Packet Queueing Delay in Wireless Networks with Multiple Base Stations and Cellular Frequency Reuse

Abstract - Cellular frequency reuse is known to be an efficient method to allow many wireless telephone subscribers to share the same frequency band. However, for wireless data and multi-media communications optimum cell layouts differ essentially from typical solutions for telephone systems. We argue that wireless radio systems for bursty message traffic preferably use the entire bandwidth in each cell. Packet queuing delays are derived for a network with multipath fading channels, shadowing, path loss and discontinuously transmitting base stations.

1 Introduction

Over the past several years, wireless communications has seen an explosive growth in the number of services and types of technologies that have become readily available. Systems for cellular telephony, radio paging and cordless telephony have become commonplace and the demand for enhanced capacity is growing. As the available radio spectrum is limited in bandwidth, network operators attempt to maximize the number of circuits that can be supported simultaneously by reusing radio channels as densely as possibly. It is expected that in the near future, wireless data and multi-media services will also become popular and that there will be an increasing need for spectrum-efficient radio networks to support such services.

Driven by spectrum scarcity, radio resource management for bursty traffic is rapidly becoming a major research topic. The problem of managing information and data flows in radio and wireless optical networks appears to be significantly different from existing techniques for wired or 'guided' communication. Decades of research on sharing communication resources among multiple users and services on wired networks has led to a wide variety of techniques for multiplexing, switching and multiple access to communication resources such as coaxial cable Local Area Networks (LAN's), e.g. [1 - 4]. The common goal of these schemes is the dynamic assignment of *bandwidth* during certain periods of *time*. Many of these multiple-access techniques are also used in radio data networks. However, it soon appeared that the performance of many random-access schemes substantially differs for guided (wired) and unguided (radio) channels, being highly dependent on the physical characteristics of the channel. A review is in [5].

Meanwhile, the aspects of *spatially* reusing scarce radio spectrum resources have mostly been addressed separately from allowing multiple users to share the same *bandwidth* - *time* resources. This is for instance illustrated by the fact that most existing mobile data networks use a cellular frequency reuse pattern, and within each cell a random-access scheme is used independently of the traffic characteristics in other cells. Results on how to dynamically assign the *space-time-bandwidth* resources in radio channels is however not yet well developed.

To future data or multi-media services, queuing and retransmission delays are far more important performance measures than outage probability. Messages lost in a fade or due to interference from other cells can simply be retransmitted, so outage probabilities *per sé* may not be an appropriate criterion. Advanced methods to find link performance, as developed for cellular telephony, can still be used, but need to be extended to address the specific Quality of Service requirements for data or multi-media traffic. This paper finds the packet delay performance in the downlink, i.e., from base station to terminals. We show that it is optimum to use the entire bandwidth in each cell. The corresponding high interference power levels from nearby transmitters in adjacent cells require a joint optimization and dynamic management of the spatial frequency reuse and the occupation of spectrum within cells.

Initially we formulate a model for packet-switched networks. The results in section 2 assume known link performance, as computed e.g. using Laplace transforms.

2 Queuing delay for packet-switched nets

We refer to other documents for a discussion of methods to compute outage probabilities in cellular links. We now apply these to the performance of packet-switched networks. Typically, each base station has a packet queue for messages to be sent to mobile terminals. Packets have a uniform length of L bits, arrive according to a Poisson process, and are destined for terminals uniformly distributed over the cell area. If a packet is received successfully the terminal sends an acknowledgement. If no acknowledgement is received, the base station repeats the packet until successful reception occurs. In a well functioning, stable system, the successful *throughput* of messages over the radio link must be equal to the flow rate of messages into the base station buffer (*arrival rate*). While in our model the arrival rate equals throughput, the *attempted traffic* on the radio channel may be much larger, due to retransmissions. This section computes the attempted traffic for a given arrival rate. We initially assume that all signals experience Rayleigh fading and shadow attenuation, independent of each other and independent from one transmission attempt to the next. The path loss is assumed constant during all retransmission attempts. Interfering signals arrive from base stations in other cells. The area-mean power of these signals is taken from (5). In a hexagonal cell layout, the distance between the centers of two co-channel cells is $R\sqrt{(3C)}$, where R is the cell radius and C is the cluster size, i.e., the number of different frequencies in use. We approximate hexagonal cells of unity size by circular cells of radius $R = 3^{3/4}\sqrt{(2\pi)} \approx 0.91$ [5].

This section derives packet queuing delays in the base station. Since the probability of successful reception $Q(r)$ decreases with increasing propagation distance r, the number of (re-) transmission attempts, M, statistically increases with increasing r. The service time S per packet is $S = M T_L$ with T_L the duration of a time slot. The bit rate per Hz is denoted as η_r, and the total system bandwidth is B_N. So, $T_L = L / (\eta_r B_T) = L C / (\eta_r B_N)$ where B_T denotes the bandwidth per cell with $B_T = B_N/C$. The maximum effective throughput in user bits per second per base station per Hz, is

$$\lambda \triangleq \frac{\eta_r}{C \, \mathrm{E}[M]} \quad (1)$$

The probability of successful reception of a data packet depends on the location of the terminal and on the activity of interfering base stations. For discontinuous transmission, the probability that an interfering base station is active is equal to the probability that its queue is non-empty. In a spatially uniform system, this probability is identical for all base stations. The number of packets waiting in the base station is modelled as an $M/G/1/\infty$ queue. Service times conditional on r are geometric with mean $T_L/Q(r)$ but the unconditional distribution of service times is not a closed-form expression. We approximate the service time of successive packets as i.i.d. random variables. This assumption, however, ignores the fact that the service times of successive packets may be correlated. For instance if one packet experiences heavy interference from transmissions by anther base station, it is more likely that the next packet will also see an active co-channel base station. Furthermore, two adjacent base stations may continue to attempt to transmit to terminals at unfavorable locations. We neglect this effect in the following analysis as a more detailed investigation would require us to compute delays in an infinite field of mutually interacting queues. This problem is notoriously hard to solve. But in section 7 we will propose an access scheme that can resolve these potential instability problems. Moreover, appendix A evaluates the effect of correlated retransmissions.

The mean value of the service time, expressed in number of slots and averaged over all locations in the cell, is

$$\mathrm{E}[M] = \frac{2}{R^2} \int_0^R \frac{r \, dr}{Q(r)} \quad (2)$$

The second moment is

$$\mathrm{E}[M^2] = \frac{1}{\pi R^2} \int_0^R \frac{1 - Q(r)}{Q^2(r)} \, 2\pi r dr \quad (3)$$

The expected queuing delay is found from the Pollacek-Khintchine expression for $M/G/1/\infty$ queues, namely

$$D = \frac{LC}{\eta_r B_N} \left[\mathrm{E}[M] + \frac{\mathrm{E}[M^2]}{\mathrm{E}[M]} \frac{P_{on}}{2(1-P_{on})} \right] \quad (4)$$

3 Results

A relevant performance measure of wireless data networks is the expected number of required transmission attempts versus terminal location. We used a typical modulation technique with receiver threshold z = 4 (6 dB), UHF groundwave propagation with β = 4 and shadowing with a spread of 6 dB (σ_s = 1.36). Service times in seconds can be obtained by multiplying resultsthis by L / (η_r B_N).

Table 2 Maximum throughput S per base station for various cluster sizes C.

C	λ/η_r	Service Time $C\mathrm{E}[M]$	$C\mathrm{E}[M \mid r{=}0.91]$
1	0.40	2.49	5.00
3	0.26	3.87	5.10
4	0.21	4.76	5.88
7	0.13	7.70	8.54
9	0.10	9.63	10.4

Results on the queuing delay in the base station, e.g., as a function of the message arrival rate λ/η_r per cell, shows that C = 1 is optimum. The relatively high bandwidth in each cell allows the base stations to empty their queues fairly rapidly, which reduced interference to other cells.

In contrast to the assumptions in the above analysis, shadow attenuation can be highly correlated for small antenna displacements. In such case shadow attenuation is likely to be almost identical during all retransmission attempts. Moreover, at high traffic loads the busy periods of each co-channel base station are likely to be much longer than the service time experienced by an individual packet. For the limiting case of identical shadow attenuation and identical interference situations during all retransmission attempts, E[M] is derived in Appendix A. Table 3 shows that under heavy traffic, C = 1 does not give favorable performance.

Table 3 Maximum achievable throughput λ/η_r with slow shadowing for various cluster sizes C. Receiver threshold z = 4 (6 dB). UHF groundwave propagation β = 4. Shadowing is 3 and 6 dB.

shadowing	3 dB		6 dB	
C	λ/η_r	P_{ON}	λ/η_r	P_{ON}
1	0.23	0.5	---	---
3	**0.29**	1	0.004	0.07
4	0.23	1	0.012	0.10
7	0.14	1	0.025	0.27
9	0.11	1	0.034	0.42
12	0.08	1	0.045	0.7
13	0.07	1	0.049	0.9
19	0.05	1	**0.050**	1
21	0.04	1	0.046	1

For 3 dB of slow shadowing, C = 3 is found as an optimum cluster size. For more severe shadowing, the network becomes very spectrum-inefficient if one requires that the base stations must continue to perform transmission attempts to terminals in deeply shadowed areas. For 6 dB of slow shadowing, a cluster size as large as C = 19 would be required. In practice, this would be unacceptable and dynamic interference protection may be needed. Also we see that for relatively small cluster sizes, it is more efficient to ensure that each base station is not fully loaded with traffic, i.e., P_{ON} < 1 for optimum throughput. This is contrast to the results for independent shadowing, where the throughput monotonically increases with P_{ON}.

3 Spatial Radio Resource Management

According to the results in the previous sections, in high-capacity spectrum-efficient packet-switched networks, adjacent base stations preferably *compete* for non-disjoint (i.e., interfering) spectrum resources. The above computations showed that spectrum efficiency is optimal if the full system bandwidth can be used in all cells. However, continuing interference between transmissions from adjacent base station may severely affect the performance of such networks. This suggests that efficient, coordinated resolution of collisions between packet retransmissions in adjacent cells is necessary to guarantee efficient performance of wireless data and multi-media networks.

4 Concluding Remarks

The analysis reported here revealed that efficient mobile packet data transmission requires entirely different spectrum reuse than telephone nets. To optimize spectrum efficiency, user capacity and network performance, presumably a new class of access schemes is needed that dynamically combine random access within one cell with protection against interfering signals from other cells. Dynamic Channel Allocation (DCA) is known to provide a means to share bandwidth-time-space resources in a more dynamic and efficient way. However, DCA primarily works on a session by session basis, whereas bursty teletraffic is presumably best supported through access schemes that assign radio resources on a packet by packet basis, or at least on a burst by burst basis.

Under certain assumptions, Contiguous Frequency Assignment (CFA), i.e., cluster size $C = 1$, can support approximately 0.4 bit/s/Hz/cell. This appeared substantially more efficient than cellular frequency reuse with $C = 3, 4, 7, \ldots$. CFA also provides the smallest packet delay at a given spatial packet throughput intensity. This suggests that, in order to ensure minimum delay at maximum user capacity, mobile radio data networks should be designed with much denser frequency reuse than typically used to ensure an outage probability on the order of 10^{-2} or 10^{-3}. This, however, results in low signal-to-interference ratios and large packet loss probabilities, which have to be addressed by appropriate data-link protocols.

In extreme cases, that transmissions to a particular terminal always see the same interfering base stations, so the number of required transmission attempts may become prohibitively large for certain terminals. This could have a detrimental effect on the entire network performance and it has motivated us to start developing new 'spatial' random access protocols.

REFERENCES

[1] L. Kleinrock, "Queuing systems. Vol. 1: Theory; Vol 2: Computer applications", New York: Wiley, 1975, 1976.

[2] D. Bertsekas and R. Gallager, "Data Networks", Prentice-Hall International, Inc., London, 1987.

[3] A.S. Tanenbaum, "Computer Networks", Prentice-Hall International Editions, London, 1989.

[4] J. Walrand, "Communication Networks: a first course", Aksen Associates, Series in Electrical and Computer Engineering, Homewood, IL, 1991.

[5] J.P.M.G. Linnartz, "Narrowband landmobile radio networks", Artech House, Norwood, MA, 1993.

Appendix A: Number of transmission attempts with slow shadowing

In contrast to the assumptions in the main body, we assume identical interference situations during all retransmission attempts of a particular packet. The number of interfering co-channel base stations I is treated as a binomial random variable (I = 0, 1, ···, 6) with mean 6 P_{ON}, where I is constant during the service time of an individual packet. Given the local-mean powers of the wanted and all active interfering signals, the expected number of retransmissions can be expressed as

$$\mathrm{E}\left[M \mid \bar{p}_0, \{\bar{p}_i\}_{i=1}^{I}, I\right] = \prod_{i=1}^{I} \frac{p_0 + z p_i}{\bar{p}_0} \quad (5)$$

If the local-mean powers of signals from interfering base station are i.i.d. log-normally distributed, we find, after solving the I-fold integration,

$$\mathrm{E}\left[M \mid \bar{p}_0, \{\bar{\bar{p}}_i\}_{i=1}^{I}, I\right] = \prod_{i=1}^{I} 1 + \frac{z \bar{\bar{p}}_i e^{\frac{\sigma_s^2}{2}}}{\bar{p}_0} \quad (6)$$

Rewriting the product into a sum of I-terms and using the property that in a log-normal channel

$$\mathrm{E}\left[\bar{p}_0^{\,-n}\right] = \bar{\bar{p}}_0^{\,-n} \exp\left(\frac{\sigma_s^2 n^2}{2}\right) \qquad n = 0, 1, \cdots \quad (7)$$

we find the closed-form expression

$$\mathrm{E}\left[M \mid \{\bar{\bar{p}}_i\}_{i=1}^{I}, I\right] = \sum_{k=0}^{I} \binom{I}{k} \left(\frac{z \bar{\bar{p}}_i}{\bar{\bar{p}}_0}\right)^{k} \exp\left\{(k + k^2) \frac{\sigma_s^2}{2}\right\} \quad (8)$$

which we average over the binomial probability mass function of I. If the position of the receiving mobile is uniform within the cell, the expected number of (re-) transmissions per message is

$$\mathrm{E}M = \sum_{I=0}^{6} \sum_{k=0}^{I} \frac{6!\, P_{ON}^{I} (1 - P_{ON})^{6-I}}{(6-I)!(I-k)!k!} z^{k} (3C)^{-\frac{\beta k}{2}} \exp\left\{(k + k^2) \frac{\sigma_s^2}{2}\right\} \frac{2 R^{\beta k}}{\beta k + 2} \quad (9)$$