Market-Driven Dynamic Spectrum Allocation: Optimal End-User Pricing and Admission Control for CDMA

Virgilio Rodriguez, Klaus Moessner, Rahim Tafazolli
CCSR, The University of Surrey
Guildford, Surrey GU2 7XH, UK
e-mail: vr@ieee.org

***Abstract*— Dynamic spectrum allocation (DSA) seeks to exploit the variations in the loads of various radio-access networks to allocate the spectrum efficiently. Previous work studies a centralised scheme in which a spectrum manager periodically re-allocates spectrum without business considerations. In the present scheme, a spectrum manager performs DSA by periodically selling to network operators short-term spectrum licenses. We target a CDMA-based radio-access technology, and delay-tolerant data applications of various data rates, on the downlink. We solve analytically the problem of the network operator, which must decide simultaneously how much spectrum to purchase, and how to charge its own utility-maximising customers in a way that encourages efficient usage, and maximises the operator's profit. We identify a specific operating point consistent with the interests of both the operator and its customers. With linear spectrum costs, and convenient units of measurement, the operator declines to serve a terminal when a product of known parameters is less than one.**

I. Introduction

Static (fixed) spectrum allocation permanently assigns a segment of the radio frequency spectrum to a radio-access networks (RAN). Static spectrum allocation can be very inefficient, in particular in the presence of highly variable bandwidth demands. Bandwidth demand can vary along the space dimension (from region to region) and along the time dimension (from hour to hour). But with a fixed spectrum allocation, for a given RAN, the region with the largest spectrum peak demand determines the spectrum demand of the entire RAN. Consequently, a substantial fraction of the spectrum may be wasted, at a given time and place.

Dynamic spectrum allocation (DSA) seeks to exploit the variations in the loads of various RAN's to allocate the spectrum efficiently. Reference [1] discusses two DSA schemes that have been previously studied. Important differences between the work reported by [1] and the present work include: (i) our consideration of data-transmitting terminals over a CDMA system, (ii) our explicit consideration of physical layer issues, and (iii) our emphasis on a decentralised solution based on pricing.

In the present work, a "spectrum manager" implements DSA by offering spectrum rights for sale on a very short-term basis. Just before the start of a DSA period, a network operator purchases spectrum, given the current state of its network. But any awarded spectrum licenses expire at the end of a specified short period, at which point the allocation process is repeated.

There are at least two ways in which this scheme could practically arise. First, it is clear that a government agency could become the "spectrum manager" mentioned above. But there is another less obvious way: Spectrum owners in a given locality could create a "spectrum managing firm". They could transfer their spectrum rights to the managing firm, while maintaining ownership of this firm. And they may instruct the managing firm to "resell" the spectrum rights on a short-term basis to the original spectrum owners themselves, (and, possibly, to new communication firms that they may approve). Of course, the managing firm's profits will eventually be distributed among its owners (the original spectrum owners themselves).

The problem of choosing the "right" pricing mechanism for short-term spectrum licenses is an interesting problem in its own right. First, it is not clear what should be the "guiding principle" of the spectrum manager: revenue-maximisation, fairness, overall efficiency, etc. Likewise, if the manager prices the spectrum too low, the operators may demand more spectrum than it is actually available. Conversely, if the price is "too high", much of the spectrum may go unused. Auctions provide an appealing alternative for the spectrum manager, which we explore in a parallel line of work. Herein we assume that the manager has already settled on an appropriate pricing mechanism, and we focus on the problem of the network operator, instead.

The present line of work has much in common with [2]. This reference focuses on a communication resource subject to congestion (an FTP server, a router, etc) and seeks both the optimal level of capacity and the optimal pricing, given some exogenous "cost function" of capacity. Our problem is similar: we assume that the operator of a single CDMA cell populated by data users can purchase "capacity" (spectrum) for short-term use, according to a "cost function" (price) set by the spectrum manager. At the start of a DSA interval, the operator must determine how much spectrum to buy, as well as how to charge its own active end-users. Both problems must be solved jointly. If the end-users are charged a relatively low price, their demand for data services will be relatively high, and so will the operators spectrum needs. If end-users are charged more, they will demand less data services, and the operator will need less spectrum. Ultimately, the operator would like to maximise its profits.

Among relevant works in the literature that have not yet been mentioned, [3] overviews some of the economic tools available to the spectrum manager (such as auctions, economic value analysis, trading, etc), and [4] explores pricing issues in the downlink of a CDMA cell.

II. THE MOTIVES OF OPERATORS AND END-USERS

A. Optimisation problem of a network operator

The main question the operator must answer is how much spectrum to purchase at a given DSA period. At the moment of the purchase decision, the operator will know the number and characteristics of the terminals operating in its network, and the details of the physical communication layer (modulation, error-control coding, mode of diversity, etc). The terminals' "demand" for services will depend on the (internal) pricing policies of the operator. Thus, the operator must determine its own pricing policy along with the amount of spectrum to be purchased from the manager. We neglect the competition among operators. The monopoly analysis provides some useful "bounds": it is the "best case scenario" for the operator, and the "worst case scenario" for the end-user. Additionally, this analysis is a useful approximation of the "oligopoly" situation often observed in practise, in which relatively few operators dominate a given region.

B. Behaviour of the terminals

We must specify the behaviour of a data terminal that can choose resources, in the presence of pricing. We focus strictly on the downlink of a single CDMA cell. Reference [5] provides the basic physical model.

We assume a QoS index similar to one proposed in [6]. It has the form $\beta_i B_i + y_i$ where (i) β_i is the monetary value to the terminal of one information bit successfully transferred (a constant for a given terminal), (ii) B_i is the (average) number of information bits the terminal has successfully transferred within a fixed length of time, say τ, and (iii) y_i is the amount of money the terminal has left after any charges and rewards are computed. This model is grounded on the micro-economic concepts of quasi-linear utility function, and partial-equilibrium analysis [7, Ch. 10].

When quality of service (QoS), x_i, costs $c_i(x_i)$, the terminal chooses x_i to maximise $\beta_i B_i(x_i) + [D_i - c_i(x_i)]$. $\beta_i B_i(x_i)$ is the "value" to the terminal of the bits it gets to transfer over the reference period (the terminal's "benefits"), and D_i is the terminal's monetary budget. D_i is just a constant for a given terminal, which limits its total expenditure. If D_i is relatively "large", it needs not be considered in the analysis. Thus, when QoS is costly, the terminal chooses QoS to maximise benefits minus costs: $\beta_i B_i(x_i) - c_i(x_i)$.

III. PHYSICAL MODEL

In this simple model, the following quantities and concepts are of interest:

1) N is the number of terminals *receiving* data simultaneously *from* a CDMA base station (BS) (downlink operation). The BS has a total downlink power constraint of $\bar{P}$.
2) R_i bps is the data rate of terminal i
3) R_C cps is the chip rate of the channel, common to all terminals. For convenience, we set $R_C = W$, where W is the total bandwidth (spectrum) allocated to the cell.
4) $G_i = W/R_i$ is the spreading (processing) gain of terminal i.
5) Information is sent in M-bit packets carrying $L < M$ information bits.
6) The frame-success rate function (FSF) yields $f_S(x_i)$, the probability of correct reception of a data packet as a function of the signal-to-interference ratio (SIR) at the receiver. Below, $f(x) := f_S(x) - f_S(0)$ replaces $f_S(x)$ to avoid certain technical problems [5]. As an example, for non-coherent FSK modulation, with packet size M=80, independent bit errors, no forward error correction, and perfect error detection, the FSF is $f_S(x) = \left[1 - \frac{1}{2}\exp\left(-\frac{x}{2}\right)\right]^{80}$. However, we stress that our analysis does *not* rely on this or any specific FSF. We assume that *all we know* about the physical layer is that the FSF has the "S" shape shown in figure 1. The

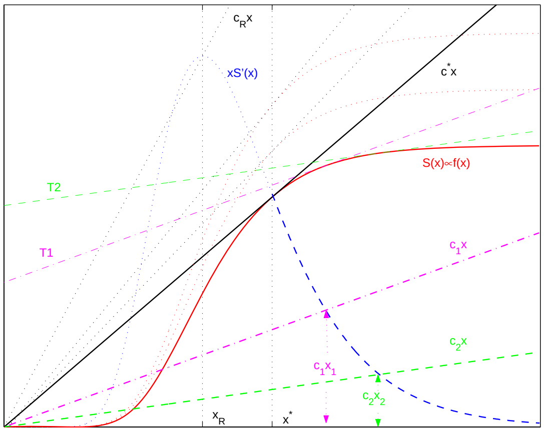

Fig. 1. Pricing for revenue maximisation: With an SIR of x, $S(x)$ represents the terminal's "benefits", or the monetary "value" of the bits it gets to transfer over a reference period. $S(x) \propto f(x)$ (the frame-success rate function (FSF)). When QoS is costly, the terminal maximises benefits minus costs, that is $S(x) - cx$. With $c = c_k$, it chooses $x = x_k$ to satisfy $S'(x_k) = c_k$ (e.g., $T1$, the tangent of S at x_1 is parallel to c_1x), provided that its cost c_kx_k does not exceed its "benefit" $S(x_k)$. The largest c for which the terminal will operate is c^*, the slope of the only tangent of S that goes through the origin. For $c_k \leq c^*$ operator's revenues are $c_kx_k \equiv x_kS'(x_k)$ (blue dash curve). The graph $xS'(x)$ is single-peaked. With the constraint $c \leq c^*$, the curve $xS'(x)$ (revenues) is maximised at $x = x^*$ corresponding to $c = c^*$. x^* does not change when S is replaced by a multiple of S; thus, the same x^* is shared by all terminals with common FSF.

technical characterisation of an "S-curve" and some useful results are given in [8].

7) Following [4] we assume that in the downlink, the CDMA signatures retain their orthogonality, and effectively eliminate intracell interference (or that it is included as part of the random noise). Thus, the received SIR is obtained as $x_i = G_ih_iP_i/\sigma^2$with P_i the downlink power, h_i the path gain, and σ^2 the average noise power at the receiver.
8) Packets received in error which cannot be corrected result in ideal re-transmissions until correctly received and confirmed.

A relatively simple analysis similar to that in [5] tells us that, on the average, the number of information bits successfully transferred by a terminal while operating at constant SIR x_i over the time interval τ is:

$$B_i(x_i) = \tau(L/M)R_i f(x_i) \tag{1}$$

IV. Optimal SIR under linear pricing

In the situation modelled in section III, the operator is assumed to charge the terminal per delivered SIR (QoS) (since the terminal performance depends on both the received power and the available bandwidth). Below, we restrict our attention to linear pricing functions. Per the discussion in section II-B, the terminal chooses its received SIR x_i to maximise benefits minus cost: $\beta_iB_i(x_i) - c_ix_i$, with B_i given by eq. 1. $\beta_iB_i(x_i)$ is just a multiple of the FSF and inherits its shape. Thus, we must investigate how to maximise an expression of the form $S(x) - cx$, where S is some S-curve.

Figure 1 illustrates the solution procedure. First, if the line cx lies entirely above S, except at the origin, the terminal should choose $x = 0$ (decline to operate), since its cost would exceed its benefit for any positive x. Otherwise, the maximising choice is a point at which the derivative of the S-curve equals c. The derivative of the S-curve is "single-peaked" (similar to the curve $xS'(x)$ shown in fig. 1). Therefore, if c is sufficiently large, no value of x can satisfy $S'(x) = c$. Otherwise, two values of x satisfy $S'(x) = c$, and the maximiser is the largest of the two, that is, the one to the right of the inflexion point of S, where the second derivative $S''(x)$ is negative.

The largest value of c for which the problem of maximising $S(x) - cx$ has a positive solution is denoted as c^*, and as shown in fig. 1, is obtained as the slope of the unique tangent line of S that goes through the origin. It is easy to see that replacing S with a multiple of S will change c^*. On the other hand, basic analytical geometry tells us that x^* must satisfy $S(x^*) = x^*S'(x^*)$, which immediately implies that replacing S with a multiple of S has no effect on the value of x^*. Thus, if S_1 and S_2 are multiple of *the same* FSF, they share the same x^* (shown also in fig. 1); that is, x^* is determined by the physical layer, through the FSF. c^* and x^* are related by $c^* = S'(x^*) = S(x^*)/x^*$

In conclusion, for any c in $[0, c^*]$, we can properly speak of a continuous function $x(c)$ that tells us the SIR value that maximises $S(x) - cx$. For $c > c^*$, the maximising choice is zero.

V. Optimal linear pricing

The preceding analysis tells us how the terminal reacts given a linear cost function set by the operator. But it is not totally obvious from the operator's point of view what is the "best" c. In addressing this issue, we shall first assume that a single terminal is active. Subsequently, we will generalise.

A. Only one terminal

As discussed above, and illustrated by fig. 1, for a given $c_k \leq c^*$, the terminal will choose an SIR (QoS level) x_k satisfying $S'(x_k) = c_k$; that is, at x_k the tangent to S is parallel to c_kx (e.g., $T1$, the tangent of S at x_1, is parallel to c_1x). Then, the resulting operator's revenue is $c_kx_k \equiv x_kS'(x_k)$

which has a single "peak" at x_R. In principle, the operator would like to drive the terminal to choose x_R, the point at which the curve $xS'(x)$ reaches its maximum. But this curve crosses S at the point x^*, which lies to the right of x_R, and it has already been established that the terminal will never operate to the left of x^* ($c > c^*$). For any $x > x^*$, $xS'(x) < x^*S'(x^*)$ as shown in fig. 1. Thus, the best the operator can do is to set $c = c^*$, and receive revenues of $c^*x^* \equiv x^*S'(x^*) = S(x^*)$.

The operator is interested in maximising profits, not revenues. It is in principle possible that the revenue-maximising choice may differ from the profit-maximising choice, because of costs considerations. However, by setting its price to c^*, the terminal is being driven to operate at x^*, the lowest SIR which the terminal finds acceptable. The smaller the SIR, the smaller the spectrum needs (for a given power constraint). Thus, by setting a price c^* the operator is both maximising revenues and minimising spectrum costs. This provides the highest achievable level of profit, while serving this terminal.

B. Many terminals

The analysis in the preceding section identifies clearly the revenue-maximising linear price, c^*, and the utility-maximising SIR value, x^*. But the analysis focus on a single terminal, and assumes that the operator knows the terminal utility function (specifically the β coefficient, which denotes the monetary value to the terminal of a correctly transferred bit). When the $\beta's$ are known to the operator, it is straightforward to extend the preceding analysis to a many-terminal situation, provided that the operator can set an individual price per terminal ("price discrimination"). The case in which terminals are non-identical, but the operator is forced to offer the same price to all terminals is more complex. And if the operator does not know the $\beta's$, all cases (even the single-terminal one) become more complicated. Below we shall continue to assume that the operator has full knowledge of the terminal's utility functions, and can set individual prices.

From the analysis summarised in the caption to fig. 1, we know that the operator will choose for terminal i a price c_i^* obtained as the slope of the only tangent to S_i that goes through the origin. ($S_i(x_i) = \beta_i B_i(x_i)$ with B_i given by equation 1).

From the discussion in section IV, we know that if the terminals share an identical FSF, f, then each S_i is a multiple of the common f, and the terminals will choose an identical SIR x^*(that is, the operator will choose c_i^* such that each terminal's best response is to choose $x_i = x^*$).

VI. Spectrum and Admission Considerations

The operator must allocate the available downlink power among all served terminals, and obtain the required amount of spectrum. For a given bandwidth, W, the allocated powers must satisfy:

$$\frac{W}{R_i}\frac{h_i P_i}{\sigma^2} = x^* \Rightarrow P_i = \frac{R_i}{W}\frac{\sigma^2}{h_i}x^* \qquad (2)$$

The power constraint requires that

$$\sum_{i=1}^{N} P_i = \bar{P} \Rightarrow W^* = \frac{x^*}{\bar{P}/\sigma^2}\sum_{i=1}^{N}\frac{R_i}{h_i} \qquad (3)$$

With $W_0 := x^*\sigma^2/\bar{P}$ we can say that serving terminal i requires a specific amount of spectrum:

$$w_i^* = W_0 R_i/h_i \qquad (4)$$

Let us suppose that the operator can purchase bandwidth W for κW. Terminal i should be served (under optimal pricing) only if its contribution to revenue exceeds the (spectrum) cost of serving it.

Under optimal pricing, terminal i would pay $c_i^*x^* = x^*S_i'(x^*) = S_i(x^*)$, which, applying eq. 1, can be written as

$$S_i(x^*) = \tau(L/M)f(x^*)\beta_i R_i \qquad (5)$$

By dividing this revenue by the amount of spectrum that terminal i requires, we obtain its contribution to revenue per unit of required bandwidth:

$$\rho_i := \tau\frac{\bar{P}}{\sigma^2}\frac{L}{M}\frac{f(x^*)}{x^*}\beta_i h_i \qquad (6)$$

In order for a terminal to be served, ρ_i should be no less than the unit cost of spectrum; that is, $\rho_i \geq \kappa$.

Evidently, we can choose a monetary unit such that $\kappa = 1$, and, for a given link layer configuration, we can set a time scale such that $\tau(\bar{P}/\sigma^2)(L/M)f(x^*)/x^* = 1$. Then, terminal i should be served only if

$$\beta_i h_i \geq 1 \qquad (7)$$

If the cost of spectrum is not linear, the admission control formula is not as neat, but the procedure is only slightly more complicated.

VII. "Optimal" link layer

Equation 6 yields ρ_i, the contribution to revenue of terminal i per unit of required bandwidth. The product of ratios $(L/M)f(x^*)/x^*$ is determined by the link layer configuration (e.g., modulation and coding). Other things being equal, the configuration providing the highest $(L/M)f(x^*)/x^*$ maximises "revenue per Hertz", when spectrum costs are linear.

VIII. Discussion

Dynamic spectrum allocation (DSA) exploits the temporal and/or regional variations in the loads of various radio access networks to allocate the spectrum efficiently. Previous work considered voice-only UMTS traffic, and studied a centralised scheme, in which a benevolent manager matches spectrum allocation to system load. We have proposed a decentralised scheme, in which a "spectrum manager" implements DSA by periodically selling short-term spectrum licenses. We have described a realistic business model which could implement our scheme. We have assumed that the spectrum manager sells spectrum at a unit price (which is presumably set to make demand equal supply). Such arrangement is plausible, for instance, when the state wants to allocate the spectrum reasonably efficiently without a significant concern for revenue, and when there is a relatively large number of spectrum buyers, none with enough power to influence the "market clearing price". We explore elsewhere the use of auctions as an allocation mechanism of short-term spectrum licenses.

In our physical model, delay-tolerant terminals operate at dissimilar data rates in the downlink of a CDMA cell. We provide analytical results whose core is summarised in the caption to fig. 1. The interests of the operator (profit maximisation) and the terminals (maximising utility, which equals benefit minus cost) meet at a specific operating point: the SIR value x^*. This number can be easily identified by drawing a tangent line from the origin to the graph of the frame-success rate function (FSF). The slope of this tangent leads to the price at which the operator must sell QoS (SIR) to a terminal. Our pricing results could be useful with or without DSA or CDMA. Even under optimal pricing, the operator may decline to serve a terminal, because what the terminal pays may be less than the cost of the additional spectrum it requires. When the operator's spectrum costs are linear, (and with the monetary unit and time scale conveniently chosen), the admission decision takes the simple form: serve terminal i only if $\beta_i h_i \geq 1$, with h_i the terminal's channel gain, and β_i its "willingness to pay". We do not impose a specific FSF, but assume that it is some S-curve, a very mild assumption. Thus, our analysis should apply to a wide variety of physical layer configurations, and in fact provides a rationale for a network operator to adjust the link layer for profit maximisation.

We have not discussed the additional functionality needed by a wireless network and its terminals in order to implement DSA. A relevant discussion is found in [1]. In particular, we would like the network to adjust its chip rate to match the spectrum allocation. Evidently, current networks and standards do not support DSA. But with the steady advance of technology, the additional functionality seems within reach. Before any adoption decision, the cost of the upgrade should be compared to the benefits of the scheme. When the demand for services varies widely over time and/or space, the performance gains of any DSA scheme are magnified, but are minimised under uniform demand. By considering a UMTS and a DVB-T operator participating in a DSA scheme, [1] reports gains approaching 40%. Those gains will, in the near future, be compared to those arising from our scheme, after we complete an analysis similar to the present one for a DVB-T operator.

We have focused on a "small island" geography, in which inter-cell interference plays no role, because it can be covered with a single cell by the participating radio-access networks, and have only considered a CDMA downlink in which intra-cell interference can be neglected. These limitations will be addressed in future reports of this work.

Acknowledgement

Performed in the framework of the EU funded project E^2R. We acknowledge the contributions of our E^2R partners.

References

[1] P. Leaves, K. Moessner, R. Tafazolli, D. Grandblaise, D. Bourse, R. Tonjes, and M. Breveglieri, "Dynamic spectrum allocation in composite reconfigurable wireless networks," *IEEE Comm. Magazine*, vol. 42, pp. 72–81, May 2004.

[2] J. K. MacKie-Mason and H. R. Varian, "Pricing congestible network resources," *IEEE J. on Selected Areas in Comm.*, vol. 13, pp. 1141–9, Sep 1995.

[3] W. Webb, "The role of economic techniques in spectrum management," *IEEE Comm. Magazine*, vol. 36, pp. 102–107, Mar. 1998.

[4] P. Liu, P. Zhang, S. Jordan, and M. L. Honig, "Single-cell forward link power allocation using pricing in wireless networks," *IEEE Trans. on Wireless Comm.*, vol. 3, pp. 533–43, Mar 2004.

[5] V. Rodriguez, "Robust modeling and analysis for wireless data resource management," *IEEE WCNC*, vol. 2, pp. 717–722, 2003.

[6] V. Rodriguez, "Efficient decentralized power control via a compensation mechanism," *IEEE/Sarnoff Symposium*, pp. 189– 92, Apr. 2004.

[7] H. R. Varian, *Microeconomic Analysis*. New York: W.W. Norton & Co., 3rd ed., 1992.

[8] V. Rodriguez, "An analytical foundation for resource management in wireless communications," *IEEE Globecom*, vol. 2, pp. 898–902, December 2003.