Performance Bounds for Dynamic Channel Assignment Schemes Operating under Varying Re-Use Constraints

Phil Whiting, Sem Borst
Bell Laboratories, Lucent Technologies
Murray Hill, NJ 07974-0636
{pawhiting, sem}@research.bell-labs.com

Abstract — We derive bounds for the performance of dynamic channel assignment (DCA) schemes which strengthen the existing Erlang bound. The construction of the bounds is based on a reward paradigm as an intuitively appealing way of characterizing the achievable carried traffic region. In one-dimensional networks, our bounds closely approach the performance of Maximum Packing (MP), which is an idealized DCA scheme. This suggests not only that the bounds are extremely tight, but also that no DCA scheme, however sophisticated, can be expected to outperform MP in any significant manner, if at all.

Our bounds extend to scenarios with varying re-use which may arise in the case of dynamic re-use partitioning or measurement-based DCA schemes. In these cases, the bounds slightly diverge from the performance of MP, which inflicts higher blocking on outer calls than inner calls, but not to the extent required to maximize carried traffic. This reflects the trade-off that arises in the case of varying re-use between efficiency and fairness. Asymptotic analysis confirms that schemes which minimize blocking intrinsically favor inner calls over outer calls, whereas schemes which do not discriminate among calls inevitably produce higher network-average blocking.

1 Introduction

Most cellular networks, such as AMPS, GSM, and TDMA IS-136, are operated according to a frequency re-use plan. The radio frequencies are divided evenly into (say) N re-use groups, which are then arranged into a regular pattern, see Lee [7]. The re-use factor N is usually determined by considering the Carrier-to-Interference Ratio (CIR) which mobiles under worst-case conditions would experience. This approach to network operation is called Fixed Channel Assignment (FCA).

Besides ensuring that the CIR is adequate, network operators are also concerned about the probability that a call is lost because all the channels are in use. Under standard assumptions of Poisson traffic, no mobility, no mobile-to-mobile calls, and blocked calls cleared, the loss probability is given by the Erlang-B formula

$$\mathrm{Erl}(\nu; C) = \frac{\nu^C}{C!} \Bigg/ \sum_{j=0}^{C} \frac{\nu^j}{j!}, \qquad (1)$$

where ν is the offered traffic and C is the number of channels per cell.

In recent years, the use of cellular services has expanded dramatically, forcing consideration of more efficient ways of using the radio frequencies. One approach to increase capacity is to allow channels to be allocated in a more flexible fashion, but still adhering to the re-use constraints. Channels may thus be taken away from cells which are being offered fewer calls, and diverted to cells which are being offered more calls. This approach to network operation is called Dynamic Channel Assignment (DCA) [7].

The flexibility of DCA schemes is an important feature, since the pattern of offered traffic cannot be exactly determined in advance, and is typically time-varying. Besides offering a potential capacity increase, DCA schemes also reduce the complexity of frequency planning, which is of particular significance in the massive deployment of micro-cells.

An important example of a DCA scheme is Maximum Packing (MP), which was introduced by Everitt and MacFadyen [1]. Under MP, a call is admitted whenever possible, even if this involves rearranging the channels assigned to calls already in progress. An exact analysis of MP on a doubly-infinite strip, in which a channel cannot be used simultaneously in two adjacent cells, is presented in Kelly [3]. The results show that even for uniform offered traffic, MP outperforms FCA, unless the load exceeds a certain critical value. Jordan & Khan [2] and Kind *et al.* [6] have reported a similar observation.

In fact, the above example is one of the rare instances in which DCA schemes allow for an exact analysis. Even so, the blocking-minimizing scheme is not known nor is it known how far FCA and MP are removed from the optimum policy. This situation motivates the construction of simple bounds for the blocking to provide insight into the performance of DCA schemes. An example is the Erlang bound, which was first derived in Whiting [13], and later studied in Raymond [10]. The Erlang bound provides a lower limit on the network-average blocking under any DCA scheme, and may be obtained as the solution to a linear program.

In the present paper, we derive bounds which substantially strenghten the Erlang bound. They may be obtained by careful selection of a reward vector w, with each call carried in cell i generating a reward w_i. Clearly, no DCA scheme can produce a greater revenue than the optimum policy with respect to this reward vector. This observation functions as the basis for our bounds.

0-7803-4383-2/98/$10.00

Now let us return to the issue of increasing the network capacity. A further extension to adopting DCA is made by modifying the frequency re-use plan or even dropping it altogether. This allows tighter re-use than would be permitted under the worst-case conditions of the plan mentioned earlier. One approach is to adopt an underlay-overlay network, where each cell now includes an inner region in which lower powers are used. This allows the channels allocated to the calls in the inner region to be re-used more frequently, while channels assigned to the outer calls continue to operate at the original re-use factor.

A second approach is to assign channels based on interference measurements, and to exploit those measurements to obtain tighter re-use than under the frequency plan. The questions of when to admit a call and of how to operate the network now need to be reasserted. In particular, the issue how the CIR is to be held at an adequate level needs to be addressed by the algorithm itself or through supplementary control mechanisms.

In these circumstances, the Erlang bound no longer directly applies, and indeed the blocking may be significantly lower than under more conventional forms of DCA. We will show how our revenue-based bounds extend to these scenarios with tightened re-use. (Xu & Akansu [14] and Zander & Eriksson [15] derive asymptotic lower and upper bounds.) Furthermore, we are once again in a position to compare our bounds with exact results for a version of MP which incorporates tightened re-use.

In summary, the paper is organized as follows. In Section 2, we present a more detailed model description, and briefly review the derivation of the Erlang bound. We then provide some basic examples illustrating how the Erlang bound may be calculated. Subsequently, we examine the achievable carried traffic region to understand why the Erlang bound may not always be tight. Section 3 introduces a reward paradigm which paves the way for the construction of sharper bounds. In Section 4, we revisit the examples studied in Section 2 to illustrate how the revenue-based bounds may be used to improve upon the Erlang bound. In Section 5, we investigate the trade-off between efficiency and fairness that arises in the case of varying re-use. Finally, Section 6 summarizes our main conclusions.

2 The Erlang bound

We first present a more detailed model description. We consider a cellular network of arbitrary topology. The cells, which are indexed by the set $\mathcal{I}$, share a pool of C channels. Users in cell i generate calls as a Poisson process of rate ν_i. All calls have exponentially distributed holding times with unit mean.

When a user generates a call, the admission policy determines whether to accept or reject it. If accepted, the call is carried for the complete duration of the holding time. In case a call is rejected, the user does not make any retrials.

We assume that the admissible states of the network satisfy the constraints $\sum_{i \in \mathcal{C}} n_i \leq C$ for all $\mathcal{C} \in \Omega$, with n_i denoting the number of calls in cell i. The set Ω is the collection of *cliques*, which are defined as the subsets $\mathcal{C}$ of $\mathcal{I}$ such that no two users within $\mathcal{C}$ can share a channel.

As shown in Whiting [13], the Erlang bound provides a lower limit on the network-average blocking under any admission scheme. It may be obtained as the solution to the following linear program

$$\min \quad \overline{B} \equiv \sum_{i \in \mathcal{I}} \nu_i B_i \Big/ \sum_{i \in \mathcal{I}} \nu_i \tag{2}$$

$$\text{sub} \quad \sum_{i \in \mathcal{C}} \nu_i B_i \geq \sum_{i \in \mathcal{C}} \nu_i \mathrm{Erl}\left(\sum_{i \in \mathcal{C}} \nu_i; C\right) \quad \text{for all } \mathcal{C} \in \Omega \tag{3}$$

$$0 \leq B_i \leq 1 \qquad \text{for all } i \in \mathcal{I}, \tag{4}$$

with the variables B_i representing the probability of call blocking in cell i under some arbitrary admission policy.

The key constraints are provided by the inequalities (3), which are obtained by considering each clique $\mathcal{C} \in \Omega$ in isolation. Since no two users within a clique can share a channel, we cannot accommodate more than C calls in any one clique simultaneously. Thus, we can never reject fewer calls in a clique $\mathcal{C} \in \Omega$ than the number of blocked call for a single group of C channels offered traffic $\sum_{i \in \mathcal{C}} \nu_i$. This number is determined by the Erlang-B formula (1).

We now provide some basic examples illustrating how the Erlang bound may be calculated.

Example 3.1: three-cell network
Consider the three-cell network depicted in Figure 1. Each cell is offered traffic at rate ν. We assume that a channel cannot be used simultaneously in two adjacent cells, i.e., the cliques are $\{1, 2\}$ and $\{2, 3\}$, labeled A and B in the figure, respectively. Thus, the clique constraints are $B_1 + B_2 \geq 2\mathrm{Erl}(2\nu; C)$ and $B_2 + B_3 \geq 2\mathrm{Erl}(2\nu; C)$. The solution to the linear program is $B_1 = B_3 = 0$, $B_2 = 2\mathrm{Erl}(2\nu; C)$, yielding the bound $\overline{B} = 2\mathrm{Erl}(2\nu; C)/3$. For $C = 2$, $\nu = 1$ for example, we obtain $\overline{B} = 0.266\cdots$.

The bound may be tightened by adding the single-cell clique constraints $B_i \geq \mathrm{Erl}(\nu; C)$, $i = 1, 2, 3$. The solution to the linear program is then $B_1 = B_3 = \mathrm{Erl}(\nu; C)$, $B_2 = 2\mathrm{Erl}(2\nu; C) - \mathrm{Erl}(\nu; C)$, sharpening the bound to $\overline{B} = (2\mathrm{Erl}(2\nu; C) + \mathrm{Erl}(\nu; C))/3$. For $C = 2$, $\nu = 1$, we obtain $\overline{B} = 0.333\cdots$. Using Markov decision theory, we find that the minimum achievable blocking in fact is $\overline{B} \approx 0.411$.

□

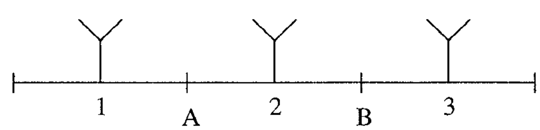

Figure 1: Three-cell network.

Example 3.2: doubly-infinite strip
Consider a similar network with uniform offered traffic as in the previous example, but now a doubly-infinite strip of cells, instead of just three, as shown in Figure 2 below. The clique constraints are $B_i + B_{i+1} \geq 2\mathrm{Erl}(2\nu; C)$ for all $i \in \mathcal{I}$. Using

an elementary limiting argument, it may be concluded that the linear program yields the bound $\overline{B} = \text{Erl}(2\nu; C)$. Adding the single-cell clique constraints does not strengthen the bound in this case.

□

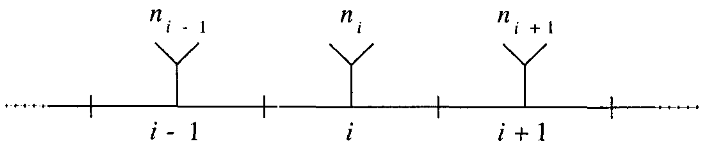

Figure 2: Linear array of cells.

Example 3.3: doubly-infinite strip with varying re-use
Consider a similar linear array of cells as in the previous example, but now a scenario with varying re-use, as illustrated in Figure 3 below. We assume that each cell i is partitioned into an inner region $(i, 1)$ and an outer region $(i, 2)$. Each of the inner regions and each of the outer regions is offered traffic at rate $\nu_1 = \alpha\nu$ and $\nu_2 = (1 - \alpha)\nu$, respectively. Calls in two different inner regions may always share a channel. Calls in outer regions cannot share a channel with any call in the two adjacent cells. The Erlang bound no longer applies at the level of cells now, but does still apply at the level of the regions. The clique constraints are $\alpha B_{i,1} + (1 - \alpha)(B_{i,2} + B_{i+1,2}) \geq (2 - \alpha)\text{Erl}((2 - \alpha)\nu; C)$ and $\alpha B_{i+1,1} + (1 - \alpha)(B_{i,2} + B_{i+1,2}) \geq (2 - \alpha)\text{Erl}((2 - \alpha)\nu; C)$ for all $i \in \mathcal{I}$. Using a simple limiting argument, it may be shown that the linear program produces the bound $\overline{B} = (2 - \alpha)\text{Erl}((2 - \alpha)\nu; C)/2$. Not surprisingly, the bound is decreasing in α, the fraction of traffic offered to the inner regions. Adding the constraints $B_{i,1} \geq \text{Erl}(\alpha\nu; C)$, the bound may be tightened to $\overline{B} = ((2 - \alpha)\text{Erl}((2 - \alpha)\nu; C) + \alpha\text{Erl}(\alpha\nu; C))/2$.

□

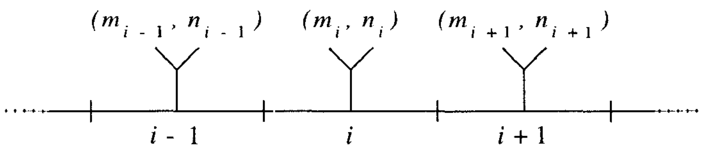

Figure 3: Linear array of cells with varying re-use.

Discussion
The Erlang bound as exemplified above may not always be tight. To understand why, we now examine the region of achievable carried traffic combinations. The clique constraints (3) underlying the Erlang bound may be rewritten

$$\sum_{i \in C} \lambda_i \leq \sum_{i \in C} \nu_i \left(1 - \text{Erl}\left(\sum_{i \in C} \nu_i; C\right)\right) \quad \text{for all } \mathcal{C} \in \Omega, \quad (5)$$

with the variables $\lambda_i = \nu_i(1 - B_i)$ representing the carried traffic in cell i under some arbitrary admission policy. Now let us return to Example 3.1. The *outer* region in Figure 4 delineates the set of all carried traffic pairs (λ_1, λ_2) that satisfy the constraints (5) for clique A. However, the *true* achievable carried traffic pairs for clique A, are demarcated by the *inner* region in the figure.

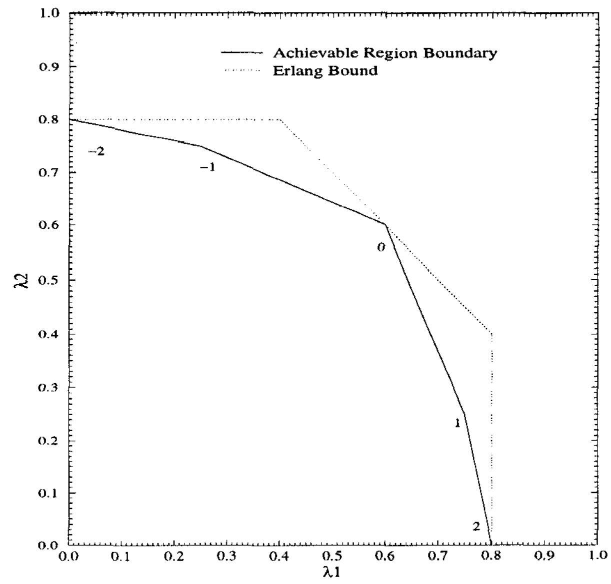

Figure 4: Achievable carried traffic region for a single group of $C = 2$ channels offered two streams of traffic of rate $\nu = 1$ each.

The piece-wise linear boundary of the achievable region may be interpreted as follows. Consider a reward vector (w_1, w_2), with w_i representing the reward generated by each stream-i call that is carried. The reward-maximizing policy is then a *trunk reservation* strategy, see Lippman [8], Miller [9]. Under trunk reservation, the calls of the lower-earning stream are rejected when there are no more than r free channels.

The carried traffic pairs for the class of trunk reservation strategies are represented by the vertices of the *inner* region in Figure 4. They are labeled with the value of the corresponding trunk reservation parameter, taken negative when used against stream-1 calls. No carried traffic pair outside the inner region is achievable, since otherwise the optimality of the class of trunk reservation strategies would be contradicted.

The Erlang bound in Example 3.1 followed from the solution $(B_1, B_2) = (0.2, 0.6)$ to the linear program. Figure 4, however, shows that the corresponding carried traffic pair $(\lambda_1, \lambda_2) = (0.8, 0.4)$ is infeasible. Thus, the Erlang bound may be strengthened if we replace the clique constraints (3) by the linear inequalities describing the boundary segments of the achievable region. This insight will be formalized in the next section.

Note that a different picture would emerge if call dropping were permitted. If pre-emption were allowed, then the achievable carried traffic pairs are exactly the vertices of the outer region in Figure 4. Thus, the Erlang bound may not be tight because it fails to exclude carried traffic combinations which are only feasible if call dropping were permitted. Allowing for pre-emption, however, appears inappropriate as call dropping should be negligibly small for any sensible admission control scheme.

3 Tighter bounds

We now proceed with a formal statement of the proposed bounds. As we have seen in the previous discussion, we may use a reward paradigm as an insightful way of characterizing the achievable carried traffic region, and thus sharpening the Erlang bound. Specifically, suppose that each call carried in cell i generates a reward w_i. For any vector $w \in \mathcal{R}_+^{\mathcal{I}}$, denote by $R(w)$ the maximum achievable mean reward rate. Clearly, no admission policy can produce a greater revenue rate than $R(w)$. This observation constitutes the basis for the next theorem.

Theorem 4.1
For any set $\mathcal{W} \subseteq \mathcal{R}_+^{\mathcal{I}}$, the carried traffic under any admission policy is bounded above by the optimum value $\lambda_{\mathcal{W}}$ of the following linear program

$$\max \quad \sum_{i \in \mathcal{I}} x_i \tag{6}$$

$$\text{sub} \quad \sum_{i \in \mathcal{I}} w_i x_i \leq R(w) \qquad \text{for all } w \in \mathcal{W}, \tag{7}$$

$$x_i \geq 0 \qquad \text{for all } i \in \mathcal{I}. \tag{8}$$

Proof
The proof follows by interpreting the variables x_i as the carried traffic in cell i under some arbitrary admission policy. The objective function (6) then exactly represents the carried traffic. Constraint (7) is satisfied since the policy cannot produce a greater revenue than the maximum achievable reward rate. Hence, the optimum value of the linear program provides an upper bound for the carried traffic under any admission policy.

□

Corollary 4.2
For any set $\mathcal{W} \subseteq \mathcal{R}_+^{\mathcal{I}}$, the carried traffic under any admission policy is bounded above by the optimum value $\mu_{\mathcal{W}}$ of the following linear program

$$\min \quad \sum_{w \in \mathcal{W}} y(w) R(w) \tag{9}$$

$$\text{sub} \quad \sum_{w \in \mathcal{W}} y(w) w_i \geq 1 \qquad \text{for all } i \in \mathcal{I}, \tag{10}$$

$$y(w) \geq 0 \qquad \text{for all } w \in \mathcal{W}. \tag{11}$$

Proof
The proof follows by observing that (9)-(11) is the dual problem to (6)-(8). Strong duality then implies that $\lambda_{\mathcal{W}} = \mu_{\mathcal{W}}$.

□

The main difficulty in evaluating the above bounds does usually not arise from solving the linear programs, but from computing the $R(w)$'s for a suitable set $\mathcal{W}$. Typically, determining $R(w)$ requires numerically solving a Markov decision problem with a state space in as many dimensions as the reward vector w has non-zero components. In certain cases, however, $R(w)$ may be obtained in closed form. For any clique $\mathcal{C} \in \Omega$ for example, $R(\chi^{\mathcal{C}}) = \sum_{i \in \mathcal{C}} \nu_i (1 - \text{Erl}(\sum_{i \in \mathcal{C}} \nu_i; C))$, with $\chi^{\mathcal{C}}$ denoting the characteristic vector of $\mathcal{C}$. From $\lambda_i = \nu_i(1 - B_i)$, we then immediately see that the inequalities $\sum_{i \in \mathcal{I}} \chi_i^{\mathcal{C}} \lambda_i \leq R(\chi^{\mathcal{C}})$ are equivalent to the clique constraints (3). Thus, for the set $\mathcal{W} := \bigcup_{\mathcal{C} \in \Omega} \{\chi^{\mathcal{C}}\}$, the above bounds coincide with the Erlang bound.

At the opposite side of the spectrum, $R(1, \ldots, 1)$ *equals* the maximum achievable carried traffic, but it is exactly the formidable complexity of calculating this quantity directly which motivated us to consider bounds. This contrast is characteristic of the trade-off between the computational complexity of determining the $R(w)$'s and the tightness of the corresponding bounds.

For any subset $\mathcal{J} \subseteq \mathcal{I}$, denote $\mathcal{R}_+^{\mathcal{J}} := \{w \in \mathcal{R}_+^{\mathcal{I}} : w_i = 0 \text{ for all } i \notin \mathcal{J}\}$. Now suppose that Π is the collection of subsets $\mathcal{D} \subseteq \mathcal{I}$ such that $R(w)$ can be obtained if $w \in \mathcal{R}_+^{\mathcal{D}}$. Define $\mathcal{W}^{\Pi} := \bigcup_{\mathcal{D} \in \Pi} \mathcal{R}_+^{\mathcal{D}}$ as the set of all reward vectors w for which $R(w)$ can be obtained. In case $\Pi \subseteq \Omega$, the collection of cliques in the network, we know that for any $w \in \mathcal{W}^{\Pi}$ the maximum reward rate $R(w)$ is achieved by some trunk reservation strategy. Occasionally, we will therefore refer to the corresponding bounds as 'trunk reservation bounds'.

Note that we cannot determine $\lambda_{\mathcal{W}^{\Pi}}$ by solving either of the above two linear programs directly, since there are an infinite number of inequalities (variables in the dual version) involved. From linear programming theory, however, we know that at most a finite number of these are relevant. We now describe two approaches to obtain $\lambda_{\mathcal{W}^{\Pi}}$ exploiting that fact.

In the first approach, we generate a finite yet exhaustive subset including all the relevant inequalities. For any subset $\mathcal{J} \subseteq \mathcal{I}$, denote $\mathcal{U}^{\mathcal{J}} := \{x \in \mathcal{R}_+^{\mathcal{J}} : \sum_{i \in \mathcal{I}} w_i x_i \leq R(w) \text{ for all } w \in \mathcal{R}_+^{\mathcal{J}}\}$. By definition, $\lambda_{\mathcal{W}^{\Pi}}$ may be obtained by maximizing $\sum_{i \in \mathcal{I}} x_i$ subject to the constraints $(x_i)_{i \in \mathcal{J}} \in \mathcal{U}^{\mathcal{D}}$ for all $\mathcal{D} \in \Pi$. Also, define $\mathcal{A}^{\mathcal{J}}$ as the convex hull of the carried traffic combinations in the subnetwork of the cells $i \in \mathcal{J}$ achievable by the class of stationary deterministic admission policies. Observe that the convex hull is a polytope, since there are only finitely many stationary deterministic admission policies.

Lemma 4.3
For any subset $\mathcal{J} \subseteq \mathcal{I}$,

$$\mathcal{A}^{\mathcal{J}} = \mathcal{U}^{\mathcal{J}}.$$

Proof
The inclusion to the right is implied by the definition of $R(w)$. The inclusion to the left holds by virtue of the optimality of the class of stationary deterministic admission policies.

□

The above lemma implies that $\lambda_{\mathcal{W}^{\Pi}}$ may be obtained by maximizing $\sum_{i \in \mathcal{I}} x_i$ subject to the constraints $(x_i)_{i \in \mathcal{J}} \in \mathcal{A}^{\mathcal{D}}$ for all $\mathcal{D} \in \Pi$. Thus, it suffices to generate the set of facet-defining inequalities of the polytopes $\mathcal{A}^{\mathcal{D}}$ for all $\mathcal{D} \in \Pi$.

In the second approach, we identify the subset of relevant inequalities more indirectly. In the dual formulation, it is quite natural to interchange the roles of the coefficients w and the variables $y(w)$. For example, fixing $y(w) = 1$ for all $w \in \mathcal{W}^*$, we find that $\sum_{w \in \mathcal{W}^*} R(w) \geq \mu_{\mathcal{W}^*} \geq \mu_{\mathcal{W}^\Pi}$ for any subset $\mathcal{W}^* \subseteq \mathcal{W}^\Pi$ with the property that $\sum_{w \in \mathcal{W}^*} w_i \geq 1$ for all $i \in \mathcal{I}$. The next theorem establishes that this in fact holds with equality for subsets $\mathcal{W}^*$ of remarkably small size.

Theorem 4.4
For any set Π, the optimum value $\lambda_{\mathcal{W}^\Pi} = \mu_{\mathcal{W}^\Pi}$ equals the optimum value V^Π of the following convex programming problem

$$\min \sum_{\mathcal{D} \in \Pi} R(w^\mathcal{D}) \tag{12}$$

$$\text{sub} \sum_{\mathcal{D} \in \Pi} w_i^\mathcal{D} = 1 \qquad \text{for all } i \in \mathcal{I}, \tag{13}$$

$$w^\mathcal{D} \in \mathcal{R}_+^\mathcal{D} \qquad \text{for all } \mathcal{D} \in \Pi. \tag{14}$$

Proof
We first prove that $V^\Pi \geq \mu_{\mathcal{W}^\Pi}$. Let $\{v^\mathcal{D}\}_{\mathcal{D} \in \Pi}$ be the optimal solution to the problem (12)-(14), so $V^\Pi = \sum_{\mathcal{D} \in \Pi} R(v^\mathcal{D})$. The statement preceding the theorem then indicates that $\sum_{\mathcal{D} \in \Pi} R(v^\mathcal{D}) \geq \mu_{\mathcal{W}^\Pi}$.

We now prove that $\mu_{\mathcal{W}^\Pi} \geq V^\Pi$. Let $\{z(w)\}_{w \in \mathcal{W}^\Pi}$ be the optimal solution to the dual problem (9)-(11), so $\mu_{\mathcal{W}^\Pi} = \sum_{w \in \mathcal{W}^\Pi} z(w) R(w)$. From optimality, we may conclude that the $z(w)$'s satisfy the constraints (10) with strict equality, since $R(\cdot)$ is an increasing function.

Let $z^\mathcal{D}(w) \geq 0$ be variables such that $\sum_{\mathcal{D} \in \Pi} z^\mathcal{D}(w) = z(w)$ for all $w \in \mathcal{W}$ and $z^\mathcal{D}(w) = 0$ if $w \notin \mathcal{R}_+^\mathcal{D}$. Now define $v^\mathcal{D} := \sum_{w \in \mathcal{W}^\Pi} z^\mathcal{D}(w) w$ for all $\mathcal{D} \in \Pi$. It is easily verified that $\{v^\mathcal{D}\}_{\mathcal{D} \in \Pi}$ satisfies the constraints (13)-(14). Plugging the $v^\mathcal{D}$'s into the objective function (12) then gives $\sum_{\mathcal{D} \in \Pi} R(v^\mathcal{D}) \geq V^\Pi$.

It remains to be shown that $\sum_{w \in \mathcal{W}^\Pi} z(w) R(w) \geq \sum_{\mathcal{D} \in \Pi} R(v^\mathcal{D})$. Note that $\sum_{w \in \mathcal{W}^\Pi} z(w) R(w) = \sum_{w \in \mathcal{W}^\Pi} \sum_{\mathcal{D} \in \Pi} z^\mathcal{D}(w) R(w) = \sum_{\mathcal{D} \in \Pi} \sum_{w \in \mathcal{W}^\Pi} z^\mathcal{D}(w) R(w)$. Now define $\zeta^\mathcal{D} := \sum_{w \in \mathcal{W}^\Pi} z^\mathcal{D}(w)$ for all $\mathcal{D} \in \Pi$. Using the fact that $R(\cdot)$ is a convex function, and that $R(\eta w) = \eta R(w)$ for any scalar $\eta \geq 0$, we obtain $\sum_{w \in \mathcal{W}^\Pi} z^\mathcal{D}(w) R(w) = \zeta^\mathcal{D} \sum_w \frac{z^\mathcal{D}(w)}{\zeta^\mathcal{D}} R(w) \geq$

$$\zeta^\mathcal{D} R\left(\frac{\sum_w z^\mathcal{D}(w) w}{\zeta^\mathcal{D}}\right) = \zeta^\mathcal{D} R\left(\frac{v^\mathcal{D}}{\zeta^\mathcal{D}}\right) = R(v^\mathcal{D}).$$

□

We now consider two special cases.

Doubly-infinite strip
Consider the doubly-inifinite strip studied in Example 3.2. Define $\lambda_{\max}$ as the maximum average amount of carried traffic in each cell. Suppose that Π^K is the set of all the subnetworks of K consecutive cells. An elementary limiting argument shows that the bound of Theorem 4.4 applies. Exploiting the symmetry and the convexity of the function $R(\cdot)$, we conclude that $\lambda_{\max}$ is bounded above by the optimum value of the following convex programming problem

$$\begin{aligned} \min \quad & R(w_1, \ldots, w_K) \\ \text{sub} \quad & \sum_{k=1}^K w_k = 1 \\ & w_1 = w_K, w_2 = w_{K-1}, \ldots \end{aligned}$$

Doubly-infinite strip with varying re-use
Consider again the doubly-inifinite strip, but now a scenario with varying re-use as described in Example 3.3. Suppose that Π^K is the set of all subnetworks of the form $\{(i+1, 2), \ldots, (i+K, 2), (i + \lfloor (K+1)/2 \rfloor, 1)\}$. Again, a simple limiting argument indicates that the bound of Theorem 4.4 applies. Exploiting the symmetry and the convexity of the function $R(\cdot)$, we find that $\lambda_{\max}$ is bounded above by the optimum value of the following convex programming problem

$$\begin{aligned} \min \quad & R(1; w_1, \ldots, w_K) \\ \text{sub} \quad & \sum_{k=1}^K w_k = 1 \\ & w_1 = w_K, w_2 = w_{K-1}, \ldots \end{aligned}$$

4 Numerical results

We now revisit the examples studied in Section 2 to illustrate how the revenue-based bounds may be used to improve upon the Erlang bound.

Example 3.1 (cont'd)
We return to the three-cell network described in Example 3.1, but we no longer assume that the offered traffic is uniform. From Theorems 4.1 and 4.4, we conclude that the carried traffic is bounded above by $\min_{0 \leq y \leq 1} V(y)$, with $V(y) = R(1, y, 0) + R(0, 1-y, 1)$. Since the function $R(\cdot)$ is convex, the function $V(\cdot)$ is convex as well. Thus, for uniform offered traffic, we may conclude from symmetry that $V(y)$ is minimal for $y = 1/2$. For $C = 2$ channels and offered traffic $\nu = 1$, we obtain an upper bound of 1.8 on carried traffic, which corresponds to a lower bound $\overline{B} = 0.4$ on blocking, tightening the Erlang bound.

Now suppose the offered traffic vector is $(\nu_1, \nu_2, \nu_3) = (1/2, 1, 3/2)$. The Erlang bound then yields $\overline{B} \approx 0.405$. Since the function $V(\cdot)$ is convex, we may use a bi-section search to find that in this case $V(y)$ is minimal for $y \approx 0.263$. This yields an upper bound of approximately 1.708 on carried traffic, which corresponds to a lower bound $\overline{B} \approx 0.431$ on blocking. Using Markov decision theory, we find that the minimum achievable blocking is in fact $\overline{B} \approx 0.439$.

□

Example 3.2 (cont'd)
We return to the doubly-infinite strip studied in Example 3.2. If we consider a subnetwork of just two cells, then we obtain $\lambda \leq R(1/2, 1/2)$. Note that $R(1/2, 1/2) = \nu(1 - \mathrm{Erl}(2\nu; C))$, so the revenue-based bound then coincides with the Erlang bound. If we consider a subnetwork of three cells, then we obtain $\lambda_{\max} \leq \min_{0 \leq y \leq 1/2} R(y, 1 - 2y, y)$. If we take a subnetwork of four cells, then we find $\lambda_{\max} \leq \min_{0 \leq y \leq 1/2} R(y, 1/2 - y, 1/2 - y, y)$. In both these cases, the convexity properties allow us to minimize the function numerically through a simple bisection search. Considering a subnetwork of five or more cells would involve solving a convex programming problem in more than one dimension.

We have conducted numerical experiments to compare the bounds with the performance of FCA and that of Maximum Packing (MP) as described earlier. The blocking for FCA is obtained from the Erlang-B formula (1). The blocking for MP is computed using the exact analytical results obtained in [3]. The results for $C = 20$ channels are shown in Figure 5.

The figure indicates that, unlike the Erlang bound, the revenue-based bounds closely approach the performance of MP. This suggests not only that the bounds are extremely tight, but also that no DCA scheme, however sophisticated, can be expected to outperform MP in any significant manner, if at all.

□

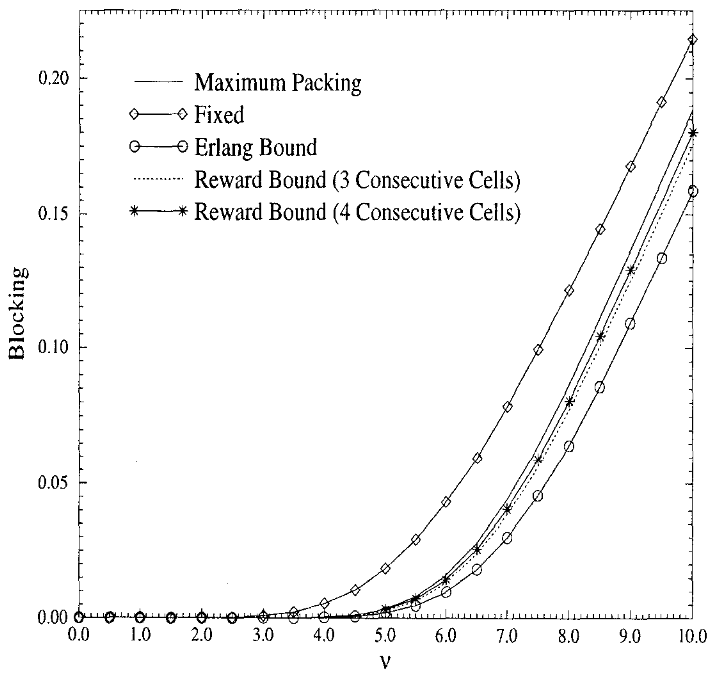

Figure 5: Erlang bound, revenue-based bounds, and performance of FCA and MP for $C = 20$ channels.

Example 3.3 (cont'd)
We return to the doubly-infinite strip with varying re-use as described in Example 3.3. If we consider a clique consisting of one inner cell and two outer cells, then we obtain $\lambda_{\max} \leq R(1; 1/2, 1/2)$ as an upper bound on carried traffic. If we consider a subnetwork consisting of two cliques with a common inner cell, then we obtain $\lambda_{\max} \leq \min_{0 \leq y \leq 1/2} R(1; y, 1 - 2y, y)$. As before, the convexity properties allow us to minimize the function numerically through a simple bi-section search. The calculation of $R(\cdot)$ in each iteration, however, is of formidable complexity for all but the smallest number of channels, and is the main obstacle in considering larger subnetworks.

We have performed numerical experiments to compare the bounds with the performance of a version of MP adapted to the varying re-use constraints. Under MP, this particular scenario may be viewed as a form of underlay-overlay cellular network, but with dynamic allocation of channels, see Lee [7]. The blocking for MP is calculated using exact results obtained by extending the analysis presented in Kelly [3]. The results for $C = 10$ channels and a fraction $\alpha = 0.3$ of traffic offered to the inner regions are shown in Figure 6.

The figure reveals that the blocking of outer calls is about twice that of inner calls for moderate values of blocking. Drawing upon the theory of loss networks, see Kelly [4], this ratio may be understood from the fact that outer calls require a channel in four cliques, whereas inner calls in only two. The revenue-based bounds now slightly diverge from the performance of MP. Although MP inflicts higher blocking on outer calls than inner calls, it does not so to the extent required to maximize carried traffic. This phenomenon reflects the trade-off between efficiency and fairness that arises in the case of varying re-use, see also Shimada *et al.* [11] and Valenzuela [12]. Schemes which minimize blocking intrinsically favor inner calls over outer calls, whereas schemes which do not discriminate among calls inevitably produce higher network-average blocking.

□

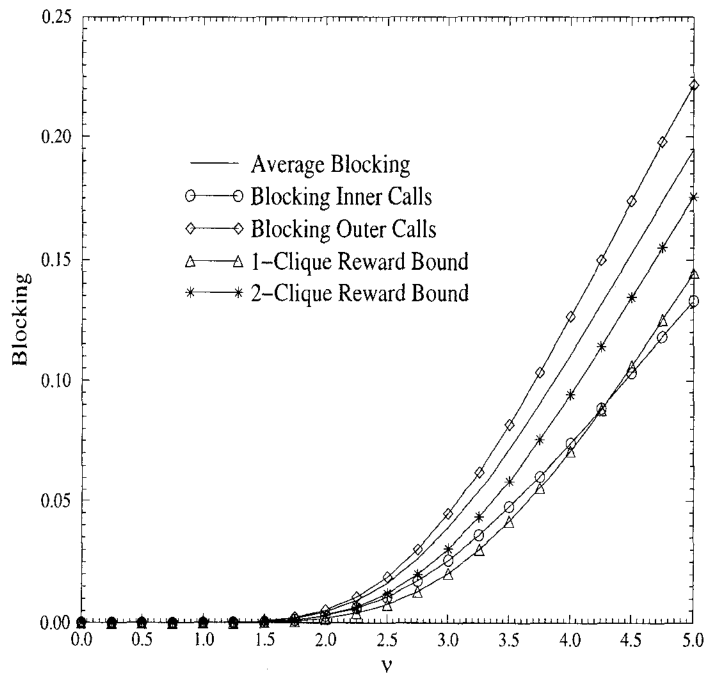

Figure 6: Revenue-based bounds and performance of MP for varying re-use with $C = 10$ channels and a fraction $\alpha = 0.3$ of traffic offered to the inner regions.

5 Asymptotic analysis

We now further investigate the trade-off between efficiency and fairness that arises in the case of varying re-use. We focus on the doubly-infinite strip with uniform offered traffic of Example 3.3. We consider a scenario in which the number of channels and the offered traffic grow large in proportion to one another, i.e., $C \to \infty$, $\nu \to \infty$, and $\nu/C = \rho$. Note that $\mathrm{Erl}(\rho C; C) \to \max\{1 - 1/\rho, 0\}$ as $C \to \infty$ for any $\rho \geq 0$.

Denote by B_1 and B_2 the blocking of inner and outer calls, respectively. Denote by λ_1 and λ_2 the carried traffic in each of the inner regions and outer regions, respectively. By definition, $\lambda_1 = (1 - B_1)\nu_1$, $\lambda_2 = (1 - B_2)\nu_2$. Note that $\overline{B} = \alpha B_1 + (1-\alpha)B_2 = (\nu_1 B_1 + \nu_2 B_2)/\nu = 1 - (\lambda_1 + \lambda_2)/\nu$.

Considering a clique of one inner cell and two outer cells, we have

$$\lambda_1 + 2\lambda_2 \leq R(1,1,1), \quad (15)$$

$$\lambda_1 \leq \alpha\nu, \quad (16)$$

$$\lambda_2 \leq (1-\alpha)\nu. \quad (17)$$

Observe that $R(1,1,1) \leq \min\{(2-\alpha)\nu, C\}$. Maximizing $\lambda_1 + \lambda_2$ subject to the constraints (15)-(17), we find that $\overline{B} \geq \beta^*$, with

$$\beta^* = \begin{cases} 0 & \rho \leq \frac{1}{2-\alpha} \\ 1 - \frac{1+\alpha\rho}{2\rho} & \frac{1}{2-\alpha} \leq \rho \leq \frac{1}{\alpha} \\ 1 - \frac{1}{\rho} & \rho \geq \frac{1}{\alpha}. \end{cases} \quad (18)$$

Define $\gamma := \min\{\alpha\rho, 1\}$. Now suppose we reserve a fraction γ of the channels for the inner calls, and leave the remaining fraction $1 - \gamma$ of the channels for the outer calls. Then $B_1 = \mathrm{Erl}(\alpha\rho C; \gamma C)$, $B_2 = \mathrm{Erl}((1-\alpha)\rho C; (1-\gamma)C)$. It is easily verified that $\overline{B}$ approaches β^* as $C \to \infty$, i.e., the bound β^* is asymptotically achievable and hence tight. Observe that this strategy only grants capacity to the outer calls that is essentially not needed by the inner calls. This strongly suggests that schemes which minimize network-average blocking will intrinsically favor inner calls over outer calls.

We now examine what the increase in blocking is if we require the blocking of inner calls and outer calls to be equal. Adding the condition $\lambda_1/\nu_1 = \lambda_2/\nu_2$ to the constraints (15)-(17), before maximizing $\lambda_1 + \lambda_2$, we find that $B_1 = B_2 = \overline{B} \geq \beta^\#$, with

$$\beta^\# = \begin{cases} 0 & \rho \leq \frac{1}{2-\alpha} \\ 1 - \frac{1}{(2-\alpha)\rho} & \rho \geq \frac{1}{2-\alpha}. \end{cases} \quad (19)$$

Now suppose we allocate a fraction $\alpha/(2-\alpha)$ of the channels to the inner calls, and assign the remaining fraction $(1-\alpha)/(2-\alpha)$ of the channels to the outer calls in each cell. Then $B_1 = \mathrm{Erl}(\alpha\rho C; \alpha C/(2-\alpha))$, $B_2 = \mathrm{Erl}((1-\alpha)\rho C; (1-\alpha)C/(2-\alpha))$. It is easily verified that B_1, B_2, and $\overline{B}$ approach $\beta^\#$ as $C \to \infty$, i.e., the bound $\beta^\#$ is asymptotically achievable and hence tight.

Define $\delta := \beta^\# - \beta^*$. From (18), (19),

$$\delta = \begin{cases} 0 & \rho \leq \frac{1}{2-\alpha} \\ \frac{\alpha}{2} + \frac{1}{\rho}\left(\frac{1}{2} - \frac{1}{2-\alpha}\right) & \frac{1}{2-\alpha} \leq \rho \leq \frac{1}{\alpha} \\ \frac{1}{\rho}\left(1 - \frac{1}{2-\alpha}\right) & \rho \geq \frac{1}{\alpha}. \end{cases}$$

This confirms that schemes which do not discriminate among calls inevitably produce higher network-average blocking.

We now analyze the asymptotic performance of MP. The Erlang fixed-point approximation for MP may be constructed as follows.

$$1 - A = \mathrm{Erl}\left(\alpha\rho CA + 2(1-\alpha)\rho CA^3; C\right),$$

$$1 - B_1 \approx A^2,$$

$$1 - B_2 \approx A^4.$$

This approximation is consistent with the earlier observation that for moderate values of blocking, i.e., $A \approx 1$, the blocking of outer calls is about twice that of inner calls.

Asymptotically,

$$A \to \min\{\frac{1}{\alpha\rho A + 2(1-\alpha)\rho A^3}, 1\}.$$

Since the Erlang fixed-point approximation is asymptotically exact, see Kelly [4], $\overline{B}^{MP} \to \alpha G + (1-\alpha)G^2$, with

$$G = \min\{\frac{-\alpha\rho + \sqrt{\alpha^2\rho^2 + 8(1-\alpha)\rho}}{4(1-\alpha)\rho}, 1\}.$$

It may be verified algebraically that $\beta^* \leq \overline{B}^{MP} \leq \beta^\#$ for all values of α and ρ. Figure 7 plots the values of β^*, $\beta^\#$, and $\overline{B}^{MP}$ as a function of ρ for $\alpha = 0.3$.

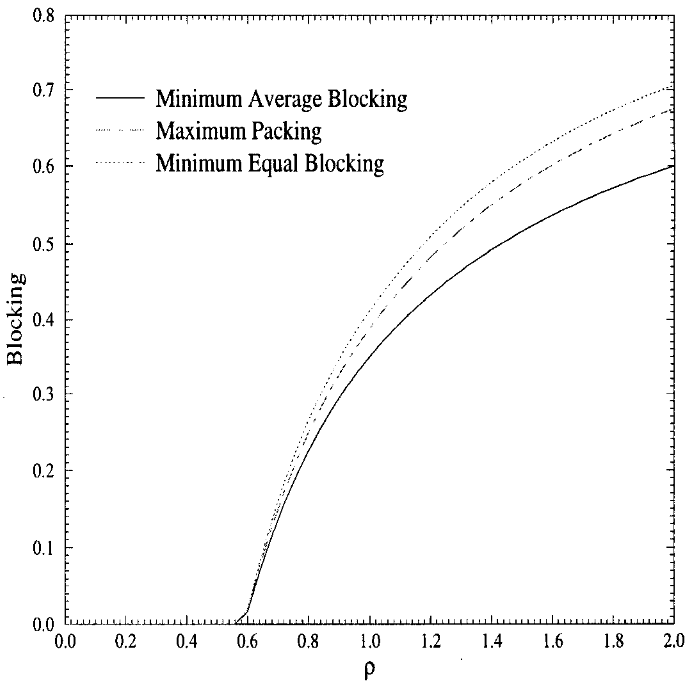

Figure 7: Minimum average-blocking β^*, minimum equal-blocking $\beta^\#$, and asymptotic performance of MP for varying re-use with a fraction $\alpha = 0.3$ of traffic offered to the inner regions.

6 Conclusion

The Erlang bound may not always be tight because it fails to exclude carried traffic combinations which are only feasible if call dropping were permitted. The 'trunk reservation bounds' which we introduced are also obtained by considering cliques of cells in the network. The construction of these bounds is based on a reward paradigm as an intuitively appealing way of characterizing the *true* achievable carried traffic region, thus exposing any infeasible combinations that may weaken the Erlang bound.

Even tighter bounds may be obtained by not considering cliques, but subnetworks of cells in which a channel may be used more than just once. In one-dimensional networks, the revenue-based bounds then closely approach the performance of Maximum Packing (MP). This suggests not only that the bounds are extremely tight, but also that no DCA scheme, however sophisticated, can be expected to outperform MP in any significant manner, if at all. The fact that such tight bounds can be obtained by considering just 3 or 4 consecutive cells in the network is striking. For a given subnetwork, no tighter bound can be obtained, since the reward paradigm completely demarcates the achievable carried traffic region.

Subsequently, we considered scenarios with varying re-use which may arise in the case of dynamic re-use partitioning or measurement-based DCA schemes. The revenue-based bounds extend to these scenarios, but the computational complexity increases further, which means that only relatively small subnetworks can be considered. In these circumstances, however, the bounds slightly diverge from the performance of MP, which inflicts higher blocking on outer calls than inner calls, but not to the extent required to maximize carried traffic. This phenomenon reflects the trade-off that arises in the case of varying re-use between efficiency and fairness. This observation is consistent with the empirical finding in Shimada *et al.* [11] and Valenzuela [12], obtained by means of simulation, of inhomogeneous blocking in distinct portions of the coverage area. Asymptotic analysis confirms that schemes which minimize blocking intrinsically favor inner calls over outer calls, whereas schemes which do not discriminate among calls inevitably produce higher network-average blocking.

In the present paper, we have not considered mobility, but calls in hand-off may be expected to experience similar high blocking as peripheral calls at set-up. Since calls already in progress should in fact receive a preferential treatment, this suggests an even greater need for a channel reservation mechanism for hand-offs in the case of varying re-use.

Finally, we add that our bounding approach not only applies to cellular networks, but also to loss networks in general. A related analysis appears in Kelly [5].

References

[1] Everitt, D.E., MacFadyen, N.W. (1983). Analysis of multicellular mobile radio telephone systems with loss. *Brit. Telecom Techn. J.* **1**, 37-45.

[2] Jordan, S., Khan, A. (1994). A performance bound on dynamic channel allocation in cellular systems: equal load. IEEE *Trans. Veh. Techn.* **42**, 3137-3143.

[3] Kelly, F.P. (1985). Stochastic models of computer communication systems. *J. Roy. Statist. Soc.* B **47**, 379-395.

[4] Kelly, F.P. (1991). Loss networks. *Ann. Appl. Prob.* **1**, 319-378.

[5] Kelly, F.P. (1994). Bounds on the performance of dynamic routing schemes for highly connected networks. *Math. Oper. Res.* **19**, 1-20.

[6] Kind, J., Niessen, T., Mathar, R. (1997). Theory of maximum packing and related channel assignment strategies for cellular radio networks. Preprint Aachen Univ. of Techn.

[7] Lee, W.C.Y. (1989). *Mobile Cellular Telecommunications Systems* (McGraw-Hill, New York).

[8] Lippman, S.A. (1975). Applying a new device in the optimization of exponential queuing systems. *Oper. Res.* **23**, 687-710.

[9] Miller, B. (1969). A queuing reward system with several customer classes. *Mgmt. Sc.* **16**, 234-245.

[10] Raymond, P.-A. (1991). Performance analysis of cellular networks. IEEE *Trans. Commun.* **39**, 1787-1793.

[11] Shimada, K., Watanabe, T., Sengoku, M. (1995). A dynamic channel assignment approach to reuse partitioning systems using rearrangement method. IEICE *Trans. Fundam.* E**78**-A, 831-837.

[12] Valenzuela, R.A. (1993). Dynamic resource allocation in line-of-sight microcells. IEEE *J. Sel. Areas Commun.* **11**, 941-948.

[13] Whiting, P.A. (1989). Performance bounds for cellular radio management schemes. In: *Proc. 6th UK Teletraffic Symposium, IEE.*

[14] Xu, Z., Akansu, A.N. (1997). A performance bound for interference adaptation dynamic channel allocation in wireless communication networks. Preprint NJ Inst. of Techn.

[15] Zander, J., Eriksson, H. (1993). Asymptotic bounds on the performance of a class of dynamic channel assignment algorithms. IEEE *J. Sel. Areas Commun.* **11**, 926-933.