PAPER

Design Requirements for Mobile Communications Systems suitable for Transmitting Best-effort IP Packets

Hideki SATOH†, *Member*

SUMMARY We derived the design requirements that wireless systems and congestion control algorithms must satisfy to transmit best-effort Internet protocol (IP) packets over wireless systems. We proved that, if these requirements are satisfied, congestion control algorithms are robust against unfairness in the systems and can provide near-maximum throughputs in various environments. From the viewpoint of the design requirements, we investigated the effect of automatic repeat request (ARQ) on the throughputs of best-effort IP connections, and showed why ARQ can improve the throughputs while too large a number of retransmissions degrade them. We also investigated the effect of variance in packet transmission rates and clarified what kind of congestion control algorithm degrades the throughputs.

key words: mobile, ARQ, TCP, IP, best-effort, congestion control

1. Introduction

Internet protocol (IP) packet communications over wireless links have been extensively studied to simultaneously accommodate high-speed data traffic and real-time traffic. Recently, it was pointed out that the burstness in transmission control protocol (TCP) traffic [1] degrades the quality of service (QoS) of real-time traffic over wireless links [2].

To improve the QoS, several slot-assignment algorithms that assign different slots to each QoS class have been proposed [2], [3]. Conventional transmission power control algorithms [4] and the above-mentioned slot-assignment algorithms are designed so that the bit error ratio is equal to a pre-determined target value. Thus, the packet transmission rates of best-effort IP connections, such as TCP connections, are restricted by the pre-determined target value; hence, high bandwidth utilization cannot be obtained [5]. To solve this problem, we developed a novel transmission power control algorithm for best-effort IP packet communications and an optimum parameter design algorithm [5]–[7]. The proposed control algorithm allows each mobile terminal to send packets to arbitrary slots without negotiation or the use of the ALOHA protocol. The bandwidth utilization achieved with this method was evaluated in fading and multi-cell environments, and it was shown that the bandwidth is used efficiently and that the QoS of real-time traffic can be guaranteed without the need to modify the TCP protocol [8], [9] or use a snoop agent [10].

It was shown that the estimation error of propagation loss degrades the throughputs of TCP connections [7]. To overcome this problem, automatic repeat request (ARQ) was used to retransmit packets over wireless links, and it was shown that ARQ can improve the throughputs. However, too large a number of retransmissions degrade the throughputs, which means that we must determine the optimum number of retransmissions depending on the channel conditions [7]. It was also shown that the throughput is dependent on congestion control algorithms and that the maximum throughput and QoS guaranteed congestion control algorithm (MAQS) provides a larger throughput than the conventional congestion control algorithm Reno does [5], [11]. However, we were unable to explain why the throughput is dependent on the type of congestion control algorithm and the number of retransmissions of ARQ. As a result, we could not determine the optimum number of retransmissions or the optimum parameters of congestion control algorithms.

In this paper, we report on our work to clarify why the throughput changes and to derive the design requirements that congestion control algorithms and wireless systems must satisfy to transmit best-effort IP packets. In Sect. 2, we describe the mathematical model of congestion control algorithms. A transmission power control method suitable for best-effort IP packet communications is also described and is compared with the conventional one. In Sect. 3, we present the design requirements for wireless systems and congestion control algorithms that must be satisfied. We prove, if these requirements are satisfied, that congestion control algorithms are robust against unfairness in the systems and that we can obtain a near-maximum throughput in various environments. In Sect. 4, from the viewpoint of the design requirements, we clarify the effect of ARQ and the variance in the packet transmission rate on the throughputs of best-effort IP connections. Finally, our conclusion is presented in Sect. 5.

Manuscript received Octorver 26, 2001
Manuscript revised February 5, 2002

†This work was done when the author was with YRP Mobile Telecommunications Key Technology Research Laboratories Co., Ltd. He is now with Communication Platform Laboratories, Research and Development Center, Toshiba Corporation. E-mail:jamisato@m.ieice.org.

2. IP Communications over Wireless Systems

2.1 Mathematical Model of Congestion Control

A TCP receiver, upon receipt of a packet, sends an acknowledgement (ACK) packet back to the TCP sender; the sender detects that a packet loss event has occurred by using some duplicate ACKs or a transmission time-out. The conventional congestion control algorithm Reno decreases the packet transmission rate when the sender observes a packet loss event, and increases it each time an ACK is received. Therefore, the packet transmission rate depends on the loss event ratio. Note that multiple packets may be lost when a packet loss event is observed.

Let $\lambda_i(n_i)$ be the transmission rate of the n_ith packet of the ith connection over the TCP layer and $r_i(n_i)$ be the probability that a loss event is observed when the ith sender receives the ACK of the n_ith packet. To construct a mathematical model describing various congestion control algorithms, we use the following performance measure that was proposed to describe MAQS [5]:

$$J(n_i) = \lambda_i(n_i)(1 - \zeta\lambda_i(n_i)^b \hat{r}(n_i)), \tag{1}$$

where $\hat{r}_i(n_i)$ is the estimated value of $r_i(n_i)$, and $b \geqq 0$ and $\zeta > 0$ are constants. Assume that a congestion control algorithm determines $\lambda_i(n_i)$ by using the following control strategy:

Control strategy 1 (Congestion control)**:** Packet transmission rate $\lambda_i(n_i)$ is determined so as to maximize performance measure $J(n_i)$.

By applying the extremal method [14] to Eq. (1), $\lambda_i(n_i)$ that maximizes $J(n_i)$ is obtained as the solution of the following equation:

$$\begin{aligned}\frac{dJ(n_i)}{d\lambda_i(n_i)} &= 1 - \zeta(1+b)\lambda_i(n_i)^b\hat{r}_i(n_i) \\ &\quad - \zeta\lambda_i(n_i)^{1+b}d\hat{r}_i(n_i)/d\lambda(n_i) \\ &= 0.\end{aligned} \tag{2}$$

To solve the above equation, we approximately express $\hat{r}_i(n_i)$ by the kth term of the Taylor series of $r_i(n_i)$ that significantly affects $r_i(n_i)$ as follows [5]:

$$\hat{r}_i(n_i) = \hat{a}_i\lambda_i(n_i)^k, \tag{3}$$

where $\hat{a}_i$ is the unknown coefficient. Let ϕ^* be defined as follows:

$$\phi^* = ((k+b+1)\zeta)^{-1}. \tag{4}$$

By substituting Eqs. (3) and (4) into Eq. (2), we obtain

$$\phi^* = \lambda_i(n_i)^b\hat{r}_i(n_i). \tag{5}$$

Let $\bar{\lambda}_i$ be the average of $\lambda_i(n_i)$ in a steady state and $\bar{r}_i$ be the average of $r_i(n_i)$ in a steady state. Assume that $\hat{r}_i(n_i) = \bar{r}_i$ in a steady state. The above equation reduces to the following equation in a steady state:

$$\phi^* = \bar{\lambda}_i{}^b\bar{r}_i. \tag{6}$$

It is very important that the congestion control algorithm be TCP-friendly (i.e., the throughput obtained by the congestion control algorithm is the same as that obtained by Reno when they are passing through the same route) [13]. So, we determine b and ϕ^* to satisfy the above requirement as described below. The performance of Reno is described as follows [2], [5], [13]:

$$\frac{3\nu^2}{2\overline{RTT}^2} = \bar{\lambda}_i^2\bar{r}_i, \tag{7}$$

where ν is the packet length, $\overline{RTT}$ is the average round-trip time, and time-out is omitted for the convenience of analysis. Therefore, for TCP-friendly congestion control algorithms, we obtain $b = 2$ and

$$\phi^* = \frac{3\nu^2}{2\overline{RTT}^2}. \tag{8}$$

Parameter ζ is determined from Eqs. (4) and (8) for a given k. The optimum value of k can be determined by using a distributed control algorithm [11].

The above discussion shows that the performance of TCP-friendly congestion control algorithms is described by Eq. (6) and that they obey Control strategy 1. The difference between Reno and MAQS is in the algorithm's ability to achieve Control strategy 1. The former determines $\lambda_i(n_i)$ based on an additive-increase and multiplicative-decrease algorithm [13] and the latter determines it based on the minimum variance control [16].

2.2 Problem with Transmitting Best-effort IP Packets over Wireless Channels

The transmission of best-effort IP packets, such as TCP packets, over wireless channels has a problem when conventional transmission power control is used [5]. Let $\bar{e}_{\mathrm{R}i}$ be the packet loss ratio of the ith connection over a wireless link in a steady state†. The purpose of conventional transmission power control methods is to adjust the transmission power so that $\bar{e}_{\mathrm{R}i}$ is equal to constant target e^*_{R} [4]. Thus, the performance of conventional power control algorithms in a steady state is expressed by

$$\bar{e}_{\mathrm{R}i} = e^*_{\mathrm{R}}. \tag{9}$$

Let $\bar{N}_{\mathrm{PLE}}$ be the number of packets lost in a loss event in a steady state. Assume that $\bar{N}_{\mathrm{PLE}}$ is unity and that

†The ratio of the number of all wireless packets lost to the number of wireless packets sent over a wireless link.

ARQ is not used. We consider only the packet loss in the target wireless link and ignore all other packet losses [5]. Thus, $\bar{r}_i$ is obtained by

$$\bar{r}_i = \bar{e}_{\mathrm{R}i}. \tag{10}$$

By substituting Eqs. (9) and (10) into Eq. (6), we get

$$\bar{\lambda}_i \cong (\phi^*/e^*_{\mathrm{R}})^{1/b}. \tag{11}$$

Because Eq. (9) shows the lower bound of $\bar{e}_{\mathrm{R}i}$, and the packet losses other than the one in the target wireless link are neglected, Eq. (11) shows the upper bound of $\bar{\lambda}_i$ and implies that the upper bound of $\bar{\lambda}_i$ does not change with wireless network conditions because it is determined by constants ϕ^*, e^*_{R}, and b. Thus, the bandwidth utilization of wireless networks varies according to the number of connections in a cell, and the bandwidth is not used efficiently when there are only a few terminals, even if sufficient bandwidth remains.

2.3 Transmission Power Control Suitable for Transmitting TCP Packets

To solve the above problem, a novel transmission power control method was proposed [5]–[7]. Let $P_i(n_i)$ be the transmission power for the n_ith packet, $P_{\mathrm{INV}i}(n_i) = P_i(n_i)^{-1}$, and $e_{\mathrm{R}i}(n_i)$ be the loss probability of the n_ith packet over a wireless link. Let $A_{\mathrm{F}i}(n_i)$ be the actual attenuation (received power/transmitted power) that the n_ith packet suffers. We assume that estimated value $A_{\mathrm{E}i}(n_i)$ of $A_{\mathrm{F}i}(n_i)$ contains estimation error $\varepsilon_{\mathrm{E}i}(n_i)$ as follows:

$$A_{\mathrm{E}i}(n_i) = A_{\mathrm{F}i}(n_i)\varepsilon_{\mathrm{E}i}(n_i), \tag{12}$$

where $\varepsilon_{\mathrm{E}i}(n_i)$ can be approximately expressed by an independent random process that has a log normal distribution with mean 1 and standard deviation σ_{err}. Let $\psi_i(n_i)$ be the new state variable defined by

$$\psi_i(n_i) = (P_{\mathrm{INV}i}(n_i)A_{\mathrm{EINV}i}(n_i)^s)^\beta e_{\mathrm{R}i}(n_i), \tag{13}$$

where $\beta \geqq 0$ and s are constants, and $A_{\mathrm{EINV}i}(n_i) = A_{\mathrm{E}i}(n_i)^{-1}$. Constant s is used to adjust the power received at the base station (BS) and decrease interference in multi-cell environments [7]. The novel transmission power control algorithm adjusts $P_i(n_i)$ according to the following control strategy:

Control strategy 2 (Transmission power control)**:** $P_i(n_i)$ is adjusted so that state variable $\psi_i(n_i)$ tracks reference input ψ^*.

If $P_{\mathrm{INV}i}(n_i)$, $A_{\mathrm{EINV}i}(n_i)$, and $e_{\mathrm{R}i}(n_i)$ converge to $\bar{P}_{\mathrm{INV}i}$, $\bar{A}_{\mathrm{EINV}i}$, and $\bar{e}_{\mathrm{R}i}$, respectively, in a steady state, Eq. (13) also converges to the following equation:

$$\psi^* = (\bar{P}_{\mathrm{INV}i}\bar{A}_{\mathrm{EINV}i}{}^s)^\beta \bar{e}_{\mathrm{R}i}. \tag{14}$$

Thus, the relationship between $\bar{e}_{\mathrm{R}i}$ and $\bar{P}_{\mathrm{INV}i}$ is determined by β, s, and ψ^*. For example, when $\beta = 0$, we obtain $\bar{e}_{\mathrm{R}i} = \psi^*$. This is the performance of the conventional transmission power control method described in Eq. (9), and ψ^* is the target of $\bar{e}_{\mathrm{R}i}$. Parameter β and reference input ψ^* are determined by the optimum parameter design algorithm [7].

Control strategy 2 works as follows. Let L be the number of TCP mobile terminals and $\bar{P}_{\mathrm{rvINV}i}$ be the reciprocal of average received power at the BS in a steady state. To use the bandwidth efficiently, $\bar{\lambda}_i$ must increase as L decreases. Assume that $\bar{N}_{\mathrm{PLE}} = 1$, that ARQ is not used, and that only the packet loss in the target wireless link is taken into account in the same manner as in Sect. 2.2 for the convenience of analysis. In a steady state, the system is described by

$$\bar{e}_{\mathrm{R}i} = f_{\mathrm{R}}(\bar{\lambda}_i, \bar{P}_{\mathrm{rvINV}i}, L), \tag{15}$$

$$\psi^* = \bar{P}_{\mathrm{rvINV}i}^{\ \beta}\bar{e}_{\mathrm{R}i}, \tag{16}$$

where $f_{\mathrm{R}}(\cdot)$ is a monotone increasing function of $\bar{\lambda}_i$, $\bar{P}_{\mathrm{rvINV}i}$, and L. By eliminating $\bar{e}_{\mathrm{R}i}$, $\bar{r}_i$, and $\bar{P}_{\mathrm{rvINV}i}$ from Eqs. (6), (10), (15), and (16), we get

$$\begin{aligned}\phi^* &= \bar{\lambda}_i^b f_{\mathrm{R}}(\bar{\lambda}_i, (\psi^*(\bar{\lambda}_i^b/\phi^*))^{1/\beta}, L)\\ &= f_{\mathrm{R}'}(\bar{\lambda}_i, L),\end{aligned} \tag{17}$$

where $f_{\mathrm{R}'}(\cdot)$ is a monotone increasing function of $\bar{\lambda}_i$ and L. Because ϕ^* is a constant, Eq. (17) shows that $\bar{\lambda}_i$ increases as L decreases. Therefore, Control strategy 2 enables high bandwidth utilization by setting ϕ^*, b, ψ^*, and β to appropriate values.

3. Design Requirements for Mobile IP Communications Systems

It was shown that the parameter design algorithm for the novel transmission power control method can determine ψ^* and β to provide a near-maximum throughput [7]. However, estimation error $\varepsilon_{\mathrm{E}i}(n_i)$ and the variance in the packet transmission rate affect the performance curve of wireless systems, and large error and variance degrade the throughputs of best-effort IP packet connections. In this section, the sufficient conditions (i.e., design requirements) that congestion control algorithms and wireless systems must satisfy are derived.

3.1 Robustness against unfairness

Consider L connections with the same path. Let $f_{\mathrm{mean}}(\sum_{j=1}^{L}\lambda_j(n_j))$ be the mean loss event probability and $f_i(\lambda_i(n_i))$ be the loss event probability of the n_ith packet of the ith sender. Loss event ratio $\bar{r}_i$ may be different for each connection even if all connections pass through the same path. This difference is referred to, in this paper, as unfairness in the systems, and it arises from the packet length, priority control, and queueing discipline. This section shows a sufficient condition for congestion control algorithms to be robust against unfairness. We express the relationship between $\bar{r}_i$ and $\bar{\lambda}_i$ by the following equation:

$$\bar{r}_i = f_i(\bar{\lambda}_i) = \theta_i f_{\text{mean}}(\sum_{j=1}^{L} \bar{\lambda}_j), \tag{18}$$

where θ_i is the unfairness factor in the system. By substituting Eq. (18) into Eq. (6), and eliminating $f_{\text{mean}}(\cdot)$, $\bar{r}_i$, and $\bar{r}_j$, we obtain

$$\bar{\lambda}_i/\bar{\lambda}_j = (\theta_j/\theta_i)^{1/b}. \tag{19}$$

This paper says that congestion control is robust against unfairness if $\bar{\lambda}_i$ is almost equal to $\bar{\lambda}_j$ for any initial state, i and $j \in \mathbf{L} = \{1, 2, \cdot, L\}$, $i \neq j$, and sufficiently small $|\theta_j - \theta_i| \neq 0$. Under the condition that $|\theta_j - \theta_i|$ be sufficiently small and not be equal to zero, Eq. (19) shows that if $b = 0$, $\bar{\lambda}_i/\bar{\lambda}_j$ becomes zero or infinity. In contrast, if $b > 0$, we obtain $\bar{\lambda}_i \cong \bar{\lambda}_j$ (i.e., the congestion control algorithm is robust against unfairness). Thus, we obtain the sufficient condition for a congestion control algorithm to be robust against unfairness as follows:

Theorem 1 (Robustness against unfairness)**:** The sufficient condition for a congestion control algorithm described by Control strategy 1 to be robust against unfairness is that the following inequality holds:

$$b > 0. \tag{20}$$

3.2 Control Error

Let $\lambda^*{}_i$ be the solution of Eq. (2) in a steady state when the ith sender knows $f_i(\cdot)$ (i.e., $\hat{r}_i(n_i) = f_i(\lambda_i(n_i))$). Because Eq. (6) is derived using Eq. (3), control error ε_i defined by

$$\varepsilon_i = \bar{\lambda}_i - \lambda^*{}_i \tag{21}$$

may arise. Let $f_i'(\lambda)$ be denoted by $df_i(\lambda)/d\lambda$. With respect to ε_i, the following theorem holds:

Theorem 2 (Control Error)**:** If Eq. (20) and the following inequalities are satisfied:

$$f_i(\lambda^*{}_i) \ll \lambda^*{}_i f_i'(\lambda^*{}_i), \tag{22}$$
$$b \ll \lambda^*{}_i f_i'(\lambda^*{}_i)/f_i(\lambda^*{}_i), \tag{23}$$
$$k \geqq 1, \tag{24}$$

we obtain a sufficiently small control error as follows:

$$|\varepsilon_i| \ll \lambda^*{}_i. \tag{25}$$

That is, the sufficient condition for Eq. (25) to hold is that Eqs. (20), (22), (23), and (24) hold.

Proof: By substituting $f_i(\lambda^*{}_i)$ and Eq. (21) into Eqs. (2) and (6), we obtain

$$1 - \zeta(1+b)\lambda^*{}_i{}^b f_i(\lambda^*{}_i) - \zeta\lambda^*{}_i{}^{b+1} f_i'(\lambda^*{}_i) = 0, \tag{26}$$
$$(\lambda^*{}_i + \varepsilon_i)^b = \frac{\phi^*}{f_i(\lambda^*{}_i + \varepsilon_i)}. \tag{27}$$

The first approximation to Eq. (27) is expressed as

$$\lambda^*{}_i{}^b + \varepsilon_i b \lambda^*{}_i{}^{b-1} = \frac{\phi^*}{f_i(\lambda^*{}_i) + \varepsilon_i f_i'(\lambda^*{}_i)}. \tag{28}$$

By rearranging the above equation for ε_i, we obtain the following quadratic equation:

$$\varepsilon_i^2 b\lambda^*{}_i{}^{b-1} f_i'(\lambda^*{}_i) + (\lambda^*{}_i{}^b f_i'(\lambda^*{}_i) + b\lambda^*{}_i{}^{b-1} f_i(\lambda^*{}_i))\varepsilon_i + \lambda^*{}_i{}^b f_i(\lambda^*{}_i) - \phi^* = 0.$$

Solving the above equation, we obtain

$$\varepsilon_i = \frac{-\lambda^*{}_i{}^b f_i'(\lambda^*{}_i) - b\lambda^*{}_i{}^{b-1} f_i(\lambda^*{}_i) \pm \{[\lambda^*{}_i{}^{b-1}(\lambda^*{}_i f_i'(\lambda^*{}_i) - b f_i(\lambda^*{}_i))]^2 + 4b\lambda^*{}_i{}^{b-1} f_i'(\lambda^*{}_i)\phi^*\}^{1/2}}{2b\lambda^*{}_i{}^{b-1} f_i'(\lambda^*{}_i)}. \tag{29}$$

From the above equation, we can prove that Eq. (25) holds if Eqs. (20), (22), (23), and (24) are true, and that Theorem 2 holds (see Appendix A). The proof of Theorem 2 is now complete.

3.3 Throughput

Let $\lambda^*_{\max}$ be the theoretical value of the packet transmission rate that provides the maximum throughput and $\bar{\lambda}_{\text{total}}$ be the total throughput obtained by congestion control algorithms that obey Control strategy 1. We obtain the following corollary with respect to the total throughput:

Corollary 1 (Maximum throughput)**:** The sufficient condition for the total throughput obtained by congestion control algorithms that obey Control strategy 1 to be near-maximum (the difference between the total throughput obtained by $\lambda^*_{\max}$ and that obtained by $\bar{\lambda}_{\text{total}}$ is sufficiently small) is that the sufficient condition in Theorem 2 holds and ζ is sufficiently large.

Proof: Applying Eq. (19) to $\bar{\lambda}_{\text{total}} = \sum_{j=1}^{L} \bar{\lambda}_j$, we get

$$\bar{\lambda}_{\text{total}} = \sum_{j=1}^{L} (\theta_i/\theta_j)^{1/b} \bar{\lambda}_i = \Gamma_i \bar{\lambda}_i, \tag{30}$$

where Γ_i is $\sum_{j=1}^{L} (\theta_i/\theta_j)^{1/b}$. By multiplying both sides of Eq. (6) by Γ_i and substituting Eq. (18) into Eq. (6), we obtain

$$(\Gamma_i \bar{\lambda}_i)^b = \frac{\phi^* \Gamma_i^b}{\theta_i f_{\text{mean}}(\Gamma_i \bar{\lambda}_i)} = \frac{\Phi^*}{f_{\text{mean}}(\Gamma_i \bar{\lambda}_i)}, \tag{31}$$

where $\Phi^* = (\phi^* \Gamma_i^b)/\theta_i$. Let b_0 be a sufficiently small positive number and $\lambda^*_{\max}$ be the solution that maximizes performance measure J when $\zeta = 1$, $b = b_0$, $L = 1$, and $\hat{r}_1(n_1) = f_{\text{mean}}(\lambda_1(n_1))$. From the definition of J, $\lambda^*_{\max}$ provides the maximum throughput and satisfies the following equation:

$$1 - (1 + b_0){\lambda^*_{\max}}^{b_0} f_{\text{mean}}(\lambda^*_{\max}) \\ -\zeta {\lambda^*_{\max}}^{b_0+1} {f_{\text{mean}}}'(\lambda^*_{\max}) = 0. \tag{32}$$

Let ε be defined by

$$\varepsilon = \lambda^*_{\max} - \bar{\lambda}_{\text{total}}. \tag{33}$$

By substituting Eq. (33) into Eq. (31), we obtain

$$(\lambda^*_{\max} + \varepsilon)^{b_0} = \frac{\Phi^*}{f_{\text{mean}}(\lambda^*_{\max} + \varepsilon)}. \tag{34}$$

Theorem 2 holds independently of the values of $b > 0$, ϕ^*, and θ_i. Therefore, by comparing Eqs. (32) and (34) with Eqs. (26) and (27), we obtain

$$|\varepsilon| \ll \lambda^*_{\max}, \tag{35}$$

if the sufficient condition in Theorem 2 holds.

Let J_0 be the total throughput obtained by congestion control algorithms that obey Control strategy 1. By applying Eqs. (6), (19), and (30) to the definition of the total throughput, we obtain the following equation:

$$J_0 = \sum_{j=1}^{L} \bar{\lambda}_j (1 - \bar{r}_j) \\ = \bar{\lambda}_{\text{total}} - ((\bar{\lambda}_{\text{total}}/\Gamma_i) \sum_{j=1}^{L} (\theta_i/\theta_j)^{1/(b-1)}) \bar{r}_i. \tag{36}$$

By substituting Eqs. (4), (30), and (33) into Eq. (6), we obtain the following equation:

$$\bar{r}_i = \phi^*/\bar{\lambda}_i^b \\ = \frac{\Gamma_i^b}{(k+b+1)\zeta(\lambda^*_{\max} - \varepsilon.)^b}. \tag{37}$$

Equation (35) holds for any ζ and L. However, $\bar{r}_i$ increases as ζ decreases as shown in Eq. (37), and J_0 decreases as $\bar{r}_i$ increases as shown in Eq. (36), even if Eq. (35) holds. Thus, to obtain a sufficiently small value of $\bar{r}_i$ and a near-maximum throughput, it is necessary for ζ to be sufficiently large. The proof of Corollary 1 is now complete.

Corollary 1 holds for any L. Thus, we obtain a near-maximum throughput for any L if the sufficient condition described in Corollary 1 holds. However, the sufficient condition in Corollary 1 does not hold if ζ is small, and the ζ of a TCP-friendly congestion control algorithm decreases as RTT decreases because the following equation holds from Eqs. (4) and (8):

$$\zeta = \frac{2\overline{RTT}^2}{3\nu^2(k+b+1)}. \tag{38}$$

Thus, we may not obtain a near-maximum throughput if RTT is short. This property is shown in Fig. 5 in Sect. 4.

Consider a conventional transmission power control algorithm that adjusts the transmission power so that $\bar{e}_{\text{R}i} = e^*_{\text{R}}$. This leads to $f'_i(\bar{\lambda}_i) = 0$ for any $\bar{\lambda}_i$, and hence, Eq. (22) does not hold. Thus, a large throughput is not obtained in most cases as described in Sect. 2.2. In contrast, a transmission power control algorithm that obeys Control strategy 2 provides the following inequality from Eqs. (15) and (17):

$$f'_i(\bar{\lambda}_i) = df_{\text{R}}(\bar{\lambda}_i, (\psi^*(\bar{\lambda}_i^b/\phi^*))^{1/\beta}, L)/d\bar{\lambda}_i \\ > 0. \tag{39}$$

Thus, we can obtain high bandwidth utilization by setting ϕ^*, b, ψ^*, and β to appropriate values as described in Sect. 2.3. This property is discussed using Figs. 8 and 9 in Sect. 4.

4. Performance Evaluation

4.1 Network Configurations

We use a network model composed of a BS with a 30-m-high antenna, wired networks with two routers, L_{M} (4, 13, and 37) mobile terminals, L_{C} fixed terminals in Router-1, and their corresponding receivers in Router-2, as shown in Fig. 1. The channel capacities of the wired links are 100 Mb/s, and packet losses occur over the wireless channels or in Router-1.

The forward error correction (FEC) encoder uses random coding [5], [15] and maps 32 bytes of information into a 64-byte wireless packet (i.e., coding rate R_{FEC}=0.5). This packet is modulated by binary phase-shift keying modulation for data modulators and spread spectrum modulators, and transmitted over a slotted direct-sequence code-division multiple access uplink. The transmission power is adjusted according to Control strategy 2, and parameters β, ψ^*, and s are set to 16.2, 4.76×10^{31}, and 1, respectively, by using a previously proposed design algorithm [7]. We use a selective-repeat ARQ to mask the loss of wireless packets from the TCP layer, where the maximum number of retransmissions is set to N_{ARQ}. We assume that each mobile terminal has a sufficiently large number of primary pseudo-noise (PN) codes with a chip rate of 20 Mc/s. The processing gain is set to 50. Each packet over the wired network consists of information parts of 16 wireless packets (i.e., the packet length over the wired network is 512 bytes). We ignore the overheads, such as packet headers. Packets from the terminals arrive at the receivers with propagation delays of D_1 ms. We assume that the terminal receives ACK packets without any loss and that they suffer the same propagation delays as the data packets do.

Each mobile terminal is located so that the distance between the closest terminals is d_{term} (see Fig.

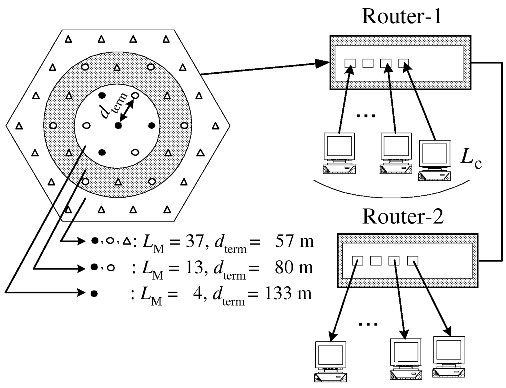

Fig. 1 Network and cell configuration.

1). The signal received at the BS is attenuated by distance (propagation loss coefficient is 3.5) and random Rayleigh fading. In order for the congestion control algorithm to work properly, the packet loss must reflect the network congestion and the packet must be transmitted according to the packet transmission rate determined by the congestion control algorithm. Hence, we use a random access protocol, which allows each mobile terminal to send packets to arbitrary wireless slots without negotiation with the BS [5]–[7]. The throughput is denoted by the bit rate of the received packet. The throughput and packet transmission rate are normalized by the product of the supremum of the channel capacity of the wireless link (20 Mb/s) and R_{FEC} (0.5). Standard deviation σ_λ in the packet transmission rate is normalized by the average packet transmission rate.

4.2 Effect of Performance Curve of Wireless Systems on Best-Effort IP Packet Communications

Corollary 1 shows that congestion control algorithms that obey Control strategy 2 provide near-maximum throughputs for any L if the sufficient condition described in Corollary 1 holds. We verified this over the network described in Fig. 1, and investigated the design requirements for wireless systems.

To evaluate the control error caused by the approximation of Eq. (3), the effect of k (the degree of the term in the Taylor series of $r_i(n_i)$ used in Eq. (3)) on $\lambda^*{}_i f_i'(\lambda^*{}_i)/f_i(\lambda^*{}_i)$ and $\varepsilon_i/\lambda^*{}_i$ (i.e., Eqs. (22) and (25)) was evaluated over the wired network with $f_i(\cdot)$ modeled by $M/M/1/K$ (that is, packets arrive at Router-1 from fixed terminals according to the Poisson process and their service interval obeys an exponential distribution). Figure 2 shows, for any $k \geqq 1$, that $\lambda^*{}_i f_i'(\lambda^*{}_i)/f_i(\lambda^*{}_i)$ is sufficiently large (i.e., Eq. (22) holds), and that $\varepsilon_i/\lambda^*{}_i$ is sufficiently small. Therefore, we find that Eq. (3) is a fairly good approximation.

We evaluated MAQS and Reno for $\sigma_{\mathrm{err}} = 0$ and $N_{\mathrm{ARQ}} = 0$ over a wireless system (i.e., $L_{\mathrm{C}} = 0$, the queueing delays in the Routers are zero, and packets are lost over wireless channels) in homogeneous environments (all terminals use the same congestion control algorithm). Figures 3 and 5 show that MAQS provides a larger throughput than Reno does †, and that variance

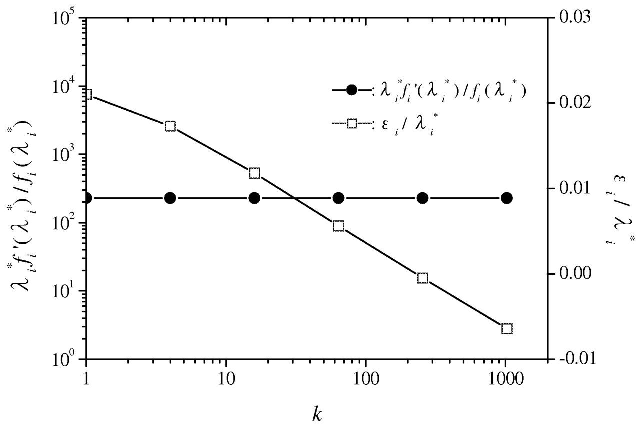

Fig. 2 Effect of k on $\lambda^*{}_i f_i'(\lambda^*{}_i)/f_i(\lambda^*{}_i)$ and control error ε_i.

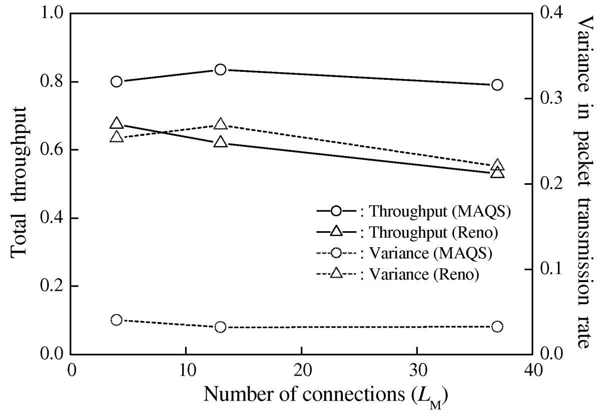

Fig. 3 Effect of number of connections on packet throughput and variance in packet transmission rate ($D_1 = 20$ ms).

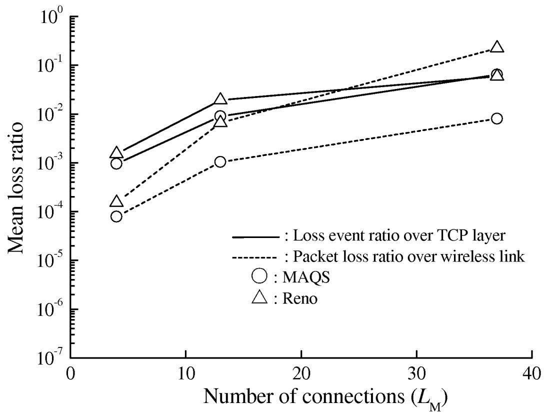

Fig. 4 Effect of number of connections on loss event ratio ($D_1 = 20$ ms).

†We showed that MAQS provides almost the same or a slightly smaller throughput than Reno does (i.e., MAQS is TCP-friendly) in heterogeneous environments (connections using different congestion control algorithms are mul-

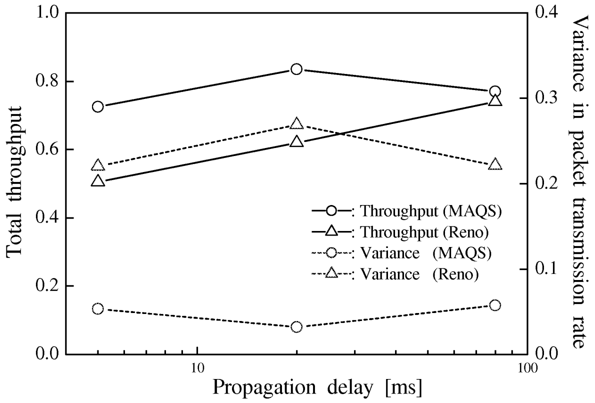

Fig. 5 Effect of propagation delay on packet throughput and variance in packet transmission rate ($L_{\rm M} = 13$).

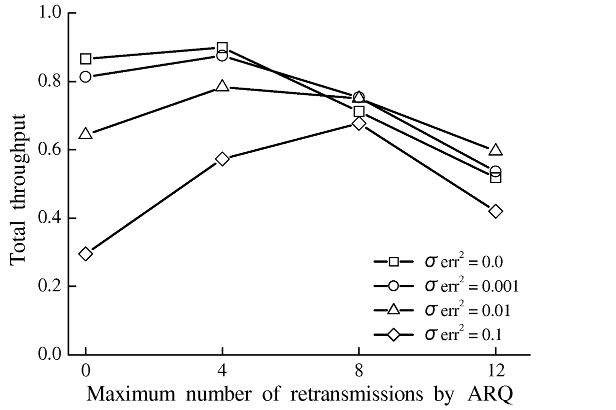

Fig. 7 Effect of ARQ on throughput for various values of variance in estimation error ($D_1 = 20$ ms, $L_{\rm M} = 13$).

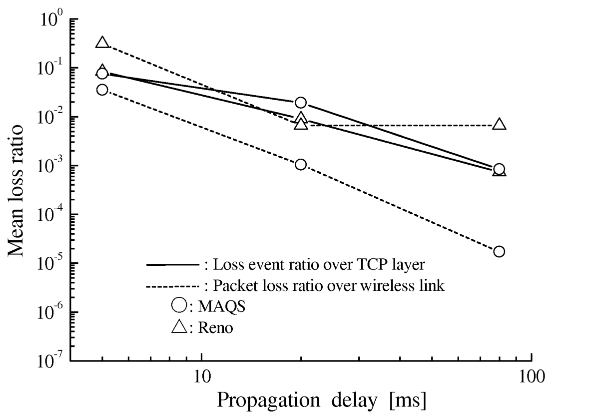

Fig. 6 Effect of propagation delay on loss event ratio ($L_{\rm M} = 13$).

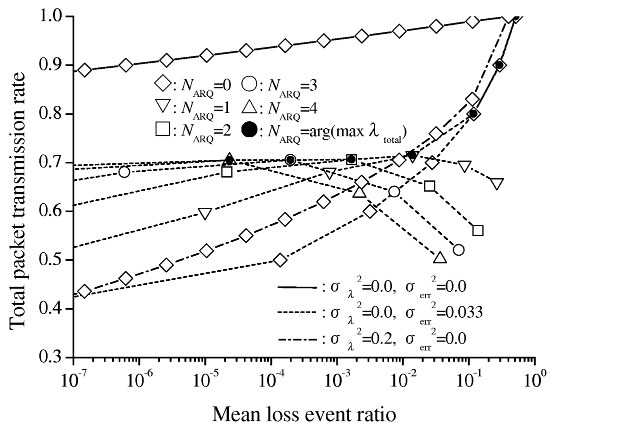

Fig. 8 Effect of ARQ and variance in packet transmission rate on loss event ratio.

${\sigma_\lambda}^2$ of MAQS is smaller than that of Reno. The mean loss event ratios of MAQS and Reno are almost equal; however, mean packet loss ratio $\bar{e}_{{\rm R}i}$ of Reno is larger than that of MAQS, as shown in Figs. 4 and 6. For $L_{\rm M} = 37$ or $D_1 = 5$ ms, the $\bar{e}_{{\rm R}i}$ of Reno exceeds 10^{-1} and lost packets degrade the total throughput. This is a reason why the total throughput of Reno is smaller than that of MAQS. However, when $\bar{e}_{{\rm R}i}$ is sufficiently small (for example, $L_{\rm M} = 4$ in Fig. 4), the throughput of Reno is also smaller than that of MAQS. This shows that there are other causes for the total throughput to be small besides the packet loss ratio. The causes are discussed below using Figs. 8 and 9.

The effect of ARQ on the total throughput of MAQS was evaluated over the wireless system. Figure 7 shows that ARQ becomes more effective as ${\sigma_{\rm err}}^2$ increases. However, when $N_{\rm ARQ}$ exceeds 4 or 8, the throughput decreases [7].

To explain the effect of ARQ and the other causes for the throughput of Reno to be smaller than that of MAQS in homogeneous environments, we evaluated packet transmission rate λ_i of the ith terminal over the TCP layer and mean loss event ratio r for various values of the packet transmission rate over a wireless link. The relationship between $\lambda_{\rm total} = \sum \lambda_i$ and r is shown in Fig. 8. When ${\sigma_\lambda}^2 = {\sigma_{\rm err}}^2 = 0$, $\lambda_{\rm total}$ is large for various values of r. In contrast, when ${\sigma_\lambda}^2 = 0.2$, $\lambda_{\rm total}$ is smaller than it is when ${\sigma_\lambda}^2 = {\sigma_{\rm err}}^2 = 0$. This is another reason why Reno, whose ${\sigma_\lambda}^2$ is larger than that of MAQS, provides a smaller throughput than MAQS does. When ${\sigma_{\rm err}}^2 = 0.033$ and $N_{\rm ARQ} = 0$, $\lambda_{\rm total}$ is also smaller than it is when ${\sigma_\lambda}^2 = {\sigma_{\rm err}}^2 = 0$. As $N_{\rm ARQ}$ increases, $\lambda_{\rm total}$ increases while r is small. However, $\lambda_{\rm total}$ decreases as $N_{\rm ARQ}$ increases while r exceeds a certain value. These results show that although ARQ reduces r, the number of packets retransmitted over the wireless link increases and $\lambda_{\rm total}$ decreases when r is large. This is the reason why ARQ improves the throughputs of best-effort IP connections and why too large a number of retransmissions degrade them.

We can also explain this from the viewpoint of Corollary 1 by using Fig. 9. When ${\sigma_\lambda}^2 = {\sigma_{\rm err}}^2 = 0$ and $N_{\rm ARQ} = 0$, $\lambda_i f_i'(\lambda_i)/f_i(\lambda_i)$ is large for various values of r. However, when ${\sigma_{\rm err}}^2$ or ${\sigma_\lambda}^2 > 0$ and $N_{\rm ARQ} = 0$,

tiplexed at the same time) [11].

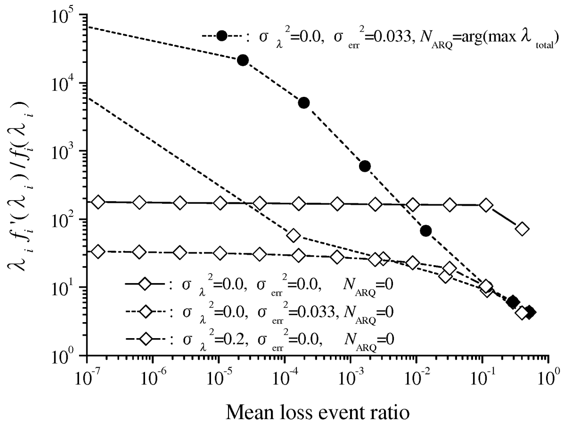

Fig. 9 Effect of ARQ and variance in packet transmission rate on $\lambda_i f_i'(\lambda_i)/f_i(\lambda_i)$.

$\lambda_i f_i'(\lambda_i)/f_i(\lambda_i)$ is smaller for $r > 10^{-4}$ (the loss event ratio obtained in Figs. 4 and 6). This means that the sufficient condition in Corollary 1 does not hold and that the throughput decreases. To improve the throughput, N_{ARQ} must be adjusted. For example, we selected N_{ARQ} that maximizes λ_{total}, and described the λ_{total}–r curve and $\lambda_i f_i'(\lambda_i)/f_i(\lambda_i)$ in Figs. 8 and 9, respectively, (they are indicated by arg(maxN_{ARQ})). As a result of this adjustment, $\lambda_i f_i'(\lambda_i)/f_i(\lambda_i)$ increased as shown in Fig. 9. Using this adjustment, it is possible to obtain an even larger throughput. Developing an algorithm to adjust N_{ARQ} is the subject of our future work.

Figure 5 shows that the total throughput of MAQS when $D_1 = 80$ ms is smaller than it is when $D_1 = 20$ ms, although the packet loss ratio when $D_1 = 80$ ms is smaller than the packet loss ratio when $D_1 = 20$ ms. This is because (1) the response time before the packet transmission rate is reduced according to the network conditions increases as D_1 increases, (2) ${\sigma_\lambda}^2$ increases as the response time increases as shown in Fig. 5, and (3) the sufficient condition in Corollary 1 is not satisfied when ${\sigma_\lambda}^2$ is large as shown in Figs. 8 and 9. In contrast, the total throughput of Reno when $D_1 = 20$ ms is larger than it is when $D_1 = 5$ ms, although ${\sigma_\lambda}^2$ also increases. This is because the packet loss ratio when $D_1 = 5$ ms is too large as shown in Fig. 6. From the above discussion, we find that the total throughput of best-effort IP connections is affected by both the packet loss ratio and the λ_{total}–r curve as shown in Corollary 1.

5. Conclusion

We derived the design requirements that wireless systems and congestion control algorithms must satisfy when best-effort IP packets are transmitted over wireless systems. We investigated the effects of ARQ and the variance in the packet transmission rate on the throughputs of best-effort IP connections over wireless systems. From the viewpoint of the design requirements described in Corollary 1, we showed why in homogeneous environments, the throughput obtained by Reno is smaller than that obtained by MAQS while in heterogeneous environments, MAQS provides almost the same or a slightly smaller throughput than Reno does. We clarified the mechanism by which ARQ improves the throughputs of best-effort IP connections and explained why too large a number of retransmissions degrades the throughputs.

References

[1] W. R. Stevens, TCP/IP Illustrated, vol. 1: The Protocols, Addison-Wesley, New York, 1994.
[2] H. M. Chaskar, T. V. Lakshman, and U. Madhow, "TCP over wireless with link level error control: analysis and design methodology," IEEE/ACM Trans. Networkings, vol. 7, no. 5, pp. 605–615, Oct. 1999.
[3] S. Choi and K. G. Shin, "An uplink CDMA system architecture with diverse QoS guarantees for heterogeneous traffic," IEEE/ACM Trans. Networking. vol. 7, no. 5, pp. 616–628, Oct. 1999.
[4] S. C. Yang, CDMA RF System Engineering, Artech House, USA, 1998.
[5] H. Satoh and T. Kobayashi, "Distributed power control for real-time and TCP packet communications over CDMA wireless link," IEICE, Trans. Fundamentals, vol. E84-A, no. 7, pp. 1702–1713, July 2001.
[6] M. Ishiba, H. Satoh, and T. Kobayashi, "Multi-class packet multiplexing scheme for wireless Internet over CDMA systems," Proc. 2001 IEEE Radio and Wireless Conference, pp. 73–76, Aug. 2001.
[7] H. Satoh, M. Ishiba, and T. Kobayashi, "CDMA transmission power control at mobile terminals for IP packet communications in fading and multi-cell environments," IEICE, Trans. Fundamentals, vol. E84-A, no. 12, pp. 3058–3067, Dec. 2001.
[8] J. W. Mark, X. Shen, Y. Zeng, and J. Pan, "TCP/IP in wireless/Internet interworking," Conference Proc. 3Gwireless'2000, pp. 438–445.
[9] M. Stangel and V. Bharghavan, "Improving TCP performance in mobile computing environments," Proc. IEEE ICC'98, pp. 584–589, June 1998.
[10] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, and R. H. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," IEEE/ACM Tans. Networking, vol. 5, no. 6, pp. 756–769, Dec. 1997.
[11] H. Satoh and M. Ishiba, "Distributed control of transient response for best-effort IP packet communications over wireless systems," IEICE Trans. Fundamentals, July 2002, to be published.
[12] W. Stevens, "TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms," RFC2001, Jan. 1997.
[13] S. Floyd, M. Handley, J. Padhey, and J. Widmer, "Equation-based congestion control for unicast applications," Comput. Commun. Rev., vol. 30, no. 4, pp. 43–56, Oct. 2000.
[14] J. S. Rustagi, Optimization Techniques in Statistics, Academic Press, London, 1994.
[15] C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana and Chicago, 1963.

[16] K. J. Astrom, Introduction to Stochastic Control Theory, Academic Press, London, 1970.

Appendix A: Boundary of Control Error

This section derives the sufficient condition for $|\varepsilon_i|$ of Eq. (29) to be sufficiently small.

Consider a larger solution of Eq. (29). By eliminating $4b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)\phi^*$ from the numerator of Eq. (29), Eq. (29) reduces to the following inequality:

$$\begin{aligned}\varepsilon_i &> \frac{-{\lambda^*}_i^{b} f_i'({\lambda^*}_i) - b{\lambda^*}_i^{b-1} f_i({\lambda^*}_i) + {\lambda^*}_i^{b} f_i'({\lambda^*}_i) - b{\lambda^*}_i^{b-1} f_i({\lambda^*}_i)}{2b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)} \\ &= -f_i({\lambda^*}_i)/f_i'({\lambda^*}_i).\end{aligned}$$

Let us assume that Eq. (22) holds. Then, we obtain

$$\varepsilon_i \gg -{\lambda^*}_i. \qquad (\text{A}\cdot 1)$$

Consider a larger solution of Eq. (29). Assume that Eqs. (20) and (23) hold. Then, we obtain

$$\begin{aligned}&({\lambda^*}_i f_i'({\lambda^*}_i) - bf_i({\lambda^*}_i))^2 \\ &= ({\lambda^*}_i f_i'({\lambda^*}_i))^2 + bf_i({\lambda^*}_i)(-2{\lambda^*}_i f_i'({\lambda^*}_i) + bf_i({\lambda^*}_i)) \\ &< ({\lambda^*}_i f_i'({\lambda^*}_i))^2 + bf_i({\lambda^*}_i)(-{\lambda^*}_i f_i'({\lambda^*}_i) + bf_i({\lambda^*}_i))\end{aligned}$$

which upon substituting Eq. (23) into the above equation yields

$$< ({\lambda^*}_i f_i'({\lambda^*}_i))^2.$$

By substituting the above equation into Eq. (29), the following inequality is obtained:

$$\begin{aligned}\varepsilon_i &< \frac{-{\lambda^*}_i^{b} f_i'({\lambda^*}_i) - b{\lambda^*}_i^{b-1} f_i({\lambda^*}_i) + \{({\lambda^*}_i^{b} f_i'({\lambda^*}_i))^2 + 4b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)\phi^*\}^{1/2}}{2b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)} \\ &< \frac{-{\lambda^*}_i^{b} f_i'({\lambda^*}_i) - b{\lambda^*}_i^{b-1} f_i({\lambda^*}_i) + \{({\lambda^*}_i^{b} f_i'({\lambda^*}_i))^2 + 4b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)\phi^* + (2{\lambda^*}_i^{-1} b\phi^*)^2\}^{1/2}}{2b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)} \\ &= \frac{2\phi^* - {\lambda^*}_i^{b} f_i({\lambda^*}_i)}{2f_i'({\lambda^*}_i){\lambda^*}_i^{b}} \\ &< \frac{1}{\zeta(1+b+k) f_i'({\lambda^*}_i){\lambda^*}_i^{b}}. \qquad (\text{A}\cdot 2)\end{aligned}$$

By substituting Eqs. (22) and (A·2) into Eq. (26), we obtain

$$\begin{aligned}1 &= \zeta(1+b){\lambda^*}_i^{b} f_i({\lambda^*}_i) + \zeta{\lambda^*}_i^{b+1} f_i'({\lambda^*}_i) \\ &\ll \zeta(1+b){\lambda^*}_i^{b}{\lambda^*}_i f_i'({\lambda^*}_i) + \zeta{\lambda^*}_i^{b+1} f_i'({\lambda^*}_i) \\ &= \zeta(1+b+1){\lambda^*}_i^{b+1} f_i'({\lambda^*}_i).\end{aligned}$$

Assume that Eq. (24) holds. Substituting the above equation into Eq. (A·2) yields the following inequality:

$$\varepsilon_i \ll \frac{(2+b)}{(1+k+b)}{\lambda^*}_i \leqq {\lambda^*}_i. \qquad (\text{A}\cdot 3)$$

Consider a smaller solution of Eq. (29). Equation (29) reduces to the following inequality:

$$\begin{aligned}\varepsilon_i &< \frac{-{\lambda^*}_i^{b} f_i'({\lambda^*}_i) - b{\lambda^*}_i^{b-1} f_i({\lambda^*}_i) - {\lambda^*}_i^{b} f_i'({\lambda^*}_i) + b{\lambda^*}_i^{b-1} f_i({\lambda^*}_i)}{2b{\lambda^*}_i^{b-1} f_i'({\lambda^*}_i)} \\ &= -{\lambda^*}_i/b. \qquad (\text{A}\cdot 4)\end{aligned}$$

By applying the above equation to $f_i(\bar{\lambda}_i)$, we obtain

$$\begin{aligned}f_i(\bar{\lambda}_i) &= f_i({\lambda^*}_i + \varepsilon_i) \\ &= f_i({\lambda^*}_i) + \varepsilon_i f_i'({\lambda^*}_i) \\ &< f_i({\lambda^*}_i) - {\lambda^*}_i f_i'({\lambda^*}_i)/b\end{aligned}$$

which upon substituting Eq. (23) yields

$$\ll 0. \qquad (\text{A}\cdot 5)$$

Because the above inequality shows that the loss event ratio is negative, Eq. (29) does not take a smaller solution if Eq. (23) is satisfied.

Therefore, based on Eqs. (A·1) and (A·3), if Eqs. (20) (22), (23), and (24) are true, the following inequality is obtained and Theorem 2 holds.

$$-{\lambda^*}_i \ll \varepsilon_i \ll {\lambda^*}_i. \qquad (\text{A}\cdot 6)$$