A HIGH CAPACITY ASSIGNMENT METHOD FOR CELLULAR MOBILE TELEPHONE SYSTEMS

Ray W. Nettleton
Stanford Telecommunications, Inc.
7501 Forbes Blvd., Suite 105
Seabrook, MD 20706

Gerald R. Schloemer
P.O. Box 307
Round Lake, IL 60073

Abstract

This paper describes a channel assignment scheme in which assignment is based on actual field strength measurements performed dynamically on the functioning system. The scheme involves no rigid partitioning of the channel set into geographic patterns, as used in conventional systems. A mutual interference criterion is applied to ensure that satisfactory service is maintained throughout the system.

The system provides capacity improvements ranging from 100% to 300% depending on the complexity of the implementation. It is implementable using any of the current or proposed modulation schemes.

Introduction

The concept of cellular telephone systems has been well established for some two decades, and actual systems have been deployed for about seven years. The success of the concept from a marketing point of view is beyond question; indeed, since its inception the industry has been scrambling to meet current demand and to plan for future demand.

The efficient use of the spectrum is of paramount importance in the quest to meet demand, since new spectrum is unlikely to be allocated to this service, and the cost of base station sites increases as the system becomes more dense. Spectral efficiency depends on two interdependent technologies; on one hand, the modulation technique used must permit many channels per MHz to be defined; and on the other, the method used to assign channels to mobiles must make efficient use of the channel resource. Many new ideas have been introduced in both of these areas in the past two decades.

The present paper concentrates on the second factor discussed above; we introduce here a new concept for channel assignment [1] that substantially increases system capacity over existing techniques. The technique can be adapted to any narrowband modulation scheme, and seems to be most suitable for the new digital technologies presently being proposed for use in cellular mobile radio.

Background: How Real is a Cell?

Existing strategies for assigning channels in a cellular mobile radio system rely on a single principle; the geographical separation of zones in which use of a given channel is permitted. In this sense the scheme is "new" only in its physical scale; the idea of "coordination zones" is central to the whole concept of channel sharing as it is practiced by the FCC and by PTTs all over the world. In the latter context, however, the zone may be as large as a city, a state, or a country; in the cellular case, it may be as small as a few city blocks.

The idea of geographical separation serves the radio regulatory process well, mainly because the radio terminals are usually fixed and, in any case, over very large distances coupling between cochannel systems is easier to predict and to control. In cellular radio, the distances are small and most terminals are mobile; there is almost never direct line of site coupling between transmitter and receiver, and propagation loss for a given link is a stochastic process.

The generally used criterion for choosing the "preferred" link between base station and mobile is that the link with lowest path loss is assigned. It is easy to show that in typical propagation conditions the probability that this preferred link is also the shortest is less than 50%. It follows that, though perhaps useful as a concept, the "cell" as a geographic entity is quite illusory.

Consider the ideal model that is often used to familiarize the newcomer with the cellular technique. In this model propagation loss is assumed to be a monotonic function of distance from the base station. (Typically an inverse fourth-power law model is used.) Thus every mobile is connected to its nearest base station, as shown in Figure 1a. The cell idea is obvious in this diagram, since no link crosses a cell boundary.

In an actual system, however, the "inverse fourth-power law" of propagation is only a mean, and the actual path loss is a log-normal variable (Gaussian when measured in dB) with standard deviation ranging from 8 to 12dB. (Here we assume

CH2379-1/89/0000/0359 $1.00 © 1989 IEEE

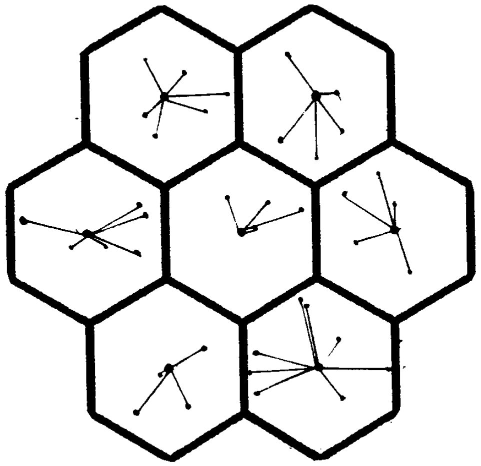

Figure 1a. Links between mobiles & base stations using an ideal propagation model

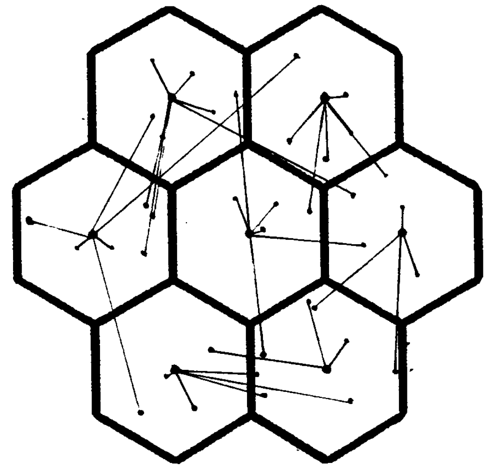

Figure 1b. Links between mobiles & base stations when shadow fading is included in the propagation model

that the short-term Rayleigh fading has been averaged out; in any case, Rayleigh fading is best considered to be a wave interference phenomenon rather than a true path loss effect.) Indeed, when the cell size is very small the conventional "inverse-fourth-power law" of propagation appears to be totally invalid [2], [3].

Under these conditions the preferred link is much less often between a mobile and its closest base station; the situation is more likely to be as shown in Figure 1b. In this diagram the hexagonal boundaries have no meaning with respect to the link patterns. If, indeed, cells can be identified, they are certainly not hexagonal, and they are almost never simply-connected figures. The cell is a nebulosity at best.

A New Technique

The new technique is quite simple; the set of available channels is not divided into disjoint partitions, but each base station is (in principle) allowed to use any of the available channels (though not all at once!) Propagation measurements from each base to each mobile are made, averaged over a short time to remove most of the variation due to Rayleigh fading. Then a procedure is used to assign channels based on an interference criterion. That is, a new call is permitted to use a given channel if it causes no interference to the calls already in progress on that channel, and if it will receive no interference from the existing calls. The assignment is made regardless of the number of calls on the channel, and regardless of the distance between base stations using the same channel.

Figure 2 illustrates the principle of the interference-checking mechanism. Suppose Mobile X is already using a channel via Base A, and Mobile Y wishes to share it using Base B. The interference is required to be 15dB below the wanted carrier.

The diagram shows the four path loss measurements that need to be made in this case, and the table below it shows how these four figures are used to derive four signal-to-interference ratios. In the illustration, all four SIRs are above the 15dB threshold, so this is a permissible sharing situation.

In passing we note that it is quite possible for the base-to-mobile links to pass the test while the mobile-to-base links fail it, or vice versa. For example, if the SIR threshold were set at 25dB, base B's signal would fail the test while all others passed. This suggests that further efficiency could be gleaned from such a system by "divorcing" the traditional channel pairs and allowing

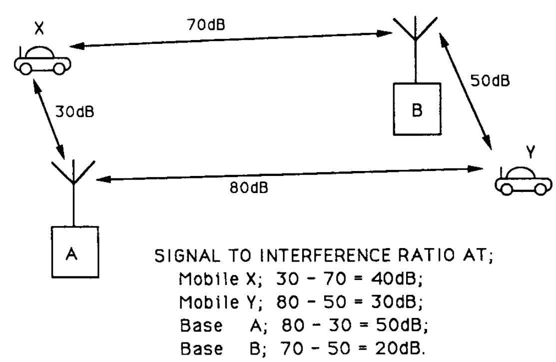

Figure 2. An illustration of the SIR calculations for the new technique

sharing independently on the two halves of the channel set. This intriguing possibility has not been investigated to date, however, and will not be discussed further here.

It should be noted that the interference calculations needed for this scheme require only the propagation measurements referred to in Figure 2, and that the existing cellular systems make these measurements now as the means of selecting the preferred link for a call. Thus the only changes called for in the cellular system infrastructure are tuneable base station radios, and a more powerful computer in the MSO to implement the selection procedure.

A Family of Implementations

The non-interference criterion outlined above is not sufficient to fully define a channel assignment technique. Specifically, it defines which channels should not be assigned to the new or handing-off call, but it is quite likely that this will result in many candidate channels that satisfy the criterion. The remaining questions are;

- How should the system decide which of the candidate channels should be selected to handle the call?

- Should the new assignment be made strictly on the basis of keeping all existing calls in place, or should some or all of the existing calls be re-assigned to further optimize the solution?

The answers to these questions are interrelated, and create an entire class of assignment techniques. The simplest possibility is to leave all existing calls in place (as is current practice) and to assign the new/handoff call to the first channel that is found to satisfy the non-interference criterion. The most complex solution is to use an optimizing technique such as linear programming to reorganize the entire system each time a new assignment is required, so as to maximize some figure of merit such as the minimum system-wide signal-to-interference ratio.

In our research we have not yet attempted to employ an optimizing technique, since for a realistic system simulation time the computational load was considered excessive. Instead we have investigated several heuristic schemes, ranging from the simplest scheme mentioned above to a complex technique that re-organizes the entire system periodically.

In this paper we report our results from the two extremes mentioned above. They were the least and the most successful of the techniques as determined by the criterion of most calls handled by a fixed number of channels. We also report on a third scheme with intermediate results and complexity to illustrate that an entire range of possibilities exist.

The Simulation

A new simulator has been written to test the efficiency of the new techniques and to compare them with existing technology. A simplified flowchart of the simulator is shown in Figure 3. We list below the primary features of the simulator.

- The simulated system consists of 19 cells of equal radius and uniform geographic distribution; see Figure 4.

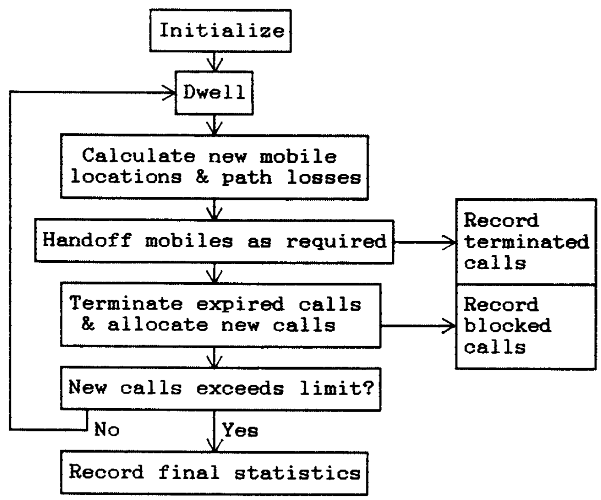

Figure 3. Simplified flowchart of the Simulator

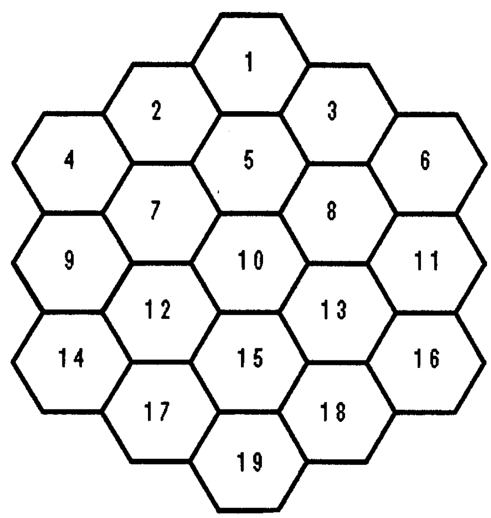

Figure 4. The nineteen-cell configuration of the simulator

- The propagation model consists of an inverse - α - law model with log-normal shadow fading having standard deviation σdB. α is hard-coded in zones; for short links it is 3.0, for intermediate links it is 3.5 and for long links it is 4.0, with smooth transitions from zone to zone. σ is variable at the program's input menu. Rayleigh (fast) fading is not modeled since the signal-to-interference measurements modeled are assumed to average these effects out. A realistic noise floor is included in the SIR calculations.

- 140 duplex channels are available for calls. Control channels are not explicitly modeled.

- Call statistics are modeled as a birth-death process, i.e. independent arrivals and exponential call durations. Mean duration of call and mean number of calls are variable at the program's input menu. The program's arrays can handle up to 1000 simultaneous calls.

- Mobile distribution is modeled as uniform within each of three concentric circles, roughly covering the center cell, the "ring" of cells surrounding the center cell, and the outside "ring". The relative density of mobiles in each ring can be set independently at the input menu.

- Mobile velocity is modeled as uniform from 0 to 100km/hour, with random directions. For simplicity, motion is held fixed during a call except that mobiles that try to wander outside the modeled system are reflected back at the system's rectangular boundary.

- The noninterference SIR criterion threshold, and the SIR threshold for initiating a handoff request, can be independently set at the input menu.

- Sampling is performed at 5 second intervals in simulated time. At each sample the position of each mobile updated, all propagation paths are recalculated and handoffs initiated as required. New calls are initiated and some existing calls are terminated. Finally statistics on mobile and call distribution, calls blocked and terminated, and link quality are accumulated before proceeding to the next sample.

- The simulation is terminated after a fixed number of call initiations has been logged. This number can be set at the input menu. The final statistics are calculated and recorded on disk for later printout.

- Optional graphics displays are available so that the experimenter can see the physical system layout in motion. This verifies the correct functioning of the simulator as well as providing useful feedback on system dynamics. A pan and zoom capability is included so that the entire system or just a portion of one cell can be viewed. Whether or not graphics output is selected, the screen display includes continuously updated statistics so that the progress of the simulation can be monitored in real time.

- The program is modular so that specific functions such as channel assignment can be rewritten with minimal disturbance to the program's architecture. The simulator is written in PASCAL and runs on a PC. It consists of some 2,500 lines of code.

- Depending on the complexity of the assignment technique and the speed of the PC, the simulator may run several times faster than real time or two orders of magnitude slower.

Analysis

Five simulations are described here. The first two are conventional systems of the type currently deployed [4]; the remainder are implementations of our new concept. The third and fifth are at opposite ends of the complexity gamut and the fourth is of intermediate complexity.

1. Conventional System, Fixed Assignment

This is the simplest embodiment of the cellular concept; it was implemented universally in the earliest systems, and is still widely used today.

The set of available channels (140 in our case) is divided into seven equal subsets (of 20) and the subsets, labeled A, B, ... F are assigned irrevocably to specific base stations, as shown in Figure 5. The assignments are made so as to minimize interference.

When a call is initiated, field strength measurements are used to determine the best base station with which the mobile should be linked (the one with the least path loss.) If a channel is free on that base station, the connection is completed. Otherwise, the call is blocked.

As it moves about the service area, the mobile monitors its own signal-to-(noise plus interference) ratio. When this ratio falls below a threshold, the mobile requests a handoff. Once again using field strength measurements, the best base station is selected and the call is transferred. If no channel is available on the new cell, the call is terminated.

2. Conventional System, Directed Retry

This system is set up exactly as described above under Fixed Assignment. It operates the same way, except for the as-

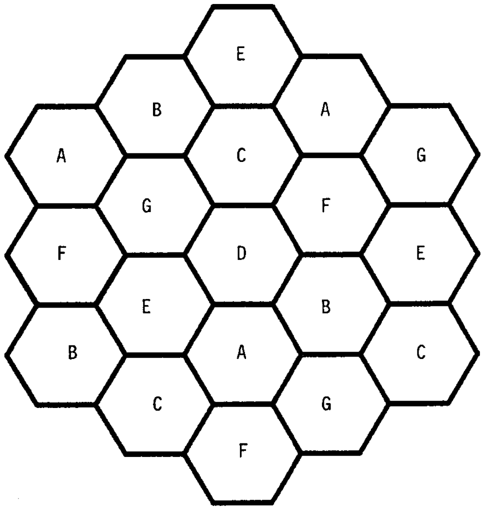

Figure 5. Division of the cells into seven-cell clusters for the fixed assignment technique

signment procedure. Here, if no channel is available at the preferred cell for a new or handoff call, the mobile is directed to the second preferred cell, that is, the cell with the second lowest path loss. This reduces the probability of blocking or termination, because each call gets two chances at finding an available channel.

3. Undivided System, No Optimization

This is the simplest embodiment of the new technique. No division of channels is made; each base station has about fifty radios, each capable of tuning to at least a large subset of the available channels.

When a call is initiated, the system determines the preferred cell for the call. Each channel not currently used at the base station of this cell is monitored to determine the signal to interference ratio that would result if the new call were to be added to the channel. The first channel to be found that satisfies the interference criterion is assigned the call. If no such channel is found the call is blocked.

As is the case with the Fixed Assignment scheme, the mobile requests a handoff when its signal quality falls below a threshold. The handoff is handled in the same way as a call initiation; if no channel is found the call is terminated.

The difference between the two schemes is that, in the Fixed Assignment scheme, blocking and termination can occur due to unavailability of a base station radio, even when the call could be tolerated from the interference viewpoint. In the new scheme the radio complement of each base station is sized to prevent this from happening until traffic loads are much higher, at which point interference considerations are already dominating the lost call statistics.

This technique allows mobiles not requiring a handoff to keep their assignments while the others are assigned new channels.

4. Undivided System - the "Target" Technique

This heuristic is very simple. A "target SIR" - typically 5dB above the threshold T - is defined. Then each mobile requiring an assignment is assigned to the channel for which the mean-squared deviation of mobile SIRs from the target is minimized.

Suppose the target is G dB, and there are N mobiles assigned to the channel. Then for each channel we calculate

$$D = \frac{1}{N+1} \sum_{i=1}^{N+1} (SIR_i - G)^2$$

... and pick the channel with the smallest D. Of course, each candidate channel must also satisfy the non-interference criterion.

This technique also allows mobiles not requiring a handoff to keep their assignments while the others are assigned new channels.

5. Undivided System - a Masking Technique

In this case the system is set up as above, but is re-organized periodically in order to maximize the traffic carried consistent with the interference criterion.

This system is so efficient at packing mobiles onto channels that it is very sensitive to changes in the path loss figures. Changes in relatively few of the links appear to necessitate a global re-organization; experiments aimed at allowing unperturbed links to keep their assigned channels, while locally reorganizing the perturbed links, have resulted in a marked, step-by-step deterioration in system capacity.

The details of the technique are quite complex, and are included in an appendix for the interested reader.

Results

Figure 6 shows some results of the simulation using the following parameters:

- Handoff SIR threshold 15dB
- Interference criterion threshold 15dB
- Shadow fading standard deviation 10dB

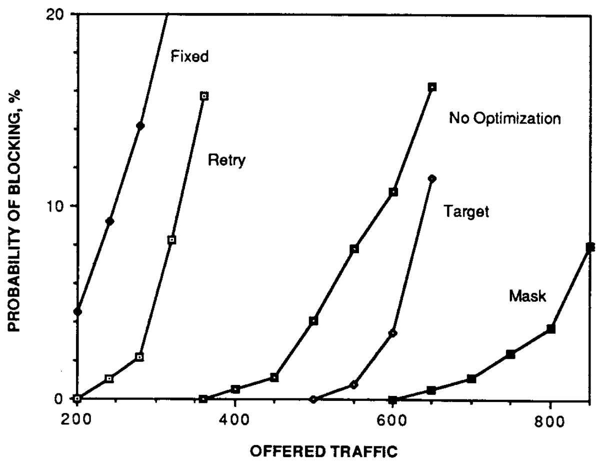

Figure 6. Results of the simulation with shadow fading σ = 10dB, handoff threshold = interference criterion threshold = 15dB

In the graph the term "blocking" includes both blocking at call initiation and termination during handoff. In each case the statistics are dominated by the termination event.

It is generally accepted that a good mobile telephone service grade is characterized by blocking levels around 4%. At this level of service, the five schemes have approximately the following total call capacities in the simulated system;

Table 1. Results Compared

Technique	Capacity	Improvement
Fixed Assignment	200	0%
Directed Retry	300	50%
No Optimization	500	150%
Target Technique	610	205%
Masking Technique	800	300%

The figures under "Improvement" indicate the improvement in capacity of the technique compared with the capacity of the Fixed Assignment technique.

Discussion & Conclusions

The results shown are for a more-or-less homogeneous system of small size. It could be the model of an entire system of modest dimensions, or it could be the dense center of a larger system. In any case, the spectral efficiency advantages of the undivided system - even in its simplest incarnation - are obvious.

Of course, spectral density is not achieved without cost. In this case there are two parameters that decline as efficiency improves.

Average signal-to-interference ratio declines as efficiency improves, as does the standard deviation of this parameter. However, no mobile suffers an SIR below the threshold, which is controllable at will as a fundamental design parameter. Of course, efficiency varies inversely with the signal quality criterion. This allows capacity vs. quality tradeoffs to be used not only in the system design phase, but also dynamically in response to increasing traffic demand.

More seriously, the average duration of a call between handoffs also declines as the efficiency improves. For the unoptimized and "target" schemes, mean time between handoffs is about half the duration of the fixed assignment and retry schemes. This is probably well within the capacity of currently available hardware.

But in the case of the masking technique, in order to achieve maximum capacity it is necessary to re-organize the system completely every sample period. (In our simulation this was every five seconds, but we doubt that this would be necessary in practice.) It is almost certainly impossible to achieve handoffs at this rate using currently available hardware.

It should also be noted that, in the new schemes, it is common for a mobile to be handed off to a new channel but not to a new base station (or even vice versa), in order to circumvent an interference problem. In conventional systems, handoff always implies a new base station <u>and</u> a new channel.

The progressively increasing capacity that results from progressively complex techniques and hardware suggests that the new technique could be implemented in phases to limit cash flow difficulties in system evolution.

Finally we note that the computational complexity implicit in the masking scheme, and in other attempts at system optimization, probably will stretch the limits of conventional computer systems, if not technologically then perhaps in terms of cost effectiveness. Very likely parallel processing, neural networks, or some other combination of advanced computational technology would be needed to implement an advanced system of this type.

Appendix: The Masking Technique

Concept

The underlying concept of the masking technique is to sort the mobiles into types according to their interference potential, then pack the maximum possible number of mobiles of one type onto each channel. The packing procedure starts with the mobiles having the least potential for interference and proceeds by admitting mobiles that have progressively more potential for interference until all possible assignments have been made. This ensures that mobiles that can be packed many to a channel are given first priority, and the "channel hogs" - those mobiles that require a dedicated channel - are most likely to be blocked.

The Masks

The mask is the tool used for sorting and packing. Two types of mask are used; the mobile mask and the channel mask. Both types of mask are arrays of symbols, with the number of elements equal to the number of cells in the system. Each element is a symbol taken from an alphabet of cardinality three.

Mobile m is assigned a mobile mask M(m,b). Element b corresponds to the b-th base station, and takes on one of the following three states;

State P; Base station b is the preferred base station, i.e. the one with the least path loss. Only one base station in each mask takes on this state, and we will call it element p.

State I; Marks the base as a potential interferer, i.e.
$L(b) < (L(p) + T)$, where
L(b) is the path loss from the mobile to base station b;
L(p) is the path loss from the mobile to the preferred base station p; and
T is the interference criterion threshold, i.e. the lowest permissible SIR.
That is, if M(m,b) is in state I, it says that a channel assigned to mobile m from base p should not be used on base b.

State N; Neither of the above. This implies only that a channel assigned to this mobile from base p *might* also be used on this base with no interference.

Path losses are expressed as positive quantities in dB.

In addition, a weight W(m) is assigned to each mask; this is just the number of elements in the mask having states P or I. Thus, a mobile having mask weight 1 is likely to be assignable to a channel that is already used several times, since it has no low-loss propagation paths to bases other than its preferred base. Conversely, a mobile with mask weight equal to the number of cells is a "channel hog"; the channel it is assigned to cannot be used anywhere else in the system. The technique starts by assigning the mobiles with the lowest weight first.

Channel c is assigned a channel mask C(c,b). Element b corresponds to the b-th base station, and takes on one of the same three states as the mobile mask, each with the same meaning. The channel mask is also assigned a weight in the same way as the mobile mask. The channel mask weight is a measure of how many bases are currently denied to mobiles wishing to have access to the channel, i.e. it measures how "full" the channel is.

Mask Merging is necessary in order to transfer the information in the mask for each mobile using a given channel onto the mask for that channel. To merge mobile mask M(m,b) onto channel mask C(c,b), we perform operations on each element b in turn of the channel mask, according to the following truth table. Here the rows correspond to the channel mask element before merging, the columns represent the mobile mask element being merged, and the table entries are the states of the merged elements.

Table 2. Truth Table for Merging a Mobile Mask into a Channel Mask

	P	I	N
P	P	P	P
I	P	I	I
N	P	I	N

In words, ranking P as most important and N as least important, each element of the merged channel mask is the more important of the old element and the corresponding element of the mobile mask being merged with it.

The following observations on the merged channel mask should be noted;

1. A channel of mask weight K □ (the number of cells in the system) could be in use by K mobiles with mask weight 1, or by one mobile with mask weight K, or by any intermediate combination.

2. A channel of weight less than K is not necessarily accessible by any new mobile, since the accumulated effect of interference might place the SIR of one or more mobiles already assigned to the channel, close to the threshold T. The channel mask does not reflect accumulated interference.

The Technique

Figure 7 gives a simplified overview of the masking Technique. It is described in narrative form below.

START

Before any channel assignments can take place, every mobile is given a mobile mask and a mask weight. This implies also that a preferred base station has been assigned to each mobile.

A test variable C is set to 1. C is the maximum weight allowed for a merged channel mask during the current loop.

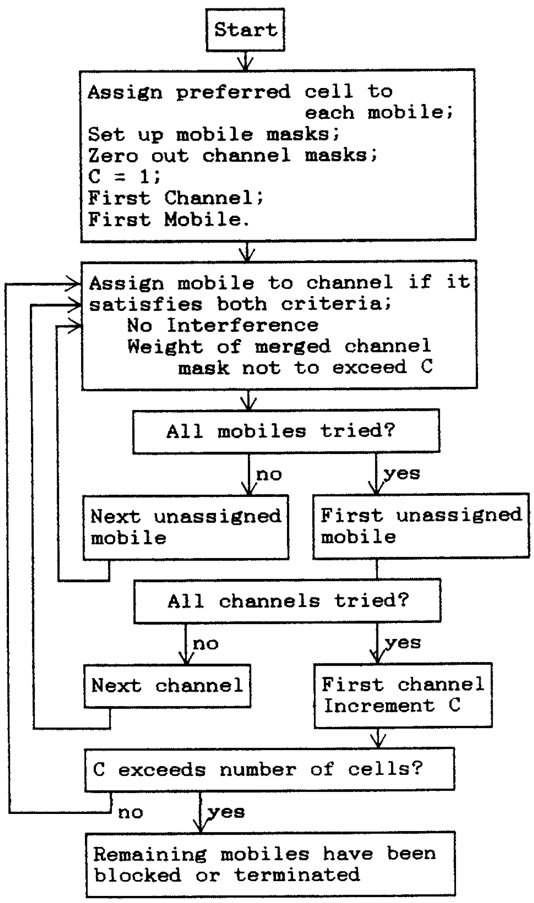

Figure 7. Simplified flowchart of the masking procedure

Starting with the first channel and the first mobile (which are in no particular order), the mobile is assigned to the channel if it satisfies the following criteria;

1. The weight of the channel mask, after merging with the mask of the candidate mobile, does not exceed C; and

2. The signal to interference ratio of every mobile using the channel, including the candidate mobile, does not fall below T. This is actually four distinct criteria;

   a) The candidate mobile's transmission to its base station must not cause interference to the other mobiles' transmissions to their base stations; and

   b) The other mobiles' transmissions to their base stations must not cause interference to the candidate mobile's transmission to its base station; and

   c) The candidate mobile's transmission from its base station must not cause interference to the other mobiles' transmissions from their base stations; and

   d) The other mobile's transmissions from their base stations must not cause interference to the candidate mobile's transmission from its base station.

If all criteria are satisfied, the candidate mobile is assigned to the channel. If not, it is left on the list of mobiles waiting to be assigned.

After all the candidate mobiles have been examined, control returns to the first unassigned mobile on the list and increments to the next channel. The above procedure is then repeated.

After all the channels have been examined with the current value of the test variable C, control returns to the lowest channel and the variable C is incremented.

If C is not greater than the number of cells in the system, the above procedure is repeated.

If C is greater than the number of cells, the procedure is terminated and assignment is complete. Any mobiles left unassigned are either blocked (if they are initiating a call) or terminated (if they are handing off.)

END

Discussion

It should be noted here that the masks are constructed using only the mobile-to-base links, even though the interference criteria are applied to mobile-to-base and base-to-mobile links. It would theoretically be possible to construct base-to-mobile masks also, but they would have to be used in a separate procedure to assign base-to-mobile links only. In the research reported on here, we have retained the convention that the two links come in inseparable pairs. Greater capacity would no doubt result if these pairs were "divorced" and the two channel sets assigned independently.

References

[1] Gerald R. Schloemer, "Method and Apparatus for Making Frequency Channel Assignment in a Cellular or Non-Cellular Radiotelephone System", U.S. Patent no. 4,736,453, April 5, 1988.

[2] J. H. Whitteker, "Measurements of Path Loss at 910MHz for Proposed Microcell Urban Mobile Systems", IEEE Transactions on Vehicular Technology 37:3, August 1988, pp. 125 - 129.

[3] P. Harley, "Short Distance Attenuation Measurements at 900MHz and 1.8GHz using Low Antenna Heights for Microcells", IEEE Journal on Selected Areas in Communications 7:1, January 1989, pp. 5 - 11.

[4] William C. Y. Lee, Mobile Communications Engineering, McGraw-Hill, 1982.