B. J. Vickers, and T. Suda, "Network border patrol: Preventing congestion collapse and promoting fairness in the internet," *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 173–186, Feb. 2004.

[22] Z. Cao, Z. Wang, and E. Zegura, "Rainbow fair queueing: Fair bandwidth sharing without per-flow state," in *Proc. IEEE INFOCOM*, 2000, vol. 2, pp. 922–931.

[23] Z. Clerget and W. Dabbous, "TUF: Tag-based unified fairness," in *Proc. IEEE INFOCOM*, 2001, vol. 1, pp. 498–507.

[24] R. Cruz, "A calculus for network delay—Part I: Network elements in isolation," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 121–141, Jan. 1991.

[25] L. Zhang, "Virtual clock: A new traffic control algorithm for packet switching networks," in *Proc. ACM SIGCOMM'90*, 1990, pp. 19–29.

[26] C. Li and E. Knightly, "Coordinated multihop scheduling: A framework for end-to-end services," *IEEE/ACM Trans. Netw.*, vol. 10, no. 6, pp. 776–789, Dec. 2002.

Gang Cheng received the B.S. degree in information engineering and M.E. degree in information and signal processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1997 and 2000, respectively, and the Ph.D. degree from the New Jersey Institute of Technology (NJIT), Newark, in 2005.

In 2000, he joined Lucent Technologies. Between January 2001 and May 2005, he was with NJIT. His research interests include Internet routing protocols and service architectures, information theory-based network optimization and protocol design, and modeling and performance evaluation of computer and communication systems. Since January 2005, he has been with VPIsystems Corporation, Holmdel, NJ, where he is focusing on the design and development of the network planning optimization algorithm.

Dr. Cheng was the recipient of the Hashimoto Prize, which is awarded annually to the "best" NJIT doctoral graduate in the Department of Electrical and Computer Engineering.

Li Zhu received the B.S. degree in physics from Nanjing University, Nanjing, China, in 1994, the M.S. degree in physics from Nanjing University, Nanjing, China, in 1997, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 2001, and the Ph.D. degree in electrical engineering from the New Jersey Institute of Technology, Newark, in 2006.

He is currently a Senior Network Engineer with Voicenet. His current research interests include quality of service in the Internet, overlay networks, and ad hoc networks.

Nirwan Ansari received the B.S.E.E. degree (*summa cum laude*) from the New Jersey Institute of Technology (NJIT), Newark, in 1982, the M.S.E.E. degree from the University of Michigan, Ann Arbor, in 1983, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 1988.

Since 1997, he has been a Full Professor with the Department of Electrical and Computer Engineering, NJIT. He coauthored *Computational Intelligence for Optimization* (Kluwer, 1997, translated into Chinese in 2000) and coedited *Neural Networks in Telecommunications* (Kluwer, 1994). He is a Technical Editor for the *Computer Communications*, the *ETRI Journal*, as well as the *Journal of Computing and Information Technology*. His current research focuses on various aspects of broadband networks and multimedia communications. He has contributed over 90 refereed journal articles, plus numerous conference papers and book chapters, and he has been frequently invited to deliver keynote addresses, tutorials, and talks.

Dr. Ansari organized (as General Chair) the First IEEE International Conference on Information Technology: Research and Education (ITRE2003), was instrumental, while serving as its Chapter Chair, in rejuvenating the North Jersey Chapter of the IEEE Communications Society, which received the 1996 Chapter of the Year Award and a 2003 Chapter Achievement Award, served as Chair of the IEEE North Jersey Section and in the IEEE Region 1 Board of Governors during 2001–2002, and currently serves in various IEEE committees including as TPC Chair/Vice-chair of several conferences. He was the 1998 recipient of the NJIT Excellence Teaching Award in Graduate Instruction and a 1999 IEEE Region 1 Award. He has been selected as an IEEE Communications Society Distinguished Lecturer (2005–2007), and he is a Senior Technical Editor for the *IEEE Communications Magazine*.