Correspondence

Performance Evaluation of a Push Mechanism for WLAN and Mobile Network Integration

Ming-Feng Chang, Lin-Yi Wu, and Yi-Bing Lin

***Abstract*—In this paper, we propose a universal mobile telecommunications system (UMTS) and a wireless LAN (WLAN) interworking solution called WLAN-based general packet radio service (GPRS) support node (WGSN), which allows a UMTS/WLAN dual mode mobile station (MS) to access heterogeneous wireless services. To reduce the power consumption of an MS, most WGSN applications are not activated at the MS, and MS-terminated services, such as incoming voice over IP (VoIP) calls, are not supported. To address this issue, a push mechanism called session initiation protocol (SIP)-based push center (SPC) was implemented in the WGSN node. For an incoming call to an MS, the SPC utilizes the UMTS short message service to activate the SIP User Agent of the MS. We study the performance of the SPC. An analytic model is proposed to derive the expected number of lost calls during the activation period. The analytic results are validated against the simulation experiments. Our study quantitatively indicates how the SPC performance is affected by the activation time and the timeout period, and we also suggest how to select appropriate values of these two factors to optimize the SPC performance.**

***Index Terms*—Session initiation protocol (SIP), short message service (SMS), universal mobile telecommunications system (UMTS), wireless LAN (WLAN).**

I. Introduction

Integration of mobile systems and wireless LAN (WLAN) systems has become an important trend in telecommunications [1]–[3]. In this approach, users can access Internet services through mobile systems [such as general packet radio service (GPRS) or universal mobile telecommunications system (UMTS)] in the area located outside of WLAN coverage, and otherwise through WLAN systems to obtain higher bandwidth. In [4], we proposed an UMTS and WLAN interworking solution called WLAN-based GPRS support node (WGSN). In the WGSN architecture, the UMTS network (1, Fig. 1) provides 3G packet switched (PS) services, and the WLAN (2, Fig. 1) provides access to the Internet. To support mobile roaming between the UMTS network and the WLANs, a mobile station (MS; see 3 of Fig. 1) must be a 3G-WLAN dual-mode terminal equipped with both a WLAN network interface card (NIC) and a 3G module.

On the network side, a WGSN node acts as a gateway between the packet data network (PDN) and the MSs. To support WGSN mobility management based on the UMTS mechanism, the WGSN node communicates with the home location register (HLR), where the subscriber data and the location information of the WLAN users are stored. In addition, the UMTS SIM-based authentication is reused for WGSN. The WGSN system has been deployed and interworked with the HLRs in MOBITAI communications (a mobile operator in Taiwan).

Manuscript received November 5, 2004; revised January 18, 2005. This work was supported in part by NSC Excellence Project NSC93-2752-E-0090005-PAE, NSC 93-2213-E-009-100, and NTP VoIP Project under Grant NSC 92-2219-E-009-032. IIS/Academia Sinica, and ITRI/NCTU Joint Research Center. The review of this paper was coordinated by Prof. W. Zhuang.

The authors are with the Department of Computer Science and Information Engineering National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C. (e-mail: lywu@csie.nctu.edu.tw).

Digital Object Identifier 10.1109/TVT.2005.861165

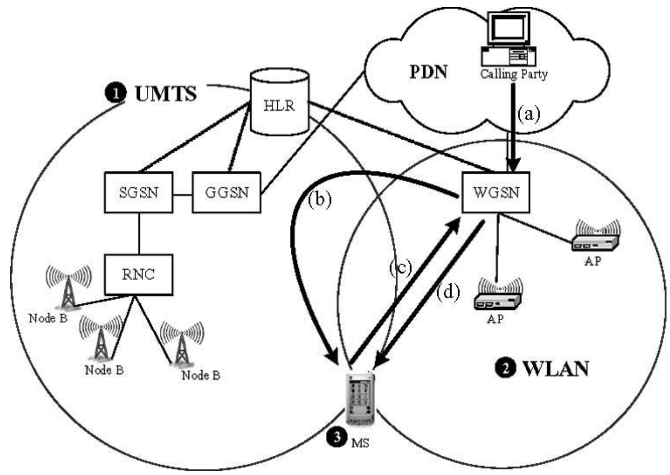

Fig. 1. WGSN Architecture (dashed lines: signaling; solid lines: data and signaling).

To reduce the power consumption of an MS, most WGSN applications are not activated at the MS (i.e., the WLAN module is turned off) until the user actually accesses them. This approach does not support "always-on" or MS-terminated services, such as incoming voice over IP (VoIP) calls [5]. To address this issue, a push mechanism called session initiation protocol (SIP)-based push center (SPC) was implemented in the WGSN node [4]. In this approach, the mobile short message service (SMS) mechanism, which consumes much less power than the WLAN modules, is always on. When a VoIP caller in the external PDN issues a call request to a WGSN MS through SIP, the request is first sent to the WGSN node [path (a) in Fig. 1]. The SPC checks if the SIP user agent (UA) of the called MS is activated. If so, the request is directly forwarded to the called MS [path (d) in Fig. 1]. Otherwise, the request is suspended, and the SPC sends a GSM short message to the MS to activate the corresponding SIP UA [path (b) in Fig. 1]. After the SIP UA is activated, the MS informs the SPC [path (c) in Fig. 1], and the call request from the caller is then delivered to the SIP UA following the standard SIP call setup procedure. The implementation details of the software modules and message flows for SPC can be found in [4]. In this paper, we present accurate analytic models to study the performance of SPC. This analytic model is validated against simulation experiments.

II. WGSN Push Mechanism

Fig. 2 illustrates the timing diagram for the execution of SIP UA activation procedure in the SPC mechanism. This procedure is initiated by the first incoming call arriving at time τ_0 (1, Fig. 2). Suppose that the SPC detects that the SIP UA of the destination MS is not activated. This incoming call is suspended at the SPC. The SPC sends a GSM short message to activate the destination MS (see b in Fig. 1) and sets a timer $T1$ for this call. The incoming call waiting for setup is referred to as the *outstanding call*. If the activation procedure is not complete before $T1$ expires, the call is dropped. In Fig. 2, the timer $T1$ for the first outstanding call expires at time τ_2 (4, Fig. 2), and the SPC receives the activation complete message from the called MS at time τ_6 (9, Fig. 2), where $\tau_6 > \tau_2$. During SIP UA activation, new incoming calls for the destination MS may arrive. If the outstanding call has not been dropped when a new incoming call arrives, then this new incoming

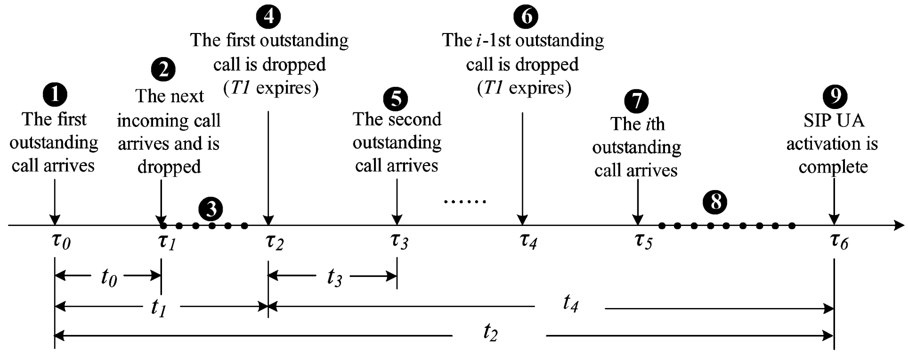

Fig. 2. Timing Diagram for SIP UA Activation (A dot "·." represents dropping of an incoming call immediately after it arrives at the SPC).

call is dropped (2, 3, and 8 in Fig. 2). Otherwise, this incoming call becomes the next outstanding call (see 5 and 7 in Fig. 2).

In this paper, we investigate the performance of the SIP push mechanism, where the expected number of lost calls during SIP UA activation is computed. The lost calls include the dropped outstanding calls due to $T1$ expiration (4, 6, in Fig. 2) and the incoming calls arriving when an outstanding call exists (see 2, 3, and 8 in Fig. 2).

III. Performance Analysis

In this section, we present an analytic model to investigate the performance of push mechanism. We make the following assumptions.

1) The incoming call arrivals are a Poisson process with rate λ; therefore, the inter-call arrival time t_0 is exponentially distributed with the density function $f_{t_0}(t_0) = \lambda e^{-\lambda t_0}$. In Fig. 2, $t_0 = \tau_1 - \tau_0$.
2) The $T1$ timeout period (denoted as t_1) has the density function. In Fig. 2, $t_1 = \tau_2 - \tau_0$. We consider $T1$ with fixed interval $1/\mu$.
3) The SIP UA activation time is denoted as $t_2 = \tau_6 - \tau_0$. In this section, we assume t_2 to be exponentially distributed with the mean $1/\gamma$, and the density function $f_{t_2}(t_2) = \gamma e^{-\gamma t_2}$. We will also consider Gamma distributed t_2 in the simulation model.

In our study, the output measure is the expected number $E[N]$ of the lost calls during the activation period.

Consider the following two cases.

Case 1) The first outstanding call is successfully set up, i.e., the activation time t_2 of SIP UA is shorter than $1/\mu$.

Case 2) The first outstanding call is dropped, i.e., the activation time t_2 of SIP UA is longer than $1/\mu$.

Let P_i be the probability that Case i occurs, and N_i be the expected number of lost calls in Case i. P_i can be expressed as

$$P_1 = \Pr[t_2 < 1/\mu] = 1 - e^{-\left(\frac{\gamma}{\mu}\right)}$$

and

$$P_2 = 1 - P_1 = e^{-\left(\frac{\gamma}{\mu}\right)}. \tag{1}$$

In Case 1, all incoming calls arrive during the SIP UA activation period t_2 are lost. Since the incoming calls are a Poisson stream, the expected number of lost calls during t_2 is λt_2. Therefore

$$P_1 \cdot N_1 = \int_{t_2=0}^{1/\mu} \lambda t_2 \times f_{t_2}(t_2)\, dt_2 = \frac{\lambda}{\gamma} - \left(\frac{\lambda}{\gamma}\right) e^{-\left(\frac{\gamma}{\mu}\right)} - \left(\frac{\lambda}{\mu}\right) e^{-\left(\frac{\gamma}{\mu}\right)}. \tag{2}$$

In Case 2, the first outstanding call and all incoming calls during the waiting time of the first outstanding call are dropped. That is, the expected number of lost calls before $T1$ expires is $1 + \lambda/\mu$. We further analyze Case 2 by two subcases in terms of the next event after $T1$ expires.

Case 2-1) The next event after the $T1$ expiration is the completion of SIP UA activation (i.e., $t_4 < t_3$ in Fig. 2). The expected number of lost calls after the drop of first outstanding call is 0.

Case 2-2) The next event after the $T1$ expiration is an incoming call request (i.e., $t_3 < t_4$ in Fig. 2), and this incoming call becomes the second outstanding call. From the residual life theorem and the memoryless property of the Exponential distribution, t_3 has the same distribution as that for the inter call arrival time t_0. That is, $f_{t_3}(t) = f_{t_0}(t) = \lambda e^{-\lambda t}$. Similarly, we have $f_{t_4}(t) = f_{t_2}(t) = \gamma e^{-\gamma t}$. Since t_3 and t_4 have the same distributions as those for t_0 and t_2, respectively, the situation seen by this outstanding call is the same as that seen by the first outstanding call. Therefore, the expected number of lost calls after the arrival of this new outstanding call is $E[N]$.

Letting P_{2-i} denote the probability that Case 2 $-i$ occurs, we have

$$N_2 = \left(1 + \frac{\lambda}{\mu}\right) + (P_{2-1} \times 0 + P_{2-2} \times E[N]) = \left(1 + \frac{\lambda}{\mu}\right) + \frac{\lambda\, E[N]}{\lambda + \gamma}. \tag{3}$$

From (1)–(3)

$$E[N] = \left[\frac{\lambda}{\gamma} - \left(\frac{\lambda}{\gamma}\right) e^{-\left(\frac{\gamma}{\mu}\right)} - \left(\frac{\lambda}{\mu}\right) e^{-\left(\frac{\gamma}{\mu}\right)}\right] + e^{-\left(\frac{\gamma}{\mu}\right)} \times \left[\left(1 + \frac{\lambda}{\mu}\right) + \frac{\lambda E[N]}{\lambda + \gamma}\right]. \tag{4}$$

By rearranging (4), we have

$$E[N] = \frac{\frac{\lambda}{\gamma} + \left(1 - \frac{\lambda}{\gamma}\right) \times e^{-\left(\frac{\gamma}{\mu}\right)}}{1 - \left(\frac{\lambda}{\lambda+\gamma}\right) \times e^{-\left(\frac{\gamma}{\mu}\right)}}. \tag{5}$$

Equation (5) has been validated against the simulation experiments. The simulation model is described in [6], and the details are omitted. The simulation results show that the discrepancies between the analytic model and simulation are small (less than 0.6% in all cases), and it implies that the derivation of the analytic model is correct.

IV. Numerical Results

Based on the analytic and simulation models in the previous section, we use numerical examples to investigate the SPC performance. Based on (5), Fig. 3 plots $E[N]$ against μ and λ. $E[N]$ is an increasing function of μ and/or λ.

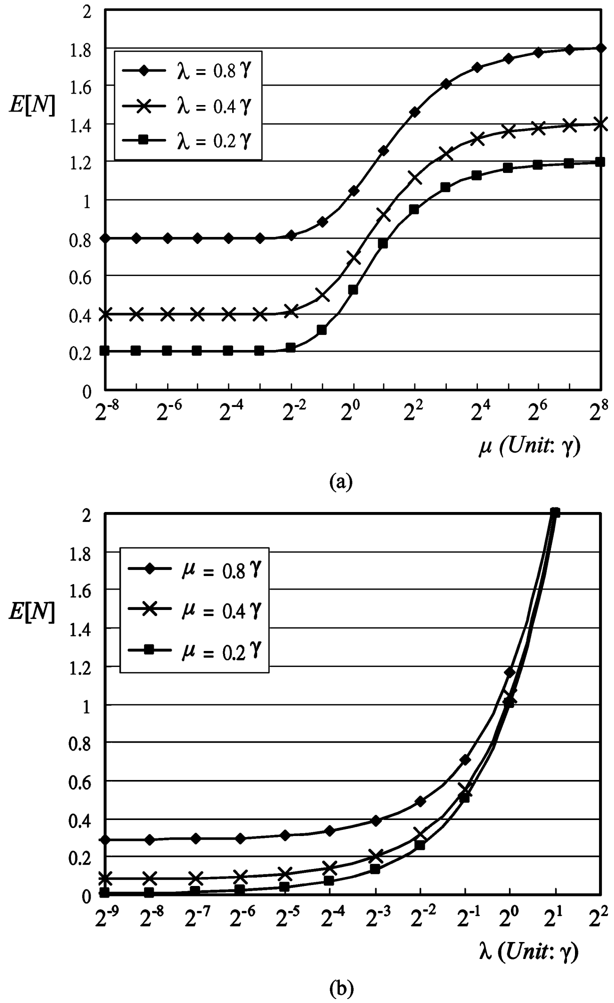

Fig. 3. Effects of μ and λ on the expected number of lost calls. (a) Effect of μ. (b) Effect of λ.

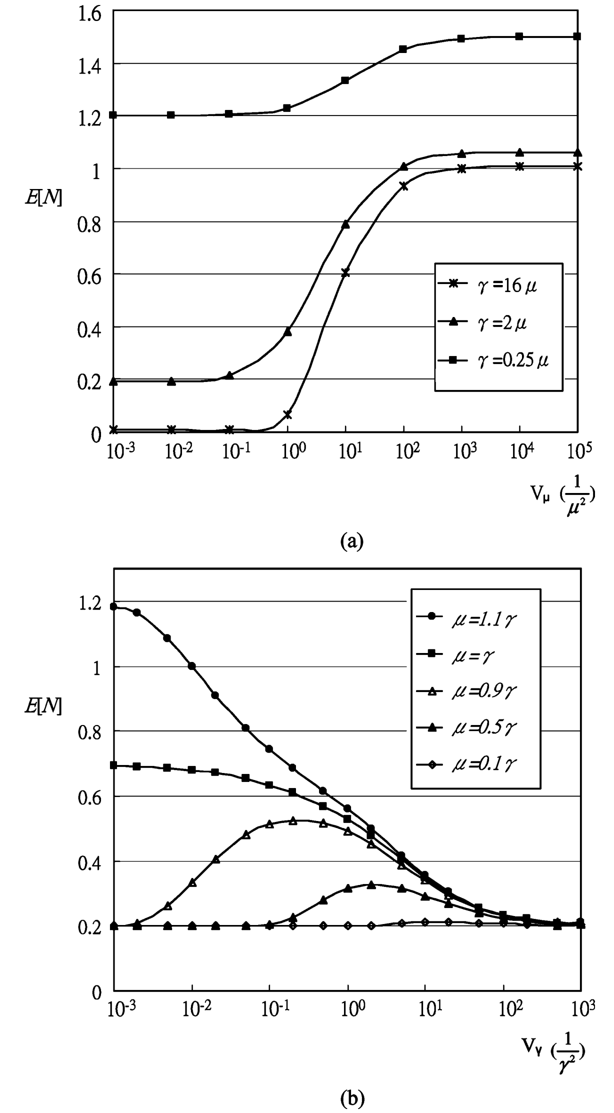

Fig. 4. Effect of Variance V_μ and V_γ. (a) Effect of $V\mu$ (Gamma-distributed T1; $\lambda = \mu/8$). (b) Effect of V_γ (Gamma-distributed activation time; $\lambda = 0.2\gamma$).

Fig. 3(a) indicates that good $E[N]$ performance can be obtained when $T1$ is set as $4/\gamma$. Any timeout period longer than $4/\gamma$ has insignificant improvement on the $E[N]$ performance. Fig. 3(b) indicates that $E[N]$ is insignificantly affected by the incoming call arrival rate λ when $\lambda < \gamma/32$. This result implies that to ensure good $E[N]$ performance, the SIP UA activation mechanism must be designed such that the activation time is shorter than 0.03125 times of the inter-call arrival time.

To extend the observations on the above examples, we use the discrete event simulation model to investigate the SIP UA activation performance with Gamma distributions. It has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions (see [7, Lemma 3.9]). One may also measure time periods in a real mobile network, and the measured data can be approximated by a Gamma distribution as the input to our simulation model. It suffices to use the Gamma distribution with different shape and scale parameters to represent different time distributions.

Suppose that $T1$ has the Gamma distribution with mean $1/\mu$ and variance V_μ. Fig. 4(a) shows the effect of V_μ on $E[N]$, where $\lambda = \mu/8$. The figure indicates that $E[N]$ is an increasing function of V_μ. When V_μ is very large, $E[N]$ is approximate to the $E[N]$ value where $\mu \to \infty$. This phenomenon is explained as follows. When V_μ increases, we will see a significantly increasing number of short $T1$ periods and insignificantly increasing number of long $T1$ periods, and the result is similar to that when $T1$ is set to a small value. On the other hand, when V_μ is small, $T1$ approaches to the fixed value $1/\mu$. Therefore, $E[N]$ is approximate to that when $T1 = 1/\mu$. Fig. 4(a) also shows that $E[N]$ is insignificantly affected by V_μ when $V_\mu < 10^{-1}/\mu^2$ or $V_\mu > 10^3/\mu^2$.

Consider the case when the activation time has the Gamma distribution with mean $1/\gamma$ and variance V_γ, and $T1$ is of fixed length $1/\mu$. Fig. 4(b) shows the effect of V_γ on $E[N]$, where $\lambda = 0.2\gamma$. As V_γ increases, $E[N]$ approaches to λ/γ. When V_γ is small (e.g., $V_\gamma < 10^{-3}/\gamma^2$), the activation time approaches to a fixed length $1/\gamma$. In this case, $E[N]$ is insignificantly affected by V_γ and is determined by the activation time and $T1$. When V_γ is small, the impact of the activation time and $T1$ on $E[N]$ is described as follows.

1) When $\mu < \gamma$ (i.e., $T1$ is very likely to be longer than the activation time), there is a high possibility that the first outstanding call would be connected, and all incoming calls arriving during the activation period are lost. Therefore, $E[N] = \lambda/\gamma$, which is 0.2, as is correctly shown in Fig. 4(b).

2) When $\mu = \gamma$ (i.e., $T1$ is roughly the same as the activation time), the first outstanding call has 50% probability to be dropped. Therefore, $E[N] = 0.5 + \lambda/\gamma = 0.7$ in Fig. 4(b).
3) When $\mu < \gamma$ (i.e., $T1$ is very likely to be shorter than the activation time), the first outstanding call is highly probably dropped, and $E[N] = 1 + \lambda/\gamma = 1.2$ in Fig. 4(b).

When $\mu < \gamma, E[N]$ increases and then decrease as V_γ increases. This phenomenon is explained as follows: When V_γ increases, more long and short activation times are observed. Long activation times result in more lost calls, while short activation times cause fewer lost calls. When V_γ is small (e.g., $V_\gamma < 10^{-1}/\gamma^2$ in the curve for $\mu = 0.9\gamma$), the impact of long activation times is more significant. Therefore, $E[N]$ increases as V_γ increases. As V_γ becomes sufficiently large, the increase of the number of short activation times is more significant than that of long activation times. Hence, $E[N]$ is dominated by the effect of short activation times, and as V_γ increases, $E[N]$ decreases and finally approaches λ/γ. For $\mu > \gamma$, when V_γ increases, the increase of the short activation times is more significant than that of long activation times. Therefore, $E[N]$ is a decreasing function of V_γ.

V. Conclusion

In this paper, we proposed an analytic model to investigate the performance of SIP push center (SPC) [4]. In this mechanism, when a caller sends an SIP request to an MS with its SIP UA is not activated, the SPC sends a GSM short message to activate the SIP UA of the called MS, and a timer $T1$ is set for this activation procedure. If $T1$ expires before the SIP UA is activated, the call is dropped. If a new incoming call arrives before the outstanding call is complete or timed out, the new incoming call is also dropped. In this paper, we focus on measuring lost calls during SIP UA activation. Our study indicates that $T1$ significantly affects the SPC performance. To obtain good $E[N]$ performance, the fixed $T1$ period should be longer than four times of that for SIP UA activation, and SIP UA activation should be completed within 0.03125 times of the intercall arrival time. We note that most callers would not await the call setup for a long time, and they would abort calls before a long $T1$ expires. Therefore, it is suggested that a call should be connected within 10 s; i.e. the SIP UA activation should be completed within 10 s, or the SPC should send a message to notify the caller that the call setup will take longer time and ask the caller to be patient.

The SPC utilizes GSM short message service mechanism. According to [8], it takes 20 s to transmit a short message to an MS. This transmission delay significantly contributes to the SIP UA activation time and may not be acceptable. Therefore, we may further shorten the SIP UA activation time by using a high-priority short messages service. Short messages with high priority would be transmitted to the destination within e.g., 3–5 s. This improvement allows SIP UA activation to be complete within an acceptable delay.

References

[1] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on 3GPP System to Wireless Local Area Network (WLAN) Interworking," Tech. Spec. TR 22.934, v.6.0.0, Sep. 2002

[2] H. Haverinen, J. Mikkonen, and T. Takamaki, "Cellular access control and charging for mobile operator wireless local area networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 52–60, Dec. 2002.

[3] A. K. Salkintzis, C. Fors, and R. Pazhyannur, "WLAN-GPRS integration for next-generation mobile data networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 112–124, Oct. 2002.

[4] W.-S. Feng, L.-Y. Wu, Y.-B. Lin, and W.-E. Chen, "WGSN: WLAN-based GPRS support node with push mechanism," *Comp. J*, vol. 47, no. 4, pp. 405–417, Jul. 2004.

[5] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York: Wiley, 2001.

[6] L.-Y. Wu, M.-F. Chang, and Y.-B. L, "Modeling the push mechanism for WGSN," in *Proc. Internal Workshop on Broadband Wireless Services and Applications 2004*, Oct. 2004.

[7] F. P. Kelly, *Reversibility and Stochastic Networks*. New York: Wiley, 1979.

[8] H.-N. Hung, Y.-B. Lin, M.-K. Lu, and N.-F. Peng, "A statistic approach for deriving the short message transmission delay distributions," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2345–2352, Nov. 2004.

Outage Probability for Maximal Ratio Combining of Arbitrarily Correlated Faded Signals Corrupted by Multiple Rayleigh Interferers

X. W. Cui, Q. T. Zhang, and Z. M. Feng

***Abstract*—Due to the functional complexity of its signal-to-interference (SIR) ratio, the outage probability of maximal ratio combining is available only for certain independent and identically distributed (i.i.d.) fading environments. In this paper, we partially relax this restriction by allowing the signal channel-gain vector to follow a general fading distribution with arbitrary spatial correlation. The problem is tackled in a novel framework. The key step is to represent the probability density of the reciprocal of the SIR, conditioned on the signal vector, as the higher order derivative of a simple exponential function in signal power, whereby a generic formula for outage probability can be determined. The application of the generic formula to Rayleigh, Rician, and Nakagami faded signals is elaborated. Numerical results are also presented for illustration.**

***Index Terms*—Correlated Nakagami fading, correlated Rayleigh fading, correlated Rician fadaing, maximal ratio combining (MRC), outage probability.**

I. Introduction

Diversity reception with adaptive array processing is an effective technique to combat multipath fading and cochannel interference, which are two major sources of performance impairment in many cellular mobile radio systems. The optimal combining (OC) maximizes the signal-to-interference-noise ratio (SINR) at the combiner's output, thereby achieving the best performance in the presence of interference, at the cost of the need to estimate the channel gain vectors for both signal and interferers. Maximal ratio combining (MRC), on other hand, is suboptimal but requires the channel gain vector only for the desired user. Therefore, it is often a more feasible choice in practice.

Outage probability is an important performance measure for a wireless system to operate in a fading environment with cochannel interference. Outage performance of MRC in a multiuser environment has been analyzed in the literature which mainly focuses on some special cases. Though simpler in its structure, the outage analysis of MRC is much more difficult than its OC counterpart. The difficulty arises from

Manuscript received November 8, 2004; revised May 10, 2005. This work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 1248/02E). The review of this paper was coordinated by Prof. G. Saulnier.

X. W. Cui and Z. M. Feng are with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084 China (e-mail: cxw2005@mail.tsinghua.edu.cn; fzm@sdp.ee.tsinghua.edu.cn).

Q. T. Zhang is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: eekzhang@cityu.edu.hk).

Digital Object Identifier 10.1109/TVT.2005.861186