Queue Dynamics of RED Gateways Under Large Number of TCP Flows

Peerapol Tinnakornsrisuphap, Armand M. Makowski
peerapol@eng.umd.edu, armand@isr.umd.edu

Department of ECE and and Institute for Systems Research
University of Maryland, College Park, MD 20742

Abstract—**We consider a stochastic model of a RED gateway under competing TCP-like sources sharing the capacity. As the number of competing flows becomes large, the queue behavior of RED can be described by a two-dimensional recursion. We confirm the result by simulations and discuss their implications for the network dimensioning problem.**

I. Introduction

One of the key mechanisms for the operation of the best-effort service Internet is the congestion-control mechanism in TCP. While there are several variations on the basic TCP congestion-control mechanism, they all have in common the *additive increase/multiplicative decrease* (AIMD) algorithm. The AIMD algorithm enables TCP congestion-control to be robust under diverse conditions. However, it is well known that with tail-drop gateways this congestion-control also leads to undesirable behavior, i.e., global synchronization. When several TCP flows compete for bandwidth in a tail-drop gateway, it has been observed experimentally that packets from many flows are usually discarded simultaneously [1], resulting in a poor utilization of the network. Active queue management algorithms such as Random Early Detection (RED) [2] were introduced to help alleviate this problem by randomly dropping packets depending on the queue size, thereby avoiding heavy congestion and prevent global synchronization.

While there are many efforts to model TCP throughput under a tail-drop assumption [3] [4] [5] [6], only a few studies have focused on modeling the interaction of RED gateways with TCP congestion-control. In [7], an analytical framework for multiple TCP flows sharing a RED gateway is developed under several potentially unrealistic assumptions. In [8], a simple analysis has been done with TCP connections operating as Poisson processes under "slow" and "fast" rates. Fixed point solutions to average TCP window sizes and queue occupancy are discussed in [9]. However, a model that can provide a good analytic understanding of TCP and RED is yet to be found. The difficulties arise from the complex behavior of TCP congestion-control, and are further compounded by the random drop mechanism and queue averaging. Detailed modeling of these characteristics results in a number of states which explodes when the number of TCP flows increase, making the analysis untractable.

In this paper, we present a stochastic model that captures the essential features of TCP, i.e., the gradual adaptive increase and the sudden decrease of transmission rate, combined with a random drop algorithm similar to RED. We analyze this ersatz model as the number of competing TCP flows becomes large, and show that the stochastic model simplifies in the limit to a two-dimensional recursion. This result suggests that with a large number of flows, it is easy for network operators to estimate the aggregate behavior of TCP flows and to dimension network resources accordingly.

The remainder of the paper is organized as follows. Section II describes the stochastic model. Section III present the main asymptotic results for the large number of TCP flows whereby the stochastic model simplifies into a simplified limiting recursion. Simulation results supporting this behavior are shown in Section IV. The conclusions of the paper are given in Section V.

II. The Model

TCP congestion-control utlizes the AIMD algorithm to provide TCP flows a fair bandwidth share [10] by using feedback obtained through acknowledgement packets (ACKs). If an ACK packet is received (i.e., a packet is successfully transmitted and acknowledged), TCP increases its transmission rate by a small, conservative amount. Otherwise, TCP interprets a lack of acknowledgement as a sign of congestion and reduces its transmission rate by half. We present an algorithm similar in spirit to the AIMD congestion-control in TCP and apply it to the model described earlier.

A. Definitions and notation

Time is assumed discrete and slotted in contiguous timeslots of equal duration normalized to a packet transmission time. We consider N traffic sources, all transmitting through a bottleneck RED gateway. The capacity of this bottleneck scales with the number N of flows, i.e., it has capacity NC packets/slot for some positive constant C. We model the RED buffer as an infinite queue, so that packet losses are due only to the random drop algorithm [1]

B. Dynamics

Fix $N = 1, 2, \ldots$ and $t = 0, 1, \ldots$. For any quantity X, we write $X^{(N)}(t)$ to indicate the explicit dependence of X on the number N of connections.

Let $Q^{(N)}(t)$ denote the number of packets in the buffer at the beginning of the timeslot $[t, t+1)$. Suppose that each source (or equivalently, TCP connection) generates at most one packet in each timeslot. So let $B_i^{(N)}(t+1)$ be a $\{0,1\}$-valued rv that encodes the number of packets generated by source i at the beginning of the timeslot $[t, t+1)$. The packet from source i, upon arrival at the RED gateway, may be rejected by the random drop algorithm (to be specified shortly). We represent this possibility by the $\{0,1\}$-valued rv $R_i^{(N)}(t+1)$ with the interpretation that $R_i^{(N)}(t+1) = 1$ (resp. $R_i^{(N)}(t+1) = 0$) if the packet is rejected (resp. accepted into the RED buffer). Given that N sources are active, the total number of packets which are accepted into the RED buffer at the beginning of timeslot $[t, t+1)$ is given by

$$A^{(N)}(t+1) := \sum_{i=1}^{N} (1 - R_i^{(N)}(t+1)) B_i^{(N)}(t+1).$$

If $Q^{(N)}(t)$ denotes the number of packets in the buffer at the beginning of the timeslot $[t, t+1)$, then $Q^{(N)}(t) + A^{(N)}(t+1)$ packets are available for transmission. Since the outgoing link operates at the rate of NC packets/timeslot, $\left[Q^{(N)}(t) + A^{(N)}(t+1) - NC\right]^+$ packets will not be transmitted during timeslot $[t, t+1]$, and remain in buffer, their transmission being deferred to subsequent timeslots. The number $Q^{(N)}(t+1)$ of packets in the buffer at the beginning of the timeslot $[t+1, t+2)$ is therefore given by [2]

$$Q^{(N)}(t+1) = \left[Q^{(N)}(t) - NC + A^{(N)}(t+1)\right]^+. \quad (1)$$

C. Statistical assumptions

In order to fully specify the model, we need to specify the statistics of the rvs $\{B_i^{(N)}(t+1), R_i^{(N)}(t+1),\ i = 1, \ldots, N;\ t = 0, 1, \ldots\}$ for each $N = 1, 2, \ldots$.

To do so we introduce the collection of i.i.d. $[0,1]$-*uniform* rvs $\{V_i(t+1), U_i(t+1),\ i = 1, \ldots;\ t = 0, 1, \ldots\}$. For each $i = 1, \ldots, N$, we take

$$B_i^{(N)}(t+1) = \mathbf{1}[U_i(t+1) \leq \alpha_i^{(N)}(t)] \quad (2)$$

where $\alpha_i^{(N)}(t)$ is an $[0,1]$-valued rv which denotes the (conditional) *transmission rate* (to be specified shortly) of traffic source i at the beginning of the timeslot $[t, t+1)$. We also set

$$R_i^{(N)}(t+1) = \mathbf{1}[V_i^{(N)}(t+1) \leq f^{(N)}(Q^{(N)}(t))] \quad (3)$$

where $f^{(N)} : \mathbb{R}_+ \to [0,1]$ denotes the *drop probability* function of the RED gateway.

To select the transmission rates we argue as follows: Suppose that source i generates no packet during timeslot $[t, t+1)$ (i.e. $B_i^{(N)}(t+1) = 0$), then the transmission rate of source i in the next timeslot remains unchanged. If on the other hand, a packet is produced by source i at the beginning of timeslot $[t, t+1)$, then either the packet is successfully transmitted ($R_i^{(N)}(t+1) = 0$), or it is dropped ($R_i^{(N)}(t+1) = 1$). In the former case, the transmission rate of source i in the next timeslot is *increased* to $\alpha_i^{(N)}(t)^{1-\varepsilon}$ (with $0 < \varepsilon < 1$). In the latter case, $\alpha_i^{(N)}(t+1)$ is *decreased* by a factor γ (or $\alpha_i^{(N)}(t+1) = \gamma \alpha_i^{(N)}(t)$), where $0 < \gamma < 1$. These two situations attempt to emulate (under the constraint that transmission rates are bounded to the unit interval) the additive increase and multiplicative decrease, respectively, of the TCP congestion-control. They can be summarized into the single equation

$$\begin{aligned} & \alpha_i^{(N)}(t+1) \\ = \ & \alpha_i^{(N)}(t)^{1-\varepsilon}(1 - R_i^{(N)}(t+1)) B_i^{(N)}(t+1) \\ & + \gamma \alpha_i^{(N)}(t) R_i^{(N)}(t+1) B_i^{(N)}(t+1) \\ & + \alpha_i^{(N)}(t)(1 - B_i^{(N)}(t+1)). \end{aligned} \quad (4)$$

For each $t = 0, 1, \ldots$, let $\mathcal{F}_t$ denote the σ-field generated by the rvs $\{Q^{(N)}(0), \alpha_i^{(N)}(0), V_i(s), U_i(s),\ i = 1, \ldots;\ s = 1, \ldots, t\}$. Note the rvs $Q^{(N)}(t)$ and $\alpha_i^{(N)}(t)$ ($i = 1, \ldots, N$) are all $\mathcal{F}_t$-measurable, so that

$$\mathbf{E}\left[B_i^{(N)}(t+1) | \mathcal{F}_t\right] = \alpha_i^{(N)}(t)$$

[1] We can account for the finiteness of the buffer by modifying the queue dynamics as is done in the footonote to recursion (1).

[2] The finiteness of the buffer in (1) can be replaced by $Q^{(N)}(t+1) = \min([Q^{(N)}(t) - NC + \sum_{i=1}^{N}(1 - R_i^{(N)}(t+1)) B_i^{(N)}(t+1)]^+, NB)$ where B denotes the buffer size per connection.

for all $i = 1, \ldots, N$, and

$$\mathbf{E}\left[R_i^{(N)}(t+1)|\mathcal{F}_t\right] = f^{(N)}(Q^{(N)}(t)).$$

III. Main Results

In this section, we analyze the model presented in Section II as the number of traffic flows increase to infinite and discuss the implications of the result.

A. The asymptotics

The discussion is carried out under the following assumptions: There exist a continuous function $f : \mathbb{R}_+ \to [0,1]$ and a constant α in $(0,1)$ such that for each $N = 0, 1, \ldots,$

(A1) $f^{(N)}(x) = f(x/N)$ $(x \geq 0)$;

(A2) $Q^{(N)}(0) = 0$ and $\alpha_i^{(N)}(0) = \alpha$ $(i = 1, \ldots, N)$.

We begin with an easy consequence of these assumptions.

Lemma 1: Assume (A1)-(A2) to hold. Then, for each $t = 0, 1, \ldots,$, the rvs $\{\alpha_1^{(N)}(t), \ldots, \alpha_N^{(N)}(t)\}$ are exchangeable for all $N = 1, 2, \ldots$.

The next proposition presents the asymptotics for the normalized buffer content as the number of TCP sources becomes large.

Theorem 1: Assume (A1)-(A2) to hold. Then, for each $t = 0, 1, \ldots$, there exist a non-random constant $q(t)$ and a rv $\alpha(t)$ such that

$$\frac{Q^{(N)}(t)}{N} \xrightarrow{P}{}_N q(t) \quad \text{and} \quad \alpha_1^{(N)}(t) \xrightarrow{P}{}_N \alpha(t) \tag{5}$$

and for every $p > 0$,

$$\tfrac{1}{N}\textstyle\sum_{i=1}^N (\alpha_i^{(N)}(t))^p \xrightarrow{P}{}_N \mathbf{E}\left[\alpha(t)^p\right]. \tag{6}$$

Moreover,

$$q(t+1) = \left[q(t) - C + (1 - f(q(t)))\mathbf{E}\left[\alpha(t)\right]\right]^+ \tag{7}$$

and

$$\begin{aligned} & \alpha(t+1) \\ = \; & \alpha(t)^{1-\varepsilon}\mathbf{1}[V(t+1) > f(q(t))]\mathbf{1}[U(t+1) \leq \alpha(t)] \\ & + \gamma\alpha(t)\mathbf{1}[V(t+1) \leq f(q(t))]\mathbf{1}[U(t+1) \leq \alpha(t)] \\ & + \alpha(t)\mathbf{1}[U(t+1) > \alpha(t)] \end{aligned} \tag{8}$$

for i.i.d. $[0,1]$-uniform rvs $\{V(t+1), U(t+1),\ t = 0, 1, \ldots\}$.

With $a_p(t) = \mathbf{E}\left[\alpha(t)^p\right]$ $(p > 0)$, we readily get

$$\begin{aligned} a_p(t+1) &= a_p(t) + (1 - f(q(t)))\, a_{(1-\varepsilon)p}(t) \\ &\quad + (\gamma^p f(q(t)) - 1)\, a_{p+1}(t). \end{aligned} \tag{9}$$

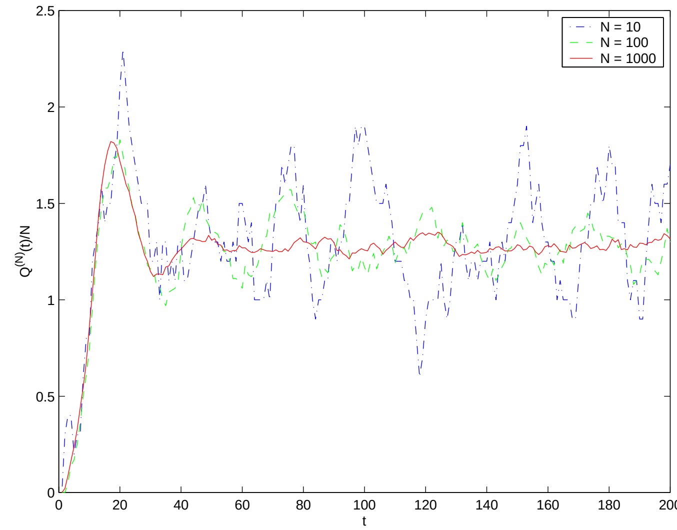

Fig. 1. The normalized queue length of Simulation 1.

B. Discussion

Theorem 1 suggests that a bottleneck queue with random-drop algorithm, under large number of TCP-like sources, can be characterized by a two-dimensional recursion giving the evolution of the normalized queue length $q(t)$ and the limiting transmission rate $\alpha(t)$. This result is not a straightforward consequence of the Law of Large Number due to the fact that (i) the transmission rates of traffic sources are *correlated* and (ii) they vary with N. However, as the number of sources increases, the dependency between any pair of sources becomes weaker so that the aggregate behavior eventually becomes deterministic. Thus, as the aggregate queue behavior scales linearly with the number of sources, the network provider could effectively dimension network resources by tacking the normalized queue behavior.

The sequence $\{(q(t), \alpha(t)),\ t = 0, 1, \ldots\}$ in Theorem 1 defines an $\mathbb{R}_+ \times [0,1]$-valued Markov process, and we expect that it admits a steady-state regime. This will be discussed elsewhere.

IV. Simulation

We simulate the system described earlier for $N = 10, 100, 1000$ with $\varepsilon = 0.1$, $\gamma = 0.5$ and $C = 0.5$; the initial conditions are $Q^{(N)}(0) = 0$ and $\alpha_i^{(N)} = 0.5$ for all $i = 1, .., N$. The drop probability is calculated through the piecewise linear function $f : \mathbb{R}_+ \to [0,1]$ given by

$$f(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{4} & 1 \leq x < 5 \\ 1 & 5 \leq x. \end{cases} \tag{10}$$

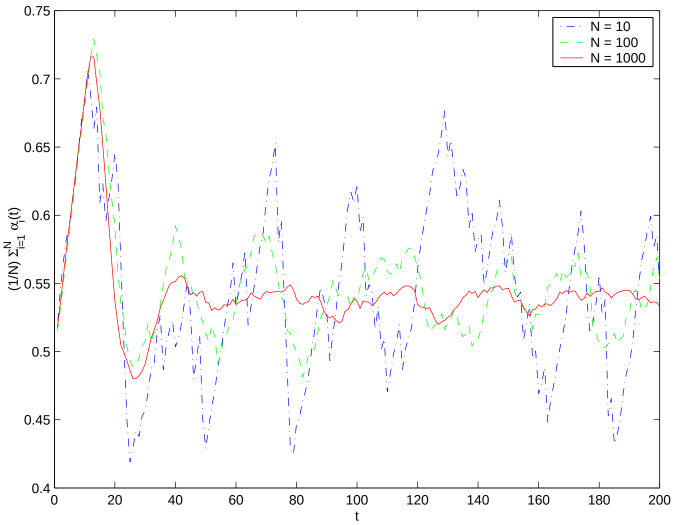

Fig. 2. The average transmission rate per user of Simulation 1.

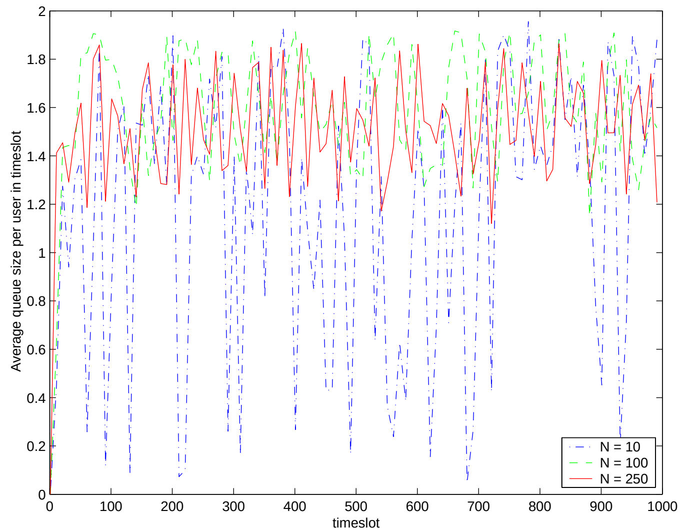

Fig. 3. The average queue size per user of ns simulation.

The simulation results are shown in Figure 1 and 2. It is clear that the fluctuation of $Q^{(N)}(t)/N$ decreases as the number of sources increases, and the same is true for the average transmission rate. With a hundred or more flows, our analytical result seems to hold reasonably well. Moreover, this simulation result also suggests the existence of the steady-state, which happens quickly after only around a hundred iterations.

We also simulate a similar system in *ns* by generating N TCP Reno connections, each of which having 100 ms round-trip delay, all competing to transmit through a bottleneck gateway with link capacity N Mbps. The TCP packet size is set to be 1500 bytes. The buffer management scheme in the gateway is RED with the following parameters: $thresh = 2N$, $maxthresh = 5N$, $p_{drop} = 1$ when queue size is greater than $5N$ and $w_q = 0.1$. The timeslot that we observed the average queue length in the bottleneck gateway is 1 second. Figure 3 shows the simulation result; a trend similar to that of Figure 1 is observed in that as the number of TCP connection increases, the fluctuation in the average queue size decreases. As time passes, the range of fluctuation settles to a certain limited range. The preliminary findings comfort our belief that our ersatz model captures some of the essential features of TCP and RED and then illustrate the important behavior of the interaction between these two mechanisms.

V. Conclusions

We have developed a stochastic model for a RED gateway under competing TCP flows. We have shown that, as the number of flows grow large, the aggregate behavior of the queue can be described by a two-dimensional recursion. We have also discussed the implications of our results to the dimensioning of the network.

Although we have yet to prove the existence of a steady-state regime for the limiting recursion, the limited simulation results (i) are compatible with the existence of such a steady-state and (ii) suggest that the rate of convergence is fast in either the number of sources (to achieve limiting behavior) and the time (to reach the steady-state).

Future work on this class of models includes (i) a proof of the existence of a steady state for the limiting dynamics and its evaluation; (ii) a derivation of a CLT complement to the basic convergence result; and (iii) the development of more accurate models (e.g., More than one packet generated per timeslot; asynchronous updating of the transmission rates; non-homogeneous population of TCP flows and continuous-time versions.) It is also interesting to see how the shape of the drop function affect the rate of convergence. Furthermore, we should be able to investigate the fairness of the competing TCP flows in this model. And finally, we believe that this model is simple enough to be extended to the scenario where the sources are stochastic, say on-off sources.

References

[1] L. Zhang and D. Clark, "Oscillating behavior of network traffic: A case study simulation," *Internetworking: Research and Experience*, vol. 1, no. 2, pp. 101–112, 1990.

[2] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397–413, 1993.

[3] M. Mathis, J. Semke, J. Madhavi, and T. Ott, "The macroscopic behavior of the tcp congestion avoidance algorithm," *Computer Communication Review*, vol. 27, no. 3, pp. 67–82, 1997.

[4] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling tcp reno performance: A simple model and its empirical validation," *IEEE/ACM Transactions on Networking*, vol. 8, no. 2, pp. 133–145, 2000.
[5] Th. Bonald, "Comparison of TCP Reno and TCP Vegas via fluid approximation," Tech. Rep. 3563, INRIA, Nov. 1998.
[6] T. V. Lakshman and U. Madhow, "The performance of TCP/IP for networks with high bandwidth-delay products and random loss," *IEEE/ACM Transactions on Networking*, June 1997.
[7] A. Abouzeid and S. Roy, "Analytic understanding of red gateways with multiple competing tcp flows," in *Proceedings of Globecom 2000*, 2000.
[8] H. M. Alazemi, A. Mokhtar, and M. Azizoglu, "Stochastic modeling of random early detection gateways in tcp networks," in *Proceedings of Globecom 2000*, 2000.
[9] A. Misra, *Dynamics of TCP Congestion Avoidance with Random Drop and Random Marking Queues*, Ph.D. thesis, University of Maryland, 2000.
[10] V. Jacobson, "Congestion avoidance and control," in *Proceedings of SIGCOMM'88 Symposium*, Aug. 1988, pp. 314–332.
[11] P. Tinnakornsrisuphap and A.M. Makowski, *Queue Dynamics of RED Gateways Under Large Number of TCP Flows: Limit Behavior and Steady-State Regime*, Technical Report, Institute for Systems Research, University of Maryland, College Park, MD 20742.
[12] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford Science Publications, 1992.

VI. Outline of a Proof of Theorem 1

A complete proof of Theorem 1 is available in [11]. Before we outline the key elements of this proof, we introduce the following notation: For each $t = 0, 1, \ldots$, the statements **[A:t]**, **[B:t]** and **[C:t]** refer to the convergences

$$[\mathbf{A}:\mathbf{t}]: \qquad \frac{Q^{(N)}(t)}{N} \xrightarrow{P}{}_N q(t) \text{ with } q(t) \text{ non-random;}$$

$$[\mathbf{B}:\mathbf{t}]: \qquad \alpha_1^{(N)}(t) \xrightarrow{P}{}_N \alpha(t);$$

$$[\mathbf{C}:\mathbf{t}]: \qquad \tfrac{1}{N}\textstyle\sum_{i=1}^N [\alpha_i^{(N)}(t)]^p \xrightarrow{P}{}_N a_p(t) \text{ non-random.}$$

The equality $a_p(t) = \mathbf{E}\left[\alpha(t)^p\right]$ in **[C:t]** readily follows under **[B:t]** since the rvs $\{\alpha_1^{(N)}(t), \ldots, \alpha_N^{(N)}(t)\}$ are exchangeable and bounded. Since the statements **[A:t]**, **[B:t]** and **[C:t]** do hold for $t = 0$, Theorem 1 will be proved by induction if the following induction step can be established.

Proposition 1: Assume (A1)-(A2) to hold as in Theorem 1. If for some $t = 0, 1, \ldots$, **[A:t]**, **[B:t]** *and* **[C:t]** *hold, then so do* **[A:t+1]**, **[B:t+1]** *and* **[C:t+1]**.

This proposition can be proved with the help of a series of lemmas. The first two lemmas are elementary, and their proofs are therefore omitted.

Lemma 2: Let U denote a [0,1] uniform rv which is independent of the $[0,1]$-valued rvs $\{X, X_n,\ n = 1, 2, \ldots\}$. If $X_n \xrightarrow{P}{}_n X$, then $\mathbf{1}\left[U \leq X_n\right] \xrightarrow{P}{}_n \mathbf{1}\left[U \leq X\right]$.

Lemma 3: For each $N = 1, 2, \ldots$, assume the rvs $\{\xi_i^{(N)}, i = 1, \ldots, N\}$ to be bounded, say $|\xi_i^{(N)}| \leq 1$ for all $i = 1, \ldots, N$, and the rvs $\{\nu_i^{(N)}, i = 1, \ldots, N\}$ to be exchangeable and bounded, say $|\nu_i^{(N)}| \leq 1$ for all $i = 1, \ldots, N$. If $\nu_1^{(N)} \xrightarrow{P}{}_N 0$, then $\frac{1}{N}\sum_{i=1}^N \xi_i^{(N)}\nu_i^{(N)} \xrightarrow{P}{}_N 0$.

The next lemma takes the first step towards proving Proposition 1.

Lemma 4: Under (A1), if **[A:t]**, **[B:t]** *and* **[C:t]** *hold for some $t = 0, 1, \ldots$, then* **[B:t+1]** *holds with $\alpha(t+1)$ related to $\alpha(t)$ by (8).*

Proof: The continuity of f and the assumed convergence **[A:t]** readily lead [12, p. 326] to

$$f^{(N)}(Q^{(N)}(t)) = f(Q^{(N)}(t)/N) \xrightarrow{P}{}_N f(q(t)), \qquad (11)$$

and Lemma 2 thus yields

$$R_1^{(N)}(t+1) \xrightarrow{P}{}_N \mathbf{1}[V_1(t+1) \leq f(q(t))]. \qquad (12)$$

Also by Lemma 2, we have

$$B_1^{(N)}(t+1) \xrightarrow{P}{}_N \mathbf{1}[U_1(t+1) \leq \alpha(t)] \qquad (13)$$

where the rv $U_1(t+1)$ is independent of the rv $\alpha(t)$. Finally, under **[B:t]**, we obtain (8) directly from (4). ■

The next two lemmas provide the final steps in the proof of Proposition 1.

Lemma 5: Under (A1), if **[A:t]**, **[B:t]** *and* **[C:t]** *hold for some $t = 0, 1, \ldots$, then*

$$\frac{A^{(N)}(t+1)}{N} \xrightarrow{P}{}_N (1 - f(q(t)))a_1(t) \qquad (14)$$

and **[A:t+1]** *holds.*

Lemma 6: Under (A1)–(A2), if **[A:t]**, **[B:t]** *and* **[C:t]** *hold for some $t = 0, 1, \ldots$, then* **[C:t+1]** *holds.*

The proofs of Lemmas 5 and 6, while involved, are not very difficult and follow the same pattern: The rvs $(1 - R_i^{(N)}(t+1))B_i^{(N)}(t+1)$, $R_i^{(N)}(t+1)B_i^{(N)}(t+1)$ and $1 - B_i^{(N)}(t+1)$ are indicator functions of mutually exclusive events. Hence, $[\alpha_i^{(N)}(t+1)]^q$ equals to the right-hand side of (9) with $\alpha_i^{(N)}(t)$ replaced by $[\alpha_i^{(N)}(t)]^q$. We can expand $A^{(N)}(t)$ and $\frac{1}{N}\sum_{i=1}^N [\alpha_i^{(N)}(t+1)]^q$ by "centering" each term by subtracting and adding back the appropriate conditional mean, say $\mathbf{1}[V_i(t+1) \leq f(q(t))]$ for $R_i^{(N)}(t+1)$ and $\alpha_i^{(N)}(t)$ for $B_i^{(N)}(t+1)$. By repeated applications of Lemma 3, each "centered" term will converge in probability to zero, and the desired results follow.