The Sub-Rating Channel Assignment Strategy for PCS Hand-Offs

Yi-Bing Lin, Anthony R. Noerpel, *Senior Member, IEEE,* and Daniel J. Harasty

***Abstract*— A new personal communications services (PCS) hand-off scheme is proposed. This scheme provides for hand-off to radio ports on which there is no free channel by "sub-rating" an existing connection. With sub-rating, an occupied full-rate channel is temporarily divided into two half-rate channels: one to serve the existing call and the other to serve the hand-off request. The blocking probabilities (combined forced terminations of existing calls and blocking of new call attempts) of this new scheme compare favorably with the standard scheme (nonprioritizing) and the previously proposed prioritizing schemes. The "costs" for this scheme are presented and discussed, as well as the additional procedural complexity of implementing on-the-fly sub-rating and the impact of continuing the conversation on a lower rate channel (which may lower speech quality or increase battery drain). Analytical models and simulations investigating the traffic impacts are presented, as are the results that show that even in the highest offered load considered a 3-min conversation in the busy hour experiences less than half a second of sub-rated conversation on average and only about 3% of the calls experience more than 5.12 s of sub-rated conversation. This scheme can increase capacity by 8–35% for systems with 1% call incompletion probability.**

I. INTRODUCTION

PERSONAL communications service (PCS) [1] is expected to provide low-power, high-quality wireless access to the public switched telephone network (PSTN) [2]. The service area of a PCS network is populated with a large number of radio ports with each providing coverage in its vicinity. This paper assumes a fixed or quasi-static channel assignment [3] where a group of channels (time slots, frequencies, spreading codes or a combination of these) are assigned to each port but the results are extensible to dynamic channel assignment schemes [4].

When a user moves from the coverage area of one port to another during a call, a hand-off to the new port is required to maintain the call quality. The forced termination probability (the probability that a hand-off is blocked) is an important criterion in the performance evaluation of the PCS network. The forced termination of an ongoing call is considered less desirable than blocking the initial access of a new call. Conventional PCS systems do not assign priority to hand-off access attempts over initial access attempts. This is called the nonprioritized scheme (NPS). If there are no channels available at a port, then all access attempts are blocked.

Several prioritizing schemes have been proposed and studied to reduce the forced termination probability. In the reserved channel scheme [5], [6], a number of channels in each port are reserved for hand-off calls. This scheme effectively reduces the forced termination probability but may significantly increase the blocking probability of new call attempts for a given number of total servers.

The queueing prioritizing schemes [5], [7]–[9] take advantage of the fact that adjacent ports in a PCS network provide coverage overlay, and there is a considerable area where a call can be handled by either of the ports. This area is called the hand-off area. When a portable with an ongoing call enters a hand-off area, it checks if there is a channel available on the new port. If not, this scheme requires that there be a way for the portable to signal to the new port its desire for a hand-off and the hand-off request is buffered in a waiting queue, and the channel on the old port is used until a new channel is available. The call is forced terminated if no channel is available in the new port by the time the portable moves out of the hand-off area, i.e., out of the range of the old port.

Previous studies [5], [7], [9] indicate that the queueing schemes effectively reduce the forced termination probability at the cost of slightly increasing the blocking probability of new call attempts. However, the total number of call incompletions (i.e., the forced terminated hand-off calls and the blocked new call attempts) for these schemes are roughly the same as NPS. Note that the performance of the queueing schemes depends on the time that a portable stays in the hand-off area and the time that a portable and network will maintain the call connection while the radio link is down or unavailable.

This paper proposes a new channel assignment scheme called the sub-rating scheme (SRS) that creates a new channel on a blocked port for a hand-off access attempt by sub-rating an existing call. Sub-rating means an occupied full-rate channel is temporarily divided into two channels at half the original rate: one to serve the existing call and the other to serve the hand-off request. The protocol required for the portable to request a hand-off at a busy port and the protocol required to sub-rate the traffic channel is described in the Appendix.

Being able to dividing two sub-rated channels from a full rate channel is considered necessary in a viable air-interface structure. For TDMA, channel sub-rating can be easily implemented: the time slots for a full-rate channel are alternatively used by two conversations (as two sub-rating channels). IS-54 [10], GSM [11], PACS [12], DECT [13], and PHP all provide protocol for sub-rating channels in this fashion. For FDMA (such as AMPS) and CDMA systems

---

Manuscript received January 26, 1994; revised August 17, 1995.

Y.-B. Lin is with the Department of Computer Science and Information Engineering, National Chiao-Tung University, Taiwan, R.O.C.

A. Noerpel and D. Harasty are with Bell Communications Research, Red Bank, NJ 07701 USA.

Publisher Item Identifier S 0018-9545(96)00185-5.

(such as IS-95 [14]), it is not as straightforward, though IS-95 may have equivalent schemes.

The performance of SRS is compared to the performance of NPS to determine if giving a priority to hand-off attempts over initial access attempts would virtually eliminate forced terminations without seriously degrading the number of failed initial access attempts. The 32-kb/s ADPCM coding rate is desirable for PCS to ensure good voice quality, low delay, and low portable power consumption. The cost of temporarily switching the call to 16-kb/s can be absorbed in two ways.

The simplest way is to use one ADPCM codec, which can operate in both a 16- or 32-kb/s mode and can be switched between the two. The 16-kb/s ADPCM quality is unattractive, however there is no impact on the channel delay, the current drain or the cost of the portable.

Alternatively, a high-quality 16-kb/s codec could be implemented along with the 32-kb/s ADPCM codec and the conversation can be switched between the two. While the high quality of the voice can be maintained, the new codec will probably have higher delay, add cost to the portable and increase the power drain during the time when it is used. There will also be a "hit" associated with the switch between the two codecs because of the different speech processing times.

For short durations during a call, if a 16-kb/s voice coder were used, then the power consumption might not be impacted significantly or the deteriorated voice quality might be tolerable. Therefore, in addition to studying the blocking probabilities, the time interval during which a call might suffer deteriorated quality is also studied. Section II describes the performance model and the results are discussed in Section III.

A potential problem for SRS is that no matter how small the deteriorated voice quality duration, it will still be bothersome and is forced on an unsuspecting user who is not directly benefiting from the sub-rating of the user's channel. However, all users indirectly benefit and enjoy reduced forced termination probabilities. Another potential problem is that portables that are equipped with the sub-rating capability may be more expensive. We note that other prioritizing schemes [5], [7], [9] also require hardware/software modifications, and the cost may be even higher than our approach. Not all portables need to be equipped with this capability for the system to realize the capacity gains associated with sub-rate hand-off.

II. PERFORMANCE MODELS

This section proposes analytic and simulation models to compare NPS and SRS.

A. The Traffic Model

This paper follows the traffic model we proposed in [7]. The model assumes that the new call attempts to a port form a Poisson process and the portable resides in the coverage area of a port for a period t_m before it moves out of the port's coverage area, where t_m is a random variable with a general distribution F_m with mean $1/\eta$. Let λ_o be the new call arrival rate to a port, μ be the call completion rate (i.e., $1/\mu$ is the mean call holding time), p_o be the blocking of the new call attempts, and p_f be the forced termination probability. Then the rate of hand-off calls λ_f is [17]

$$\lambda_h = \frac{\eta(1-p_0)[1-f^*_{\mathrm{m}}(\mu)]\lambda_o}{\mu[1-(1-p_f)f^*_{\mathrm{m}}(\mu)]}$$

where

$$f^*_{\mathrm{m}}(s) = \int_{t=0}^{\infty} e^{-st} f_{\mathrm{m}}(t)\,dt$$

is the Laplace–Stieltjes Transform for the distribution $F_{\mathrm{m}}(t)$. If we follow Wong's mobility model [15] where $F_m(t)$ is exponentially distributed, then

$$f^*_{\mathrm{m}}(\mu) = \frac{\eta}{\mu+\eta}$$

and

$$\lambda_h = \frac{\eta(1-p_o)}{\mu+\eta p_f}\lambda_o. \tag{1}$$

Let p_{nc} be the probability that a call is not completed by either blocking (as a new call attempt) or forced termination (as a hand-off call). Then [17]

$$p_{nc} = 1-(1-p_o)\left\{1-\left(\frac{\eta}{\mu}\right)\frac{p_f[1-f^*_{\mathrm{m}}(\mu)]}{1-(1-p_f)f^*_{\mathrm{m}}(\mu)}\right\}.$$

If $F_{\mathrm{m}}(t)$ is exponentially distributed, then

$$p_{nc} = 1-\frac{1-p_o}{1+\dfrac{\eta}{\mu}p_f}. \tag{2}$$

If $\mu \gg \eta$, then a portable always completes the call before move, and $\eta p_f/\mu = 0$. Thus, $p_{nc} = p_o$ (i.e., the incomplete calls are blocked new call attempts). If $\eta \gg \mu$, then a call never completes and is eventually blocked (i.e., $p_f = 1$ and $\eta p_f/\mu \to \infty$). Thus, $p_{nc} = 1$. Hong and Rappaport [5] observed that p_{nc} is more sensitive to p_o than p_f. Their observation is not true in general. Equation (2) indicates that as η/μ increases, the impact of p_f on p_{nc} increases, and the impact of p_o decreases.

B. The Nonprioritized Scheme

We have derived p_f, p_o, and p_{nc} for NPS in [7]. The results are summarized below. In NPS, $p_f = p_o$. Suppose that the call holding time t_c is exponentially distributed with the density function $f_c(t_c) = \mu e^{-\mu t_c}$. The channel occupancy time is the minimum of the call holding time (note that the call holding time for a hand-off call has the same distribution as an originating call because of the memoryless property of the exponential distribution) and the remaining portable residual time. In other words, the density function $f_{\mathrm{co}}(t)$ of channel occupancy time distribution is

$$\begin{aligned} f_{\mathrm{co}}(t) &= \int_{t_c=t}^{\infty} f_{\mathrm{c}}(t_c) f_{\mathrm{m}}(t)\,dt_c \\ &\quad + \int_{t_m=t}^{\infty} f_{\mathrm{c}}(t) f_{\mathrm{m}}(t_m)\,dt_m \\ &= (\mu+\eta)e^{-(\mu+\eta)t}. \end{aligned} \tag{3}$$

The net traffic to the system is $\lambda_o + \lambda_h$ where λ_o is the arrival rate for new call attempts and λ_h is the arrival rate for hand-off calls. From the Erlang-B formula

$$p_o = \frac{\dfrac{(\lambda_o + \lambda_h)^c}{(\mu + \eta)^c c!}}{\displaystyle\sum_{i=1}^{c} \frac{(\lambda_o + \lambda_h)^i}{(\mu + \eta)^i i!}} \tag{4}$$

where c is the number of channels in a port. Since $p_o = p_f$ in NPS, (1) is rewritten as

$$\lambda_h = \frac{\eta(1 - p_o)\lambda_o}{\mu + \eta p_o}. \tag{5}$$

The probability p_o can be obtained by an iterative algorithm (similar to the one to be described in the next subsection) using (4) and (5).

C. The SRS Scheme

Let c be the number of (32 kb/s) channels. By assuming that the inter-call arrival times, the call holding times, and the portable residence times are exponential distributed, the sub-rating scheme can be modeled by a Markov process with states $s(i)$, $0 \le i \le 2c$. For $0 \le i \le c$, $s(i)$ represents that there are i 32-kb/s busy channels. For $0 \le i < c$, less than c channels are busy, and the next arrival call (either a new call attempt or a hand-off call) is accommodated to occupy the next free channel. Thus, the Markov process moves from $s(i)$ to $s(i+1)$ with rate $\lambda_o + \lambda_f$. At state $s(i)$ where $0 < i \le c$, the first busy channel is released with rate $i(\mu + \eta)$ because the channel occupancy times are exponentially distributed with rate $(\mu+\eta)$[see (3)]. In other words, the Markov process moves from $s(i)$ to $s(i-1)$ with rate $i(\mu + \eta)$. For $c < i \le 2c$, $s(i)$ represents that there are $2c - i$ 32-kb/s busy channels, and $2(i-c)$ 16-kb/s busy channels. When all 32-kb/s channels are busy, SRS blocks the new call attempts, but accommodates hand-off calls if $c \le i < 2c$. That is, when a hand-off arrives, SRS sub-rates a 32-kb/s channel into two 16-kb/s channels to accommodate the hand-off call. If $i = 2c$, all channels are sub-rated, and new hand-off calls are forced terminated. For $c < i \le 2c$, the Markov process moves from $s(i)$ to $s(i-1)$ with rate $i(\mu + \eta)$ as explained previously. Note that when a 32-kb/s (16 kb/s) busy channel is released, SRS upgrades two (one) 16-kb/s channels to 32 kb/s (and every channel is still occupied). Let π_i be the steady state probability for $s(i)$. Then

$$\pi_i = \begin{cases} \dfrac{\lambda_o + \lambda_h}{i(\mu + \eta)} \pi_{i-1}, & 0 < i \le c \\ \dfrac{\lambda_h}{i(\mu + \eta)} \pi_{i-1}, & c < i \le 2c \end{cases}$$

$$= \begin{cases} \dfrac{(\lambda_o + \lambda_h)^i}{i!(\mu + \eta)^i} \pi_0, & 0 < i \le c \\ \dfrac{(\lambda_o + \lambda_h)^c \lambda_h^{i-c}}{i!(\mu + \eta)^i} \pi_0, & c < i \le 2c \end{cases} \tag{6}$$

and

$$\pi_0 = \left[\sum_{i=0}^{c} \frac{(\lambda_o + \lambda_h)^i}{i!(\mu + \eta)^i} + \sum_{i=c+1}^{2c} \frac{(\lambda_o + \lambda_h)^c \lambda_h^{i-c}}{i!(\mu + \eta)^i} \right]^{-1}. \tag{7}$$

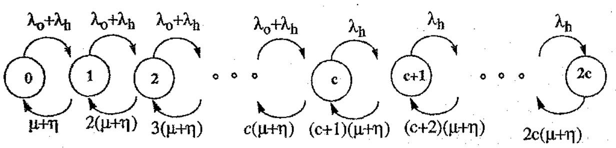

Fig. 1. The state diagram for SRS.

The probabilities p_o and p_f are expressed as

$$p_o = \sum_{c \le i \le 2c} \pi_i$$

and

$$p_f = \pi_{2c}. \tag{8}$$

The expected number of portables in conversation is

$$E[N] = \sum_{0 \le i \le 2c} i\pi_i.$$

The expected number of portables using 16-kb/s channels in conversations is

$$E[N_s] = \sum_{c < i \le 2c} 2(i - c)\pi_i.$$

The degradation of the voice quality D is defined as $D = N_s/N$ and

$$E[D] = E\left[\frac{N_s}{N}\right] = \sum_{c < i \le 2c} \frac{2(i - c)\pi_i}{i}. \tag{9}$$

Therefore, D is the probability that a call is sub-rated. We may also interpret D as follows. During a conversation, the subscriber is expected to experience degraded voice quality for D portion of the call holding time.

The output measures p_{nc}, p_o, p_f and $E[D]$ are obtained by iterating (1), (6)–(8) as follows (note that in the following steps, δ is used to test if the computation converges):

```
**Input:** $c$, $\mu$, $\eta$, and $\lambda_o$.
**Output:** $p_f$, $p_o$, $p_{nc}$, and $E[D]$.
**Step 0.** $\lambda_f \leftarrow 0$, $\delta \leftarrow 1$, $\lambda_h \leftarrow 0$.
**Step1.** If $|\delta| < 0.00001\lambda_h$ go to Step 4.
    Otherwise go to Step 2.
**Step 2.** Compute $\pi_0$ using (7).
    Compute $\pi_i$ using (6). Compute $p_o$ and $p_f$
    using (8).
**Step 3.** Compute new $\lambda_h$ using (1).
    Let $\delta$ be the difference between the old $\lambda_h$ and
    the new $\lambda_h$. Go to Step 1.
**Step 4.** Compute $p_{nc}$ using (2).
    Compute $E[D]$ using (9).
```

For all cases studied in this paper, the above procedure converges.

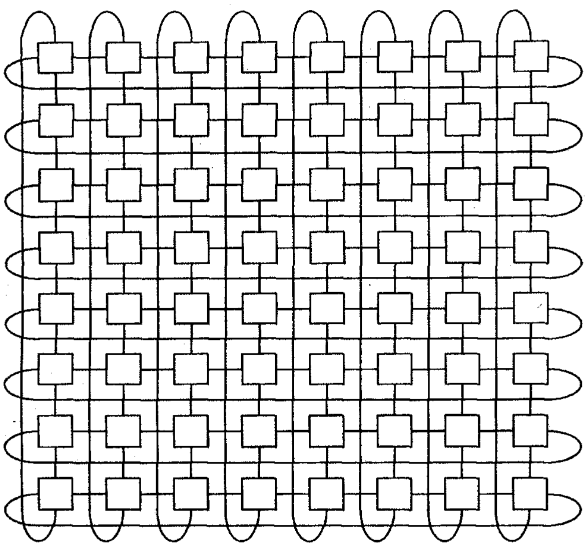

Fig. 2. The PCS network topology considered in the simulation.

D. Simulation

Our analytic models are validated against the simulation experiments. To simulate a very large PCS network, we proposed a wraparound topology for simulation [16]. This approach eliminates the boundary effect occurs in an unwrapped topology. An 8 × 8 wrapped mesh topology is considered in this paper (cf. Fig. 2). The mobility behavior of portables in the simulation is described by a two-dimensional random walk proposed in [17]. In this model, a portable stays in the coverage area of a port for a period of time that has an exponential distribution with mean $1/\eta$. Then, the portable moves to one of the four neighboring port coverage areas with the same routing probabilities 0.25. Initially, every port coverage area has $N = 80$ portables. We assume $c = 10$ channels per port. Two million incoming calls are simulated to ensure that the confidence interval of the 99% confidence level of p_{nc} is less than 3% of the mean value $E[p_{nc}]$. Fig. 3 compares p_{nc} for the analytic results and the simulation results. The curves indicate that the analytic results are consistent with the simulation experiments. Note that the values for the simulation results are smaller than the analytic results. This phenomenon is due to the *busy line* effect [18]: if a new call arrives when a portable is in a conversation, then the new call is not connected. The dropped call is not considered blocked in the simulation. The analytic models for NPS and SRS do not capture the busy line effect. If the number of portables is much larger than the number of channels, the busy line effect disappears and the discrepancy between the analytic model and the simulation model can be ignored.

III. DISCUSSION

This section discusses the performance of SRS.

A. Comparison of SRS and NPS

Fig. 4 plots the p_o curves. For the same mobility η and the offered load ρ, the blocking probability for a new call attempt (i.e., p_o) for NPS is smaller than SRS. This phenomenon is consistent with the previous studies [5], [7], [9] that any hand-off scheme that gives priority to the hand-off calls will increase the blocking probability of the new call attempts. Note that for NPS, p_o decreases as η increases. When η increases, the system experiences larger hand-off arrivals and shorter channel occupancy times where the total offered load (new call attempts plus the hand-off calls) is about the same. A previous study [19] indicates that for the same offered load, the system utilization improves by decreasing the service times, which explains why p_o decreases as η increases in NPS. On the other hand, for SRS, p_o increases as η increases. This phenomenon is due to the fact that λ_f increases as η increases. Since SRS gives priority to the hand-off calls, more new call attempts tend to be blocked as λ_f increases.

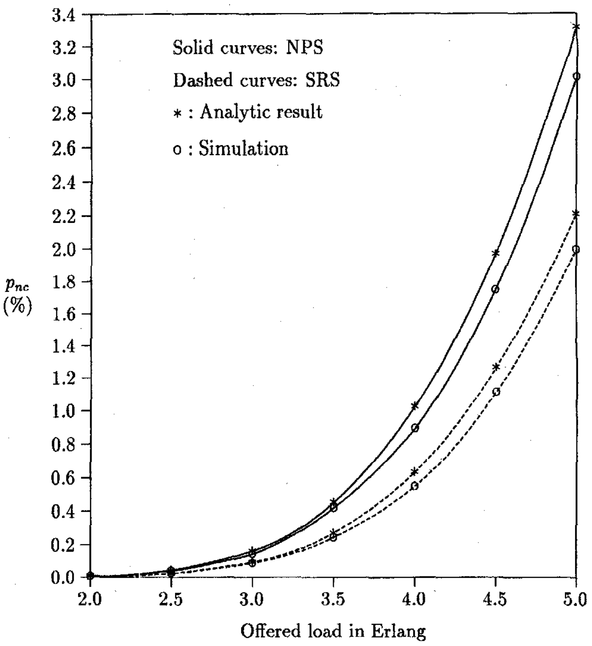

Fig. 3. The comparison of p_{nc} (the probability of an incomplete call) for the analytic results and the simulation results ($\eta = \mu$, $c = 10$).

Table I lists p_f for SRS. Since there are extra c (16 kb/s) channels available to accommodate the hand-off calls, it is not surprising that hand-off calls are almost never forced terminated in the offered load range from 2–5 Erlangs. Fig. 5 plots the p_{nc}. The curves which indicate that for the same η value and offered load, p_{nc} for SRS is smaller than NPS. For both NPS and SRS, p_{nc} increases as η increases. For SRS, $p_{nc} \simeq p_o$ because $p_f \simeq 0$ (see Table I). Note that as the mobility η increases, the benefit of SRS over NPS becomes significant. If we assume that the offered load of the system is engineered at 1% call incompletion probability, then SRS carries 8% more offered load (than NPS) for $\eta = \mu$, and 35% more load for $\eta = 9\mu$.[1]

Fig. 6 plots $E[D]$ for different offered loads. The figure indicates that at high mobility $\eta = 9\mu$ and heavy offered

---

[1] Our study assumes that the offered load is engineered at 1% blocking probability. The extra offered load carried by SRS (compared with NPS) is defined as the engineered offered load of SRS minus the engineered offered load of NPS divides the engineered offered load of NPS.

TABLE I
THE FORCED TERMINATION PROBABILITY p_f FOR SRS

Offered load ρ	2 Erlangs	3 Erlangs	4 Erlangs	5 Erlangs
$p_f(\eta = \mu)$	5.69538×10^{-15}	6.8818×10^{-12}	7.44758×10^{-10}	1.96796×10^{-8}
$p_f(\eta = 3\mu)$	3.28473×10^{-13}	3.95688×10^{-10}	4.21604×10^{-8}	1.06217×10^{-6}
$p_f(\eta = 9\mu)$	2.03382×10^{-12}	2.44519×10^{-09}	2.57518×10^{-7}	6.28067×10^{-6}

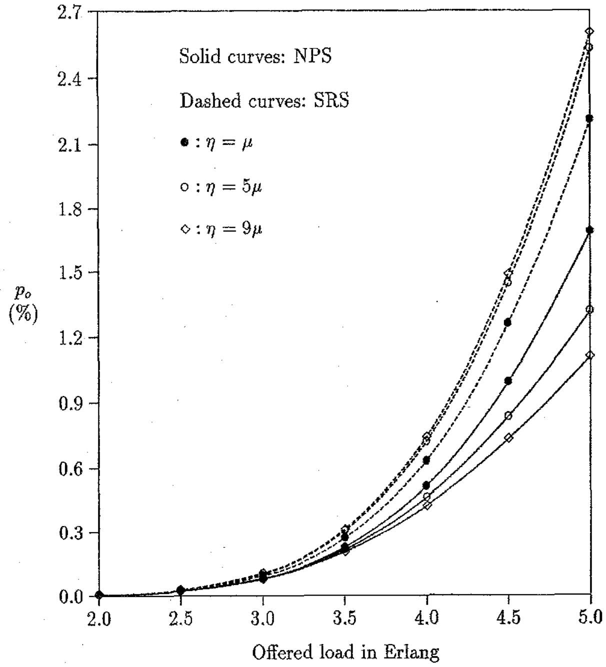

Fig. 4. The comparison of p_o (the probability of a new call attempt being blocked) for NPS and SRS ($c = 10$).

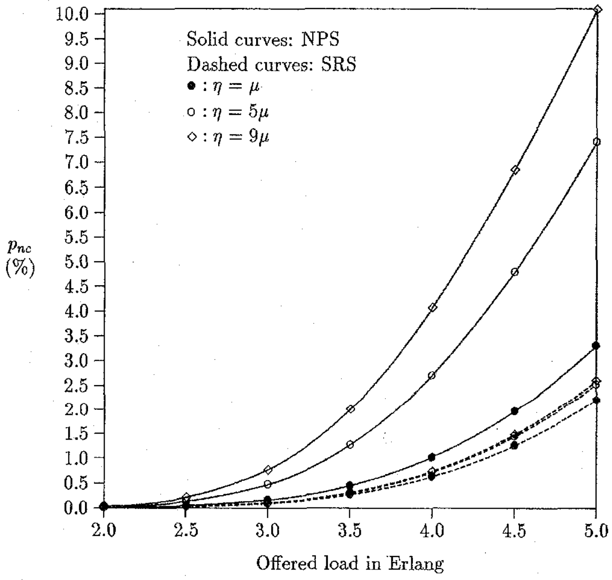

Fig. 5. The comparison of the incompletion probability p_{nc} for NPS and SRS ($c = 10$).

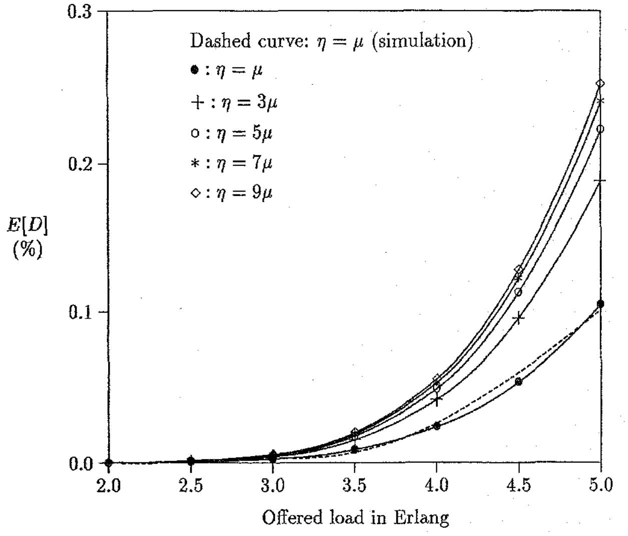

Fig. 6. The effect of the mobility η on $E[D]$, the expected fraction of sub-rated calls ($c = 10$).

load (5 Erlangs), a subscriber only experiences degraded voice quality for a very short period of time (less than 0.25% of the call holding time; i.e., on the average, if the subscriber makes a 3-min call, only 0.45 s are in low voice quality situation). The probability distributions of D are given in Table II. Consider the worst case where $\eta = 9$ and $\rho = 5$. The probability that $\Pr[D \geq 0.18] = 1. -\Pr[D = 0] = 1 - 0.990102 < 1\%$. In other words, less than 1% of calls will experience a low-quality voice period longer than 32.4 s for 3 min mean call holding time. Our simulation experiments indicated that less than 3% of calls will experience a low-quality voice period longer than 5.12 s for 3-min mean call holding time. The low D value is due to the facts that SRS only allows hand-off calls to access the sub-rate channels and that the sub-rate channels are upgraded back to 32-kb/s channels immediately after conversations are completed.

B. Control of Quality Degradation

$E[D]$ is controlled at a low level in SRS for two reasons: 1) we only sub-rate the channels for hand-off calls, not the new call attempts, and 2) when channels are released, the sub-rated channels are upgraded back to full-rate channels immediately. Thus, sub-rating is only a temporarily phenomenon in SRS. Fig. 6 indicates that the SRS mechanism guarantees low $E[D]$ values. However, the variance of the D distribution may not be sufficiently small (for example, the network designer may not be satisfied about "3% of the 3-min calls experience more than 5.12 s of sub-rated conversations." The variance of the D distribution can be controlled by limiting the number c_S of channels available for sub-rating. Our simulation experiments indicate that for $c_S = 1$, "0.5% of the 3-min calls experience more than 5.12 s of sub-rated conversations" for the offered load of 5 Erlangs with $c = 10$ and $\eta = 9\,\mu$. Fig. 7 (this figure

TABLE II
THE PROBABILITY DISTRIBUTION FOR D (THE FRACTION OF CALLS THAT EXPERIENCE SUB-RATING). (a) $\eta = \mu$. (b) $\eta = 9\mu$

D	2 Erlang	3 Erlang	4 Erlang	5 Erlang
0	0.999996	0.999874	0.998877	0.995197
0.18	6.31×10^{-7}	2×10^{-5}	1.71×10^{-4}	6.98×10^{-4}
0.33	9.64×10^{-8}	4.58×10^{-6}	5.18×10^{-5}	2.6×10^{-4}
0.46	1.03×10^{-8}	7.31×10^{-7}	1.09×10^{-5}	6.79×10^{-5}
0.57	9.08×10^{-10}	9.68×10^{-8}	1.92×10^{-6}	1.46×10^{-5}
0.67	7.06×10^{-11}	1.12×10^{-8}	2.98×10^{-7}	2.79×10^{-6}
0.75	4.96×10^{-12}	1.18×10^{-9}	4.16×10^{-8}	4.8×10^{-7}
0.82	3.2×10^{-13}	1.15×10^{-10}	5.34×10^{-9}	7.58×10^{-8}
0.89	1.92×10^{-14}	1.03×10^{-11}	6.37×10^{-10}	1.11×10^{-8}
0.95	1.07×10^{-15}	8.7×10^{-13}	7.1×10^{-11}	1.52×10^{-9}
1	5.69×10^{-17}	6.88×10^{-14}	7.44×10^{-12}	1.96×10^{-10}

(a)

D	2 Erlang	3 Erlang	4 Erlang	5 Erlang
0	0.999993	0.999747	0.997675	0.990102
0.18	1.13×10^{-6}	3.58×10^{-5}	3.00×10^{-4}	1.16×10^{-3}
0.33	3.12×10^{-7}	1.47×10^{-5}	1.64×10^{-4}	7.81×10^{-4}
0.46	5.99×10^{-8}	4.24×10^{-6}	6.24×10^{-5}	3.64×10^{-4}
0.57	9.53×10^{-9}	1.01×10^{-6}	1.97×10^{-5}	1.41×10^{-4}
0.67	1.33×10^{-9}	2.12×10^{-7}	5.48×10^{-6}	4.81×10^{-5}
0.75	1.68×10^{-10}	4.02×10^{-8}	1.37×10^{-6}	1.48×10^{-5}
0.82	1.96×10^{-11}	7.02×10^{-9}	3.17×10^{-7}	4.20×10^{-6}
0.89	2.12×10^{-12}	1.13×10^{-9}	6.81×10^{-8}	1.10×10^{-6}
0.95	2.14×10^{-13}	1.71×10^{-10}	1.36×10^{-8}	2.71×10^{-7}
1	2.03×10^{-14}	2.44×10^{-11}	2.57×10^{-9}	6.28×10^{-8}

(b)

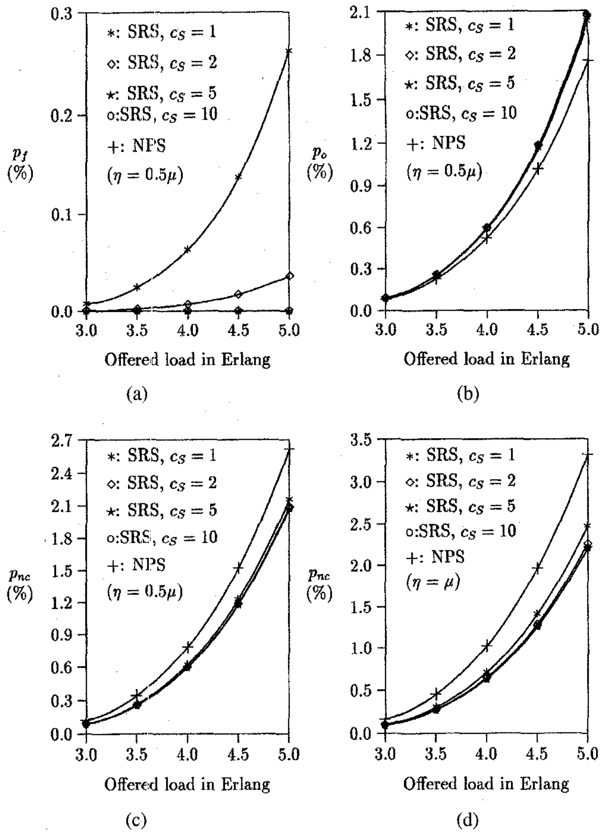

Fig. 7. The performance of SRS with different number of channels for sub-rating. (a) Forced termination. (b) Originating call blocking. (c) Call incompletion. (d) Call incompletion.

was first published in [20]) shows the performance of SRS with different c_S values. The figure indicates that SRS effectively reduces the blocking probabilities (over NPS) with $c_S = 1$. For $c_S > 5$, increasing the number of channels for sub-rating does not improve the performance.

C. Comparison of SRS with Other Priority Schemes

Fig. 8 (this figure is a modification of a figure published in [20]) compares SRS with the reserved channel scheme (RCS), the queueing priority scheme (QPS), and NPS. In [20], we showed that by increasing the number c_R of reserved channels, p_{nc} significantly increases (and is much worse than NPS). Thus, our comparison assumes that c_R in RCS is 1. In QPS, we assume that the queueing policy is first-in-first-out. We also assume that the expected time T_H that a portable stays in the hand-off area is $0.1/\mu$ (i.e., for a call of 3 min, $T_H =$ 18 s). We assume that $c_S = 1$ in SRS. The figure indicates that the blocking probabilities of SRS compare favorably with the other schemes.

Several aspects of the hand-off schemes were discussed in [20]. An unexploited issue regards wireless data communication. For wireless data communication applications such as file transfer, the most important output measure of PCS is p_{nc}. If a file transfer is forced terminated, then the file must be re-transferred. Thus, reducing p_{nc} is important. In this aspect, both SRS and QPS provide good performance (low p_{nc} values). Another aspect of file transfer is that the signal of the radio channels should be strong enough to limit the data re-transmission rate. In a QPS hand-off, either the link is transferred too early (as soon as the user enters the hand-off area) or is transferred too late (when the user is about to leave the hand-off area), and the signal strength of the radio link during hand-off may not be as strong as a normal radio link. This implies that data error rate is high and frequent data re-transmission may be expected. On the other hand, in SRS hand-off, the signal strength of a sub-rated channel is not degraded. The user may experience slow file transfer rate (of the sub-rated channel), but the data re-transmission rate does not increase and the total system resource utilization is better than QPS. Thus, SRS may be a better hand-off scheme in wireless data communication.

IV. CONCLUSION

A new strategy that reduces forced terminations of calls has been presented. By sub-rating existing calls on busy ports to create new sub-rate channels for hand-off access attempts, virtually all forced terminations are eliminated. The penalty has been shown to be a reduction of voice quality during the time that the links are sub-rated to accommodate the hand-off call. If an ADPCM voice coder is used for both the 32- and the 16-kb/s operation, then the impact is degradation in voice quality. However, there will be no increase in portable

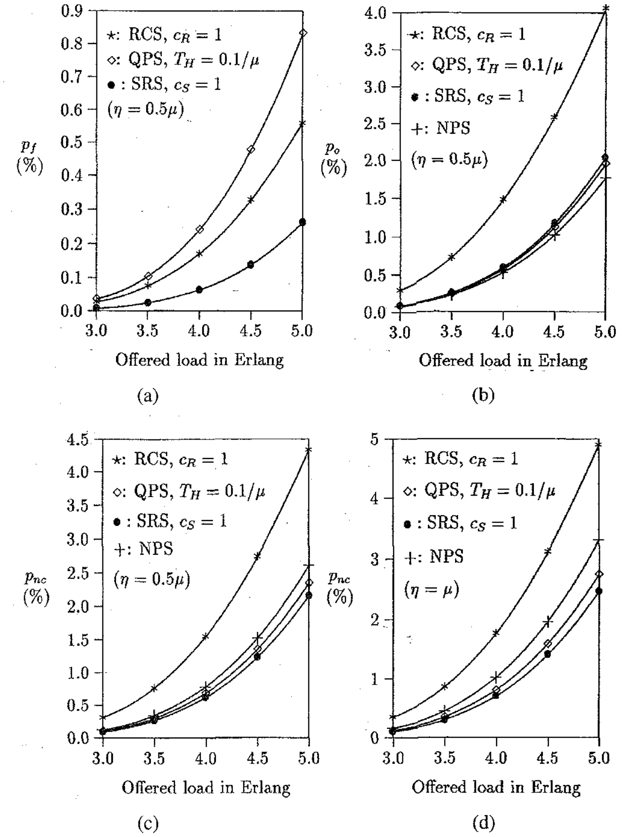

Fig. 8. Comparison of SRS with other priority hand-off schemes. (a) Forced termination. (b) Originating call blocking. (c) Call incompletion. (d) Call incompletion.

power consumption and no increased cost of the portable. On the other hand, if another type of 16-kb/s voice coder were substituted for the 32-kb/s ADPCM coder during the interval when sub-rating was employed, then there may be very little degradation in the intelligibility of the speech but there will be penalties due to portable complexity, power consumption, and delay.

Analytical models and simulations investigating the traffic impacts show that even in the highest offered load considered, i.e., 5 Erlangs per port and 10 servers per port, a 3-min conversation in the busy hours experiences less than half a second of sub-rated conversation on average, and only about 3% of the calls experience more than 5.12 s of sub-rated conversation. Our scheme reduces the forced termination probability more effectively and carries more offered load compared with the previously proposed schemes [5], [6], [9]. This scheme can increase capacity by 8–35% for systems with 1% call incompletion probability.

A potential problem for SRS is that no matter how small the deteriorated voice quality duration, it will still be bothersome and is forced on another user who is not directly benefiting from the sub-rating of the user's channel. However, all users indirectly benefit and enjoy reduced forced termination probabilities. Another potential problem is that portables that are equipped with the sub-rating capability may be more expensive. We note that other prioritizing schemes [5], [8] also require hardware/software modifications, and the cost may be even higher than our approach.

There is another application for the sub-rating protocol in connection with its application to the PACS air interface standard (based on Bellcore WACS [12]). The system signaling channel will typically be a sub-rate channel of 8- or 16-kb/s bandwidth, which leaves a 16- or 24-kb/s channel (normally not available for conventional 32-kb/s voice). This sub-rate channel capacity can be reserved for hand-off access attempts without having to disturb another user's call. This effectively increases the number of servers at a port by one.

APPENDIX

In order for a PCS radio system to take advantage of a priority based scheme for hand-off and initial access channel assignment, there must exist some way for the portable to signal to the blocked port of the need for a traffic channel. For most radio systems, intended for use in the emerging technologies band, portable controlled hand-off is used. Portable controlled means that the request for a hand-off is sent to the new port from the portable. Portable controlled hand-off is often referred to as automatic link transfer (ALT) to distinguish its substantially different characteristics from network controlled hand-off used by analog cellular systems.

In order to implement SRS as well as other prioritizing schemes for systems with ALT, a radio system must have a physical channel—a system signaling channel—for the portable to request the link transfer even when all traffic channels are in use. This channel should always available and cannot be used as a traffic channel. Some PCS radio systems already reserve a channel for other purposes (e.g., system broadcast channel or SBC of PACS), which can be shared by the prioritizing hand-off procedure. For systems with conventional hand-off procedures, the reserved channel is not necessary because the request is made through the network.

The radio air interface also requires a signaling protocol so that the portable can inform the network through the busy port of the access request for the hand-off. Furthermore, there needs to be a mechanism to sub-rate an existing 32-kb/s call in order to free up an additional 16-kb/s channel for the hand-off access request. Note that all of these protocols will take place at layer 2, the link layer. There may be some calls for which it is inappropriate to sub-rate the channel, however this analysis assumes that all calls can equally be sub-rated.

The message flow diagram for the sub-rated ALT procedure is given in steps 1–3 in Fig. 9: 1) the hand-off access request, 2) sub-rating an existing call, and 3) assigning the newly created sub-rate channel to the portable requesting the hand-off. The sub-rate channels are switched back to full-rate channels immediately after some occupied channels are released as shown in steps 4 and 5 in Fig. 9. In this figure, portables are referred to by the more general term of subscriber unit (SU), which includes mobile units for vehicles, portable handsets for pedestrians, and fixed wireless access units for wireless local loop applications.

When the portable, $SU_{hand\text{-}off}$, detects the need for a hand-off, it attempts to seize an available traffic channel.

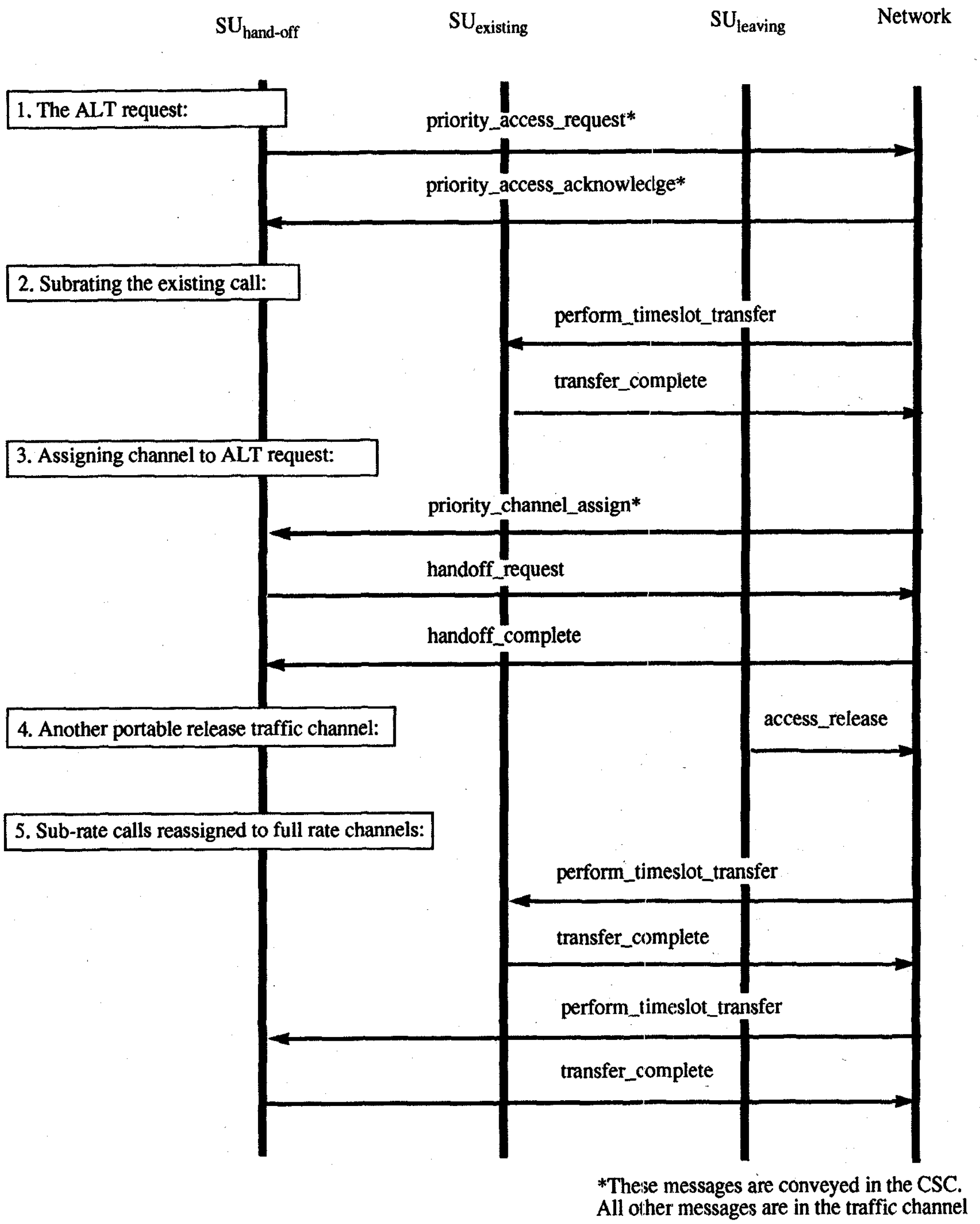

Fig. 9. Message flow diagram for sub-rating automatic link transfer requests.

If no traffic channels are available, then step 1 in Fig. 9 is performed. The portable synchronizes to the SBC and transmits a *priority_access_request* message. The new port responds with either a *priority_channel_assign* message or a *priority_access_acknowledge* message. In the former case, the port has a nonbusy channel that can immediately make available to the portable. In the latter case the port does not have an available channel and is simply acknowledging the receipt of the request message. The portable must continue to monitor the SBC for a *priority_channel_assign* message (in this case if no channel is available within a time-out period, the hand-off call is forced terminated). These messages are given generic pneumonic names but have parallels in many air interface protocols.

At step 2, an existing caller, $SU_{existing}$, is sent a *perform_timeslot_transfer* message commanding it to perform a time slot transfer to a sub-rate channel. This action frees up a sub-rate channel for $SU_{hand\text{-}off}$, the portable requesting the hand-off. This message is acknowledged by the transmission of the *transfer_complete* message.

At step 3, $SU_{hand\text{-}off}$ is informed of the newly made available sub-rate traffic channel via the *priority_channel_assign* message. After receiving it, the portable synchronizes to the available channel and transmits a *hand-off_request* message.

When a user $SU_{leaving}$ terminates an existing call or performs its own hand-off away from the port, its occupied channel is released. The released channel is either a full-rate channel or a sub-rate channel. Without loss of generality, let the channel be full-rate. This channel is not made available for access. Instead, two sub-rate channels are switched back to full-rate channels. At step 4, $SU_{leaving}$ transmit an *access_release* message and releases its channel. At step 5, the released full-rate channel is assigned to either $SU_{existing}$ or $SU_{hand\text{-}off}$ and both of these users now enjoy full rate transmission.

It follows that this access protocol should be generalized to also include emergency access. The message elements of the *priority_access_request* and the *priority_access_acknowledge* messages should include an access random number to resolve collisions and to temporarily identify the portable requesting priority access, the type of priority access, i.e., a hand-off or a 911 call, for example, and a requested channel rate. The portable should identify that it does have a sub-rating capability. The *priority_channel_assign* message requires these message elements in addition to the channel assignment. The channel assignment should include frequency, time slot and sub-rate channel code that identifies the channel which the portable is to use to request the hand-off. In the case of hand-off requests, the access random number can be a radio call identifier, or some number that identifies the call on the old port.

The *perform_timeslot_transfer* message is used to command a portable to switch to a new time slot on the same port. The new time slot could be a sub-rate channel of the currently used full-rate traffic channel. This same message would be sent to return both calls to 32-kb/s time slots once a traffic channel became available.

REFERENCES

[1] D. C. Cox, "Personal communications—A viewpoint," *IEEE Commun. Mag.*, vol. 128, no. 11, pp. 8–20, 1990.

[2] ——, "A radio system proposal for widespread low-power tetherless communications," *IEEE Trans. Commun.*, vol. 39, no. 2, pp. 324–335, Feb. 1991.

[3] J. C.-I. Chuang, "Autonomous adaptive frequency assignment of TDMA portable radio systems," *IEEE Trans. Veh. Technol.*, vol. 40, no. 3, pp. 627–635, Aug. 1991.

[4] ——, "Performance issues and algorithms for dynamic channel assignment," in *IEEE GLOBECOM'92*, Orlando, FL, Dec. 6–9, 1992.

[5] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and no-protection handoff procedure," *IEEE Trans. Veh. Techol.*, vol. VT-35, no. 3, pp. 77–92, Aug. 1986.

[6] C. H. Yoon and K. Un, "Performance of personal portable radio telephone systems with and without guard channels," *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 911–917, Aug. 1993.

[7] Y.-B. Lin, S. Mohan, and A. Noerpel, "Queueing priority channel assignment strategies for handoff and initial access for a PCS network," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 704–712, 1994.

[8] S. Tekinary and B. Jabbari, "Handover policies and channel assignment strategies in mobile cellular networks," *IEEE Commun. Mag.*, vol. 29, no. 11, 1991.

[9] ——, "A measurement based prioritization scheme for handovers in cellular and microcellular networks," *IEEE J. Select. Areas Commun.*, pp. 1343–1350, Oct. 1992.

[10] EIA/TIA, "Cellular system dual-mode mobile station—Base station compatibility standard," Tech. Rep. IS-54, EIA/TIA, 1992.

[11] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications.* Palaiseau, France, 1992.

[12] Bellcore, "Generic criteria for version 0.1 wireless access communications systems (WACS) and supplement," Bellcore, Tech. Rep. TR-INS-001313, Issue 1, 1994.

[13] ETSI, *Digital European Cordless Telephone Common Interface, Version 05.03,* 1991.

[14] EIA/TIA, "Mobile station-base station compatibility standard for dual-mode wideband spread spectrum cellular system," Tech. Rep. IS-95, EIA/TIA, 1993.

[15] W. C. Wong, "Packet reservation multiple access in a metropolitan microcellular radio environment," *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 918–925, 1993.

[16] Y.-B. Lin and V. K. Mak, "Eliminating the boundary effect of a large-scale personal communication service network simulation," *ACM Tran. on Modeling and Computer Simulation,* vol. 4, no. 2, 1994.

[17] Y.-B. Lin, "Reducing location update cost in a PCS network," submitted for publication, 1993.

[18] Y.-B. Lin and W. Chen, "Impact of busy lines and mobility on cell blocking in a PCS network," in *ICCCN '94,* 1994, pp. 343–347.

[19] Y.-B. Lin, Y.-J. Lin, and V. W. K. Mak, "Allocating resources for soft requests—A performance study," *Information Sciences,* vol. 84, nos. 1 and 2, pp. 39–65, 1995.

[20] Y.-B. Lin, S. Mohan, and A. Noerpel, "Channel assignment strategies for hand-off and initial access for a PCS network," *IEEE Personal Commun. Mag.*, vol. 1, no. 3, pp. 47–56, 1994.

[21] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. I.* New York: Wiley, 1966.

Yi-Bing Lin, for a photograph and biography, see this issue, p. 121.

Anthony R. Noerpel (S'76–A'80–M'84–SM'88) received the M.S.E.E. degree from the New Jersey Institute of Technology, Newark, NJ in 1977 and the B.A. degree in mathematics from Rutgers University, New Brunswick, NJ in 1972.

He has worked for Bellcore since its founding in January 1984. He led the effort to develop the higher-layer protocols for the personal access communications system (PACS) radio air interface. He managed, edited, and was the principal author of the generic criteria that detail requirements and signaling protocol for this air interface, which became the basis for the PACS PCS air interface standard document. He is a co-inventor of a time division duplex version of PACS for unlicensed PCS operation called PACS-UB, and he chairs the JTC validation and verification process for this air interface standard. He is currently working on digital data communications protocols for PACS and the interoperability of multiple wireless access systems and services. From June 1977 to December 1983, he worked for Bell Telephone Laboratories. In addition, he has worked on communications satellite systems and has several patents on hybrid mode feed antennas and multimode couplers.

Daniel J. Harasty received the B.S. and M.S. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1987.

He has been at Bellcore since 1988, where he has worked on ISDN and video services and is now in the Radio Research Department. He has been a stalwart participant in T1's and TIA's Joint Technical Committee (JTC) on the PCS Air Interface, where he served as Chair of the Signaling Subcommittee and was a principal architect of the signaling and data protocols for the personal access communications system (PACS) air interface standard. He is continuing his research in wireless data services, focusing on radio and network architectures and protocols that support wireless access. Prior to working at Bellcore, he worked at RCA's David Sarnoff Research Center, where he investigated digital video and audio enhancements for television receivers including graduate research on noncausal recursive filtering to reduce ghosting due to feed-forward multipath channels.