Bandwidth Reservation strategies for mobility support of wireless connections with QoS guarantees

Ruston Hutchens

Network Programmer, Information Technology Services
La Trobe University, Melbourne, VIC 3086Australia

Samar Singh

Dept. Computer Science and Computer Engineering
La Trobe University, Melbourne, VIC 3086Australia.
`{R.Hutchens, S.Singh} @latrobe.edu .au`

Abstract

This paper examines QoS guarantees for bandwidth in mobile wireless networks, with a focus on reducing dropped connections on handoff. We can achieve this by reserving bandwidth for connections that might move into a cell from a neighbouring one. We develop a novel framework for analysing issues relevant to handoff. The principal novelty of this framework is the use of an arbitrary planar graph (network) to model the adjacency relationships of cells in the network. Mobility patterns of the mobile stations are then captured by simple probabilities for moving to a neighbouring cell, leading to a notion of a shadow cluster that is very easy to implement. Three strategies for bandwidth reservations are then proposed and investigated. They range from simple (used as a baseline for comparison) to more complex ones using the shadow cluster concept. We study these strategies by simulation and attempt to determine important parameters and quantify what is gained by added complexity and other variations. Our results show that the topology has an important influence on the results of a reservation policy. However, one of our strategies (the optimistic one) appears to be uniformly superior, because it adapts to changing network topologies as well as movement patterns.

Keywords: Wireless networks, handoff dropping, bandwidth reservation.

1 Introduction

We consider here a typical cellular network of the future, with mobile hosts requesting connections with QoS guarantees on bandwidth. The mobiles require different amounts of bandwidth, depending on the nature of the applications that are running. Because bandwidth is a scarce resource in wireless networking, it is necessary to allocate it carefully amongst competing connections. Thus not only must we consider new calls that are being initiated, but also those that may move into a cell from a neighbouring one. Users can freely roam within a network's area of coverage, and experience a large number of handoffs during a typical session. These users expect good quality of service from the system, e.g. low delay, small call-dropping and blocking probabilities. This therefore requires that sufficient bandwidth be reserved for callers from adjacent cells, so that when they move, bandwidth will be available, and their call will not be dropped. We must balance the requirement that bandwidth be available for established calls that wish to move, while also providing bandwidth for new calls. It may be better to reject a number of connections at initialisation (increased “call blocking”), rather than have to drop them at a later stage. On the other hand, too large a reservation would lead to under-utilisation of the network. While much work has been done on this issue, most of it is narrowly focussed on mobile telephony.

The work of Levine et al. [1], brings together a number of ideas expressed in earlier papers, focussing on:

- Handoff in cellular ATM networks [6]
- Wireless networks [7] [8]
- Mobility patterns of wireless users.

The main idea of [1] is the concept of *Shadow Clusters*: every mobile terminal with an active wireless connection exerts an influence upon the cells (and their base stations) in the vicinity of its current location and along its direction of travel [1]. As an active mobile moves, the region of influence also moves, following the active mobile to its new location. As different calls cast their shadows over cells, the ‘sum of the shadows’ is used to determine if new calls should be admitted. The shadow cluster mechanism will only admit calls that are considered likely to complete. In order to reduce the number of connections that are dropped at handoff due to insufficient bandwidth being available in the new cell, *bandwidth reservation in the shadow cluster* is introduced. Two network topology scenarios are discussed; the one-dimensional highway case and a standard two-dimensional hexagonal structure.

Algorithms are provided to calculate the shadow for each mobile connection in its ‘home’ cell, and the other cells around it, for both the one-dimensional highway case, and

---

the two-dimensional hexagonal cellular structure. In computing the reservations to be made in the shadow cluster, *significant weight is placed on the speed and direction* the mobile is travelling in. In practice, the network cannot know much about this until *after* the call has been admitted, since the mobile may have just recently been switched on, or it may register its presence to the network only occasionally, to reduce the registration overhead. Further, to gather this information requires much co-ordination amongst the base-stations, especially as it may be changing all the time. Thus, finally, only the one-dimensional scheme is evaluated for performance in [1].

Choi and Shin [2] follow up [1], and also use the one-dimensional highway case. They however propose an adaptive bandwidth reservation scheme, which does not require any knowledge of the mobile's speed and direction at call initiation, but only the bandwidth required. The simulation of the highway in [2] takes into account the fact that there would be no calls entering from either 'end' of the highway, and therefore artificially joins the two ends together. The authors also make an assumption that calls will only move in one direction, and never turn around. They propose to determine the amount of bandwidth to reserve using an adaptive handoff scheme that estimates the optimum amount of bandwidth to reserve in a particular cell. The scheme will react to handoff drops in a particular cell, and will attempt to adapt the amount of bandwidth reserved to ensure that this requirement is met (handoff drops per time period for a given cell do not exceed the predetermined limit). Three variations of this Admission Control scheme are proposed in [2]. Ramanathan et al. [3], propose a dynamic scheme called ExpectedMax, which basically computes the maximum expected value of the bandwidth requirement in a cell, taking into account information from neighbouring cells. This scheme is compared with a modified version of the work of Yu and Leung [8], which was originally restricted to mobiles requiring a single type of connection, and with various "fixed" types of schemes. ExpectedMax uses past history statistics to adjust *reservation thresholds* using some theory, and is therefore similar only in spirit to our pessimistic scheme (described below). The simulations performed are also more comprehensive but are concerned with typical scenarios such as "evening rush hour" and so on. Recently, others have again considered predicting the direction in which mobiles are moving and using that for Call admission control and bandwidth reservation. However, our attempt will be to see what can be achieved without directly attempting to predict the mobility of calls. We make further comparisons with Ramanathan et al. [3] during and after the description of our schemes.

In the current paper, we make three main contributions for studying this question.

1) Topology: Previous work [1][2][3] has focussed on either a straight-line scenario in the simplest case, or a regular, hexagonal structure. Both of these are rather limiting, and do not accurately model the real world. In reality, base stations are erected in a highly irregular topology which depends on many factors such as availability of the location, likely density of users in the area etc. To capture this reality, our topology is represented as a (arbitrary) planar network. Details on how we obtain these planar networks are discussed in the section "Developing a mobile wireless network topology". Our results indicate that the topology can have an influence on the performance of the reservation scheme.

2) Generality: As a generalisation, we dispense with the notion of user "classes" and allow mobiles to request any amount of bandwidth within a certain range. This is equivalent to having as many classes as possible values of requested bandwidth. This allows us to define a general reservation scheme quite independent of the "class" unlike in [3]. Furthermore, we avoid having specific scenarios in our simulations such as " rush hour" etc. Rather, we generate movement patterns at random and attempt to find a strategy that is uniformly superior to the others under all such variations. And finally, we dispense with "blocking probability" as the chief measure of the efficacy of the strategies and introduce the notion of "goodput", which is the total time of a completed call multiplied by its bandwidth, summed over all such calls. We argue that blocking probability has less meaning in the context of variable bandwidth requests because, calls requesting more bandwidth are more likely to be blocked under almost any scheme. For example, [3] is forced to compute the blocking probabilities separately for each class, for making comparisons amongst schemes, and when there are many classes, this obviously becomes cumbersome. Our approach also contrasts with the approach taken in [3], where by having different thresholds for different classes, the scheme tries to be "fair" and equalise the blocking probabilities. We would argue that this notion of "fairness" is questionable.

3) Bandwidth Reservation Schemes: We give a different, much simplified, and more generic quantification of the shadow cluster idea, which is far easier to obtain by actual measurement.

- It is independent of the particular mobile and depends only on the particular cell in which the mobile is located.
- It does not depend on estimation of the speed or other characteristics of the mobile.

It is simply the statistical pattern of movements out of the cell into neighbouring cells. This pattern can be estimated from past history by the Base station and indeed be updated continuously. It is similar to the data collected by the proposed scheme of [3]. However, using this new notion of the shadow cluster, we provide two generic reservation schemes, which could be refined in various directions, as done in [3], by using a more sophisticated model. There is a pessimistic (conservative) scheme and an optimistic scheme and we compare these with a baseline "optimised scheme" which is actually ***impractical,*** but gives some idea of the best possible performance. We have used the simple strategy as our baseline. We have optimised the simple strategy by running the simple strategy many times with different reservation thresholds. Clearly this is not possible to do in the real world. The optimistic scheme uses the idea of overbooking in a very simple and general way, which is also consistent with a reduction in messaging overhead. It turns out that our optimistic scheme comes quite close to the "optimised" one. The advantage to this optimistic scheme is that it does not require any form of tuning (as the fixed "optimised" one does).

The rest of this paper is organised as follows. In the next section we describe in greater detail the analytical framework developed by us and the assumptions contained therein. In section 3, we discuss some details about the simulations carried out by us, the methodologies used in them and the quantities to be measured. Section 4 then discusses the results obtained, and Section 5 has detailed comparisons with the work of [3] and section 6 contains the conclusions with some indications of future work.

2 Analytical Framework

2.1 Obtaining the Network Topology

Choi and Shin [2], and Levine et al. [1] develop a simulation, in which they attempt to simulate a highway, which can effectively be represented in a one-dimensional space, and make suggestions for a two-dimensional hexagonal topology, which is used by [3]. Whilst this is an improvement, there is the certainly the possibility to develop a topology that is more akin to a real world environment. When we move from a sparsely populated region into more densely populated area, the density of base-stations will increase, thus increasing the number of neighbours each base station has. The use of exactly six neighbours for every cell may actually bias the results in some way, and show one reservation strategy to be better than another, when for another topology, the opposite is true. In our view, an arbitrary planar graph or network is necessary to simulate the mobile wireless network, with the nodes or vertices representing the cells in the network, and the edges representing the adjacency relationships between the cells. That is, an edge between two nodes indicates that the two cells are neighbours and have a common boundary. It is now easy to see that the network should be planar if the adjacency relationships are to be representative of real networks constructed ***on the surface of the earth.*** While it is possible that cells may overlap, a mobile terminal can only talk to one base station at a time, therefore this is not an issue. Similarly, physical land features are irrelevant, as the terminal can only communicate with one base station at a time. We only concern ourselves with which base station the terminal is communicating with, not where the nearest base station is. Indeed there may be a base station that is geographically nearer to the one that the mobile is communicating with, but hidden behind a hill. We use a set of randomly generated planar graphs to test our strategies so that any systematic bias is avoided.

2.2 Other Assumptions

We make a number of assumptions about the total system. These are reasonable ways to model the system, but they need to be spelled out and explained for completeness. A strategy that works well for many different such graphs would then be expected to work well on any other graph.

System Wide Assumptions

1. The statistical probabilities for call movement out of a cell are generated at simulation set-up, and remain fixed for the entire simulation run. This is designed to model network movement patterns that would be gathered in a real mobile wireless network. There is a separate probability that the mobile will move or not move. That represents, in a sense, the "speed" of the mobile.
2. At the end of the simulation, all calls will simultaneously terminate. This simplifies the problem of trying to account for calls that are still ongoing at the time the simulation ends.

Cellular Assumptions

1. All cells have the same total bandwidth available. We do not consider the effects of fading etc. and assume perfect transmission. We also assume that any amount of this bandwidth can be allocated to a mobile at connection (call) set-up, subject to some granularity. Thus we are considering a future system in which a call may specify how much bandwidth it requires and ask for a guarantee that it will be provided.
2. Calls can only be dropped during a handoff (base-stations and mobile hosts never fail)
3. Cells will have at least one neighbouring cell (and preferably do not have more than 10-15)
4. At most one connection request is generated in each cell for one given time period. This is not restrictive as the time period can be made very small with respect to the length of a call.

Call Assumptions

- Calls model CBR connections in the ATM world. Bandwidth is determined at call initialisation, and is fixed for the duration of the call.
- Each call can have a different bandwidth, where the bandwidth can be any integer up to CALL_BW where CALL_BW is the largest bandwidth a call can request as specified in the simulation configuration file.

- Calls will only talk to one base station at any one time (i.e. the complexities of “soft handoff” are suppressed).
- Once a call has moved into a cell, it will remain there for a minimum of one time period. Once again this is fine if the time period is very small.

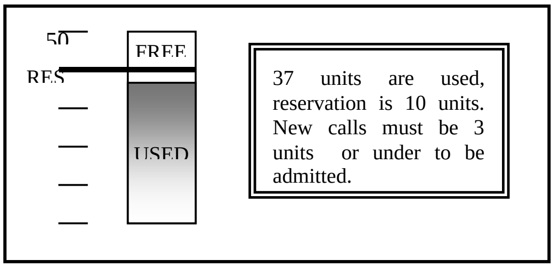
2.3 Reservation schemes.

We propose three simple schemes for bandwidth reservation. The first is extremely simple and used only for comparison. The other two are based on our idea of simplified shadow clusters, and requires that all the neighbours of any cell keep a certain amount of bandwidth in reserve for each of the calls in that cell. One of our two strategies is pessimistic or conservative in operation, while the other is optimistic and hopes to get away with overbooking. Thus the conservative strategy will hardly ever drop an existing call, but will under-utilize the bandwidth in the system, while the optimistic one will attempt to overcome this problem at the cost of having to drop more calls occasionally.

2.3.1 “Optimised” Fixed reservation Technique

This proposes to reserve a set amount of bandwidth in each cell. In our simulation this value is predetermined at set-up time, and does not alter. We set a threshold value for the system. If the *admission* of a *new* call would exceed this threshold value, the call is blocked. Obviously, this scheme is much too simple to be much good. However, we use it to serve as a benchmark by “optimising “ its performance in an unrealistic way.

Since the simple scheme has a fixed reservation percentage, we carry out a search over the percentage values to get the *best one for any given scenario.* This “tuned” version of the fixed scheme is then used for comparison with the other schemes. However, from a practical viewpoint, it represents an unrealistic “ideal” situation, in which the statistics of the call movement are known beforehand and we can find the best threshold value for them. It is interesting to note that this simple “optimisation” does better than any of the other schemes proposed by us, irrespective of changes in other parameters such as topology and movement patterns.



37 units are used, reservation is 10 units. New calls must be 3 units or under to be admitted.

Figure 1: Fixed Bandwidth reservation strategy

It is useful to compare our use of the fixed scheme with that in [3], where the figure 5% is chosen arbitrarily and compared with the proposed ExpectedMax Scheme.

2.3.2 Pessimistic Reservation Strategy

This is a more complex scheme and is based on our simplified shadow cluster. Rather than a simple test to determine if the admission of a new call overshoots a set value, we introduce the concept of a separate, dynamic *bandwidth reservation index*. This index (BRIndex), tries to guarantee that BRIndex more units of bandwidth will be available for calls already in the system, (but in other neighbouring cells) in preference to new calls. When the network accepts a new call, or a call moves into a cell from a neighbouring cell, the neighbouring cells of the newly occupied cell are notified and a specific amount of bandwidth is reserved. The exact amount reserved in each neighbouring cell is determined by:

- the bandwidth of the call
- the probability that the call will move to that cell

In fact, these two items are multiplied to determine that amount to reserve. This amount is added to BRIndex More formally, q_i is the probability that the cell will move (at all) in the given time period, p_{ij}, is the probability that a call moves from cell i to cell j, if it does move, and b_k is the bandwidth requested by call k. Then, if call k is in cell i (either initiated as a new call or entered from a neighbouring cell) cell j is required to reserve $b_k\ q_i\ p_{ij}$ units of bandwidth and add it to BRIndex.

As soon as the call is terminated, or the call moves out of this cell, the neighbours are informed to remove the amount of bandwidth that was reserved for this particular call. In the case of a movement, the new neighbours make reservations according to the same scheme, but using the probabilities of the new location. Now, when such a bandwidth reservation is being made, an additional check is made to determine if, the additional bandwidth being reserved would increase the BRIndex to more than the total amount of unused bandwidth. In this conservative scheme, the *reserved bandwidth* must be equal to, or less than the amount of *unused bandwidth.* **If this is not so, the call is dropped.** This happens at call admission as well as on handoff. It is easy to see, in this scheme that every call has a fraction of its bandwidth requirement reserved (and left unused) in the neighbouring cells. This fraction depends on the probability of movement, which in our case is the same for all cells and held constant at 0.4. Thus, the actual utilisation of capacity will never exceed about 71% on a system-wide basis.

Once again, it is useful to compare this with the way the reservation amount is computed in [3], which uses a more complex expected value computation, starting from the same probabilities of moving from some cell i to neighbouring cells j. Our computation is extremely simple, independent of any update period, and has a low message overhead as discussed below. Nevertheless, the ExpectedMax strategy of [3] does resemble our conservative strategy in its broad features.

2.3.3 Optimistic Reservation Strategy

This is very similar to the pessimistic strategy, except that we now optimistically hope that overcrowding will be reduced in neighbouring cells before a call moves. Thus,

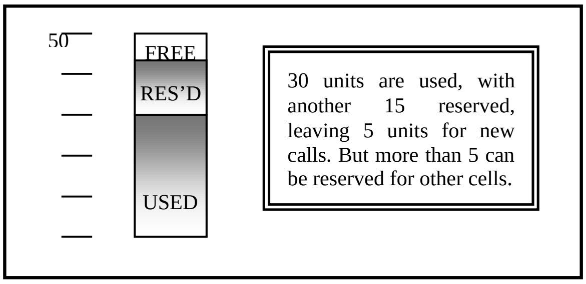


Figure 2: Optimistic bandwidth reservation strategy

in this optimistic scheme, it is possible for the sum of the used and reserved bandwidths to exceed the total available, provided this is caused by movement of calls already within the network (growing the reserved bandwidth allocation), but it cannot be caused by the admission of new calls. In this way, we give greater importance to calls already initiated compared to new calls. Thus, a new call request can be blocked in two ways: either there isn't enough bandwidth to allocate to it in the home cell, or the neighbouring cells cannot reserve enough bandwidth to allow movement. The BRIndex figure is variable, and reflects the load in cells immediately surrounding the current cell, and the probabilities that calls will move into the current cell (captured in the shadow cluster probabilities of those cells and the movement probability). Clearly, this scheme permits more bandwidth to be reserved in a cell than is actually available. We are *optimistically* hoping that calls will not move to the overcrowded cell, either by moving elsewhere, by hanging up, or by staying still. In the meantime, no new calls will be accepted by that cell and the chances are that the overbooking will be relieved by calls moving out or terminating. It turns out that this hope is justified to a large extent.

2.4 Signalling Overhead of the Schemes.

Notice that these two schemes require some signalling overhead between neighbours, but this occurs only at handoff or at call admission. Thus, the base station needs to send a message to each neighbour when a new call enters the cell. Further, the message is simply a request to reserve a certain amount of bandwidth and there is an ack or nak for that request. Similarly, if it is a handoff, the base station, which has handed off the call, also needs to send a message asking each neighbour to unreserve the requisite amount of bandwidth. Considering all the communication that takes place during call admission and handoff anyway, this extra overhead should be negligible, particularly as it can be out of band. In the case of the optimistic scheme even the acks and naks are not required. The base station assumes the reservation is made.

Another aspect of our schemes is the concept of the shadow cluster, which is obtained as the probability for any cell that a call will move to each of its neighbours respectively. These probabilities can be updated continuously, based on past statistics. All that is required is to count the number of calls that have crossed over to a neighbouring cell in a given time period. This does not involve anything more than incrementing a counter at the base station at each handoff. Similarly, the probability that the call will move out of the cell is again based on raw statistics of the past period and does not require any sophisticated computation, as in [3].

2.5 Choice of parameters to be measured.

All work that has been done on simulating a mobile wireless network up to date has focussed on reducing the number of handoff drops, and has measured only this quantity, or the blocking probabilities. We add another measure, “goodput”, or the network throughput of calls that completed successfully. This is calculated as the amount of bandwidth used, times the duration of the call. Some simulations also refer to “Dropput”, which is the network throughput that calls achieved, before they were dropped.

If it were only the number of handoff drops that we were concerned about, it would be simple to give preference to calls with smaller bandwidth, while allowing those with large bandwidth to drop. Similarly, a simple *blocking probability* measure is not really meaningful, as we need to take account of the bandwidth requested by the call. A call that requests a large chunk of bandwidth is more likely to be blocked. This can be captured perhaps, by finding an “average blocked bandwidth”, but it is not clear how this figure is to be interpreted. As pointed out already, it is necessary in [3] to have separate blocking probabilities for each “class” of call. Secondly, there is always a trade-off between the blocking probability for a handoff and the blocking probability for a new call. Thus every strategy is characterised by (at least) two probabilities i.e. a vector. Since vectors can only be partially ordered, the question that naturally arises when no ordering is possible, is: which strategy is better? We have attempted to get over this difficulty by using the Goodput as the metric. While all these other statistics have indeed been gathered from the simulations, because it is simple to do so, we concentrate here on the “goodput”.

Goodput captures the *blocking probability* aspect (as it measures effective utilisation) as well as the handoff dropping aspect, since dropped calls occupy bandwidth but do not add to Goodput**. Therefore, to simplify the discussion, and to have a metric that gives a clear ranking, the results of the simulations are presented in terms of goodput only.** It is again worth noting that the results of our simulations indicate that the use of goodput leads to a ranking of our strategies, that is robust against changes in other parameters, such as the topology and the movement patterns.

3 Implementation.

We have simulated a mobile wireless network from the perspective of calls entering, moving through and exiting the network. The first step in carrying out a simulation is the consideration of the topology of the mobile wireless network itself. After we have determined the network topology, we can then simulate call flow in the wireless network, implementing the appropriate bandwidth reservation policy.

We have been able to obtain many, large randomly generated planar networks, using the planarity testing algorithm of Tarjan,1973, [4] which was implemented by Mehlhorn, Mutzel and Näher in the LEDA package, [6].

3.1 Developing the Simulation

The simulation program, written in C, takes the output of the LEDA graph package, and simulates a mobile wireless network with that topology. Whilst the output of LEDA is not a bi-connected planar graph, it is a trivial task to make it so in our simulation.

3.2 Running the Simulation

Once we have the network, and have generated the movement probabilities, we can start the simulation. In each time-period, we inspect all nodes, and the calls that are currently at that node. One of the assumptions we make is that a call can only reside in a single cell. When a call moves from one cell to another, the handoff from the source to destination is *instant*, provided there is sufficient bandwidth to accommodate the call in the destination cell. Another assumption is that once a call has moved to a cell, it will remain there for one time period. To implement this we set a timer at call admission. This timer is checked when this call becomes a candidate to move, or hang-up (terminate). At least one time period must have passed before the call becomes eligible for movement or hang-up. If a call becomes eligible for movement, and moves, the timer is reset. This is to avoid a *"Silly Movement Syndrome"*. This describes a situation where a call's location or state is transformed an excessive number of times in one time period.

```
Set-up Simulation
for (simulation_length)
 for (all_nodes)
   {
     if(new_call)check avail
bandwidth/reservation
      if (ok_to_admit)
            ACCEPT
      else
            REJECT
      for (call_list)
       if (move) check avail
bandwidth /reservation
        if (ok_to_move)
            MOVE
        else
            DROP
       if (hang-up)
            HANG-UP

    }
End Simulation
```

Simulation Pseudocode

Once we have finished the call initiation phase (if appropriate), we process the current call to see if it should move (probabilities for call initiation, movement and hang-up are read from a configuration file). If it should move, we can determine (with the given probabilities) where to move the call. First we test to see if the destination cell can accommodate the call. If not, the call is dropped, and the dropput counter is incremented with the amount to bandwidth the call has used during its connect time to date. If we are using the *optimistic* or *pessimistic* reservation strategies, we also need to reserve bandwidth with the new cell's neighbours. As a minor point, we note that reservations are all in integers and real values obtained in any computation are truncated. If the call does not move, we check for call hang-up. In such a case, if we are using the *optimistic* of *pessimistic* bandwidth reservation strategies, we must also remove the bandwidth reservations from the neighbouring cells.

This process is repeated for all calls in the original cell's call list. After we finish processing the call list, any calls that remain are to stay in the current cell, and we move on to the next cell, and repeat the process. This is repeated as many times as required by the length of the run.

3.3 Simulation Length

In our investigations, we examined a number of parameters that could affect the results of the simulation. All results were verified using 99% Confidence Intervals, on simulation length runs of 50,000 time intervals, with mean values gathered from sub-runs of 1,000 time intervals. As a result of calculations of confidence intervals, we report all results for simulation lengths of 50,000 cycles.

It is worth noting in what follows, that the term *graph* is used to refer to a planar graph or network that represents a mobile wireless network topology. The term *plot* refers to the graphical representation of a particular set of values with respect to another.

3.4 Random number seed and scenarios

The initial seed value chosen makes more of a significant impact on the results of the simulation. This is due to the fact that so much of the simulation ***set-up*** relies on the generation of random numbers. Thus each seed corresponds to a certain scenario in terms of movement probabilities. Using 99% confidence intervals for the fixed reservation scheme, with all parameters other than the seed value fixed, we found the values of the goodput varied as shown below:

Seed	Lower Bound	Upper Bound
1	124,152	126,025
938322360	119,114	121,356
938322580	110,248	112,412

The initial network set-up is clearly an important issue. Therefore we use these three default seed values to illustrate the results throughout the discussion. We will see however that other factors (such as network topology graph) have a greater role in the results of the simulation. The initial seed can be seen as modelling the time of day/ day of week of the simulation. Mobility patterns will certainly vary depending upon whether it is a weekday or a weekend. Different seed values are used for completeness, to verify that the initial seed does not play a role in the ***ranking*** of goodput values for our reservation strategies.

3.5 Use of different network topology graphs

The use of different planar networks as a basis for the simulation has the greatest effect. The paper uses our two larger networks, Network1 (63 nodes/163 edges) and Network3 (51 nodes/133 edges) to illustrate the results. The larger networks are all well-connected and generally present a realistic picture of what a mobile wireless network of a regional city, or a slice of a larger capital city might look like.

4 Results

4.1 Fixed reservation strategy "optimisation"

The fixed reservation strategy was run with a random seed value of one, and simulation length of 50,000. The Cell bandwidth was set at 50 units, with individual call bandwidth up to 10 units. Simulations were run with bandwidth reserved for calls already in the system set at 0%, 20%, 40%, 60%, 80% of the available bandwidth. After analysis of the results for network1, further simulations were run for 36%, 42%, 44% and 46%. These determined that a fixed bandwidth reservation strategy, reserving 42% of bandwidth for calls already in the system produced optimal network bandwidth goodput for network1.

After analysis of the results for network3, further simulations were run for 26%, 34%, 36%, and 38%. This determined that 36% was the optimal reservation value for this network. This tells us already that the actual network used has a strong influence on the results. For, although the difference in Goodputs for the two networks and any given strategy, can be explained by their differences in size, the fact that the optimal goodput is achieved at different thresholds for the two networks is significant. Thus, a fixed reservation strategy with 42% reservation performs better than one with 36% on network1 *but the results are reversed on network3.*

4.1.1 Random number seeds and different topology networks

A simulation using the fixed bandwidth reservation strategy, with the parameters shown in the table was made for two networks, and three random number seeds. The networks used were Network1 and Network3, with the random number generator seeded with the values 1, 938322360 and 938322593 (referred to as seed 1, seed 360 and seed 593 in plots). The simulation was run with bandwidth reservations from 1 to 49. The following plot shows goodput (throughput for all calls that terminated successfully) for different levels of bandwidth reservation for one of the seeds.

Cell Bandwidth	50 units
Maximum call bandwidth	10 units
P(new call)	0.5
P(call move)	0.4
P(call stay)	0.6
P(call hang-up)	0.05
Simulation Length	50,000

We can see that the main factor affecting the goodput values is the network topology graph itself. The highest goodput values were all achieved for network1. This is indicated quite clearly as for **all** seed values, the goodput curves for network1 are consistently higher than for network3. This can to some extent be explained by the difference in the sizes of the two networks. Whilst the actual seed value has an effect on the curve, it does not affect the curves for different networks in a uniform way. Network1 has curves from two seed values that lie in overlapping 99% confidence intervals, and one outside. Network3 has three separate curves that are ordered completely differently from network1. This is entirely what one might expect, given that the seed values will have a random (unpredictable) effect on the movement statistics that are generated at set-up time. Note however, that topology and set-up together can have a significant effect, as will become even clearer later.

4.1.2 Investigation of new call arrival probability

The effect of altering the probability that a new call would arrive at a given node during any time period, was also investigated. The same parameters as above were initially used in this simulation, with the additional fixed reservation threshold set at 33 units (i.e. 33% reserved). Simulation was carried out for both network1 and network3. Plots of the results are displayed in Figure 3, for goodput of the two networks investigated. Call arrival probability was varied from 0.01 to 0.99. It was observed that the two curves displayed similar trends, with the peak for network1 higher than network3. The mobile wireless network becomes saturated with calls when call arrival is set between 0.3 and 0.4 calls per time interval, after which the goodput increases only slightly. We could normalise the goodput for the two networks by dividing by the number of cells (nodes in the network), but this does not change things much. Further investigation of call arrival and movement statistics was also made.

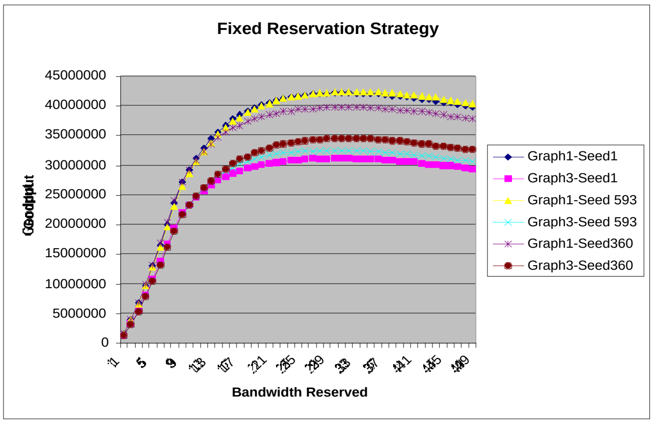


Figure 3: The effect of varying bandwidth reservation for the fixed strategy

4.2 Pessimistic reservation strategy

The pessimistic reservation strategy was run with our standard random seed values of 1, 938322360 and 938322593. The simulation length was 50,000 time intervals. Simulation parameters were as before, with the exception of the pre-reserved bandwidth. The pessimistic strategy was allowed to determine the amount of bandwidth that would be reserved for calls already in the network, according to the shadow cluster concept explained earlier. The performance of this scheme was substantially below either the (ideal) fixed reservation scheme or the optimistic scheme, as is presented below for network1 and network3.

Seed Value	Network1	Network3
1	15,821,958	12,013,981
360	13,732,870	14,136,984
593	15,823,510	13,144,142

This scheme has an interesting property: if a call has reserved bandwidth in neighbouring cells, the chances of there being sufficient bandwidth available for it to move into that cell are very good. i.e. the number of calls actually dropped is very low, but of course this comes at a price. This strategy is not very good when examined in terms of goodput. It could have potential uses for investigations involving other types of call traffic (non-CBR), such as VBR and Best Effort. In such cases, reserved bandwidth can be given over to other connections that require it, until such time as it is required for an entering call. If this is done, the total goodput of this strategy may not be very much worse, when the network is congested and demand for bandwidth from Best Effort connections is high.

One more fact worth noting is that for seed 360 this strategy actually produces a (significantly) higher goodput on network3 than on network1, whereas this has never been true for any other strategy. This demonstrates again the possible effects of a combination of topology and movement patterns (or set-up).

4.3 Optimistic reservation strategy

The optimistic reservation strategy was run with our standard random seed values and other parameters as before. The optimistic strategy determined the amount of bandwidth that would be reserved for calls already in the network, using the BRIndex as explained in section 2. The resulting Goodput is presented below:

Seed Value	Network1	Network3
1	39,493,495	29,115,325
360	37,412,610	33,067,445
593	39,662,058	31,447,748

The effects of topology are again worth noting here. Firstly, the difference in goodput between the two networks for seeds 1 and 593 are too large to be explained simply by differences in size. Secondly, seed 360 produces the lowest goodput of the 3 seeds on network1 but the highest on network3.

As with the pessimistic scheme, there is little we can 'adjust' in the optimistic scheme. Unlike the fixed reservation scheme, we cannot 'adjust' the level of reservation as this is handled by the optimistic algorithm itself (any adjustments would require a change to the algorithm). This becomes an advantage when choosing an *adaptable* strategy for network bandwidth reservations, as there is less information that must be known *a priori*. Goodput curves for Network1 and Network3, using our three standard random number seed values are plotted in Figure 5.

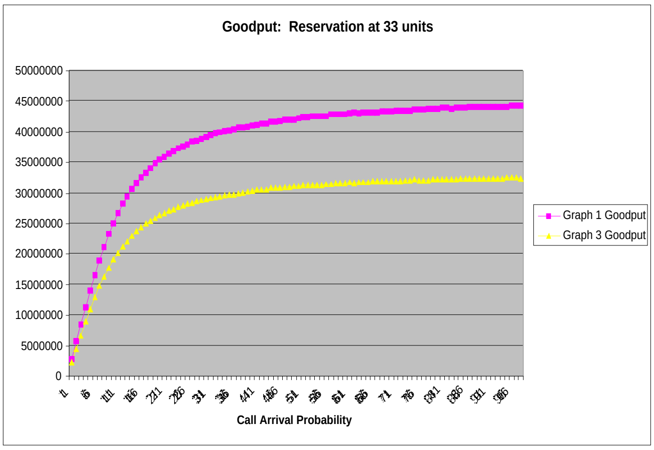


Figure 4: Goodput for varied call arrival probabilities

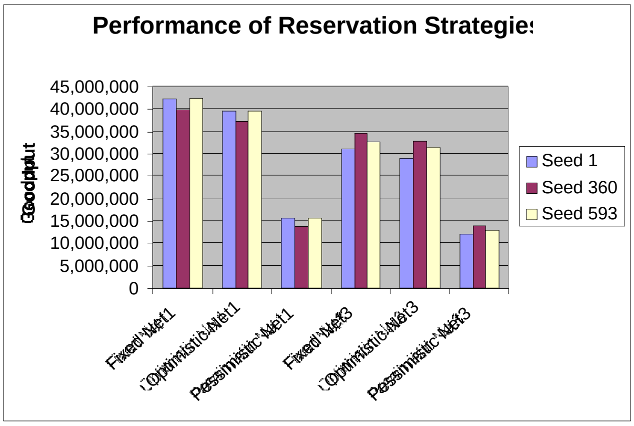


Figure 5: Comparison of proposed bandwidth reservation schemes

4.4 Comparison of the three proposed bandwidth reservation schemes

Figure 5 allows a comparison of the Goodput given by the (optimised) Fixed reservation scheme, Optimistic Reservation scheme and Pessimistic Reservation scheme. The figures for the Fixed scheme refer to the *best* values that were given for that particular network/seed combination. Recall that the *best* values for the fixed scheme were obtained by running multiple simulation runs, with identical network topology, call arrival and movement conditions, and represents a *benchmark* rather than a realistic strategy. A clear ranking of the three methods, purely in terms of goodput, is shown by the plots on Fig. 5. Because the three seeds have been used in each case, we see that the ranking is independent of the set-up probabilities, or the network topology.

Similarly, the table in Figure 6, shows a comparison, taking the optimised fixed scheme as 100% and then showing what percentage of the goodput could be achieved by the optimistic scheme, in the various cases. We can see here that the percentage values are fairly high, so that we cannot expect any significant improvement from our schemes unless the optimised fixed scheme itself can be substantially improved upon.

Seed Value	Fixed – Network1&3	Optimistic –Network1	Optimistic –Network3
1	100%	93.41%	93.22%
360	100%	93.82%	95.58%
593	100%	93.28%	96.50%

Figure 6: Performance comparison of Fixed and Optimistic schemes

5 Discussion of the results

It is useful to discuss the significance of our results in the light of the work done in [3]. Because of a number of differences in the methodology as well as in the final simulation scenarios the results there are not directly comparable to ours. However, some general comments can be made. Firstly, the comparisons there appear to be arbitrary in the case of the fixed strategy, and as we have seen, at least in the case of Goodput, a reservation of around 30% or 40% would have been more appropriate. Secondly, the results there are inconclusive from our point of view, as the proposed scheme always has the lowest blocking probability for handoffs but often has the highest one for new calls. Indeed in some cases, the rankings change with changes in load (new call rate). On the other hand, the results obtained by us appear to be quite unequivocal and indicate that refinement of the pessimistic strategy (which might lead to something like the ExpectedMax strategy) is less promising than a refinement of the optimistic one. Our results also suggest that any refinement in the model of the optimistic strategy may not really yield much at all, because we have compared with an "optimised" strategy, which always does better, but only marginally so. Another indication of this is the fact that, this optimised strategy comes close to achieving what might be considered the "saturation" value of good put for each graph. This can be seen from Figure 4 for example. Finally, as in many other situations in computer science, a "conservative" or "completely safe" strategy turns out to be too expensive in terms of resource, while the "overbooking" strategy is superior, because the overbooking is reduced by the time the call moves into a cell.

6 Conclusions and future work

This paper has examined bandwidth reservation schemes that attempt to minimise call dropping on handoff. We have presented a novel, most general, simplified framework and two new generic schemes. All our schemes are evaluated using a planar graph to model mobile wireless network topology, rather than one-dimensional or two-dimensional regular structures. In our view, the preferred solution would be to develop an adaptive bandwidth reservation algorithm that is able to accommodate different network topologies, and adapt to all changes in the usage patterns, without requiring mobility estimation of individual mobiles. We have used extensive simulations to rank the effectiveness of our schemes, as well as to determine the parameters of significance. In doing so, we have introduced the notion of "goodput" as a measure of effectiveness, in view of the difficulty of interpreting "blocking probability", when different calls request different bandwidths and are treated differently by the network as a result.

Our results show that network topology is a significant determinant of the outcome, and that the use of regular hexagonal cell structures could lead to results which may be rather different from what would occur in the real network. In particular, it is possible that one strategy that appears better than another in one scenario may actually perform worse under a different topology and with different movement patterns. Purely from the view of Goodput for CBR-type connections, the optimistic strategy using our simplified notion of shadow cluster, outperforms the pessimistic one in all situations and performs close to the *"optimised"* fixed strategy without requiring any form of tuning. In other words it is robust and adapts quite well to changing topologies and movement patterns. We are investigating the effect of estimation error in the probabilities. Also, we have not yet investigated the effect of time-varying movement patterns of the mobiles on these results, which is equivalent to the previous situation in a sense. Our schemes currently appear to define bandwidth reservation policies for the network as a whole, rather than taking the specific situation of each cell into account. However, our definition of the shadow cluster implicitly accounts for factors such as network topology, cell utilisation and current network load for each cell, but perhaps a finer local tuning is possible. In particular, the use of the "movement probability" is somewhat simplistic and needs to be traded off against the call termination probability in a more refined way. Changes in the relative performance in the presence of other VBR and Best Effort Traffic also need to be investigated.

7 References

[1] D. A. Levine, I. F. Akyildiz & M. Naghshineh, "A Resource Estimation and Call Adaptation Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept", IEEE/ACM Trans. Networking, Feb. 1997, Volume 5 Number 1, pp.1-12.

[2] Sunghyun Choi & Kang G. Shin. "Predictive and Adaptive Bandwidth Reservation for Hand-Offs in QoS-Sensitive Cellular Networks", ACM SIGCOMM 1998, pp.155-166.

[3] P. Ramanathan, K.M. Sivalingam, P.Agrawal, and S.Kishore, "Dynamic Resource Allocation Schemes During Handoff for Mobile Multimedia Wireless Networks", IEEE JSAC, July 1999, Vol 17 No. 7, pp.1270-1283.

[4] Tarjan, RE 1981: A unified approach to path problems, JACM 28, pp.577-593

[5] Mehlhorn, Kurt, 1977, Data structures and Algorithms 2 – Graph Algorithms and NP-Completeness. ETACS Monographs on Theoretical Computer Science, Springer-Verlag, Berlin].

[6] St. Näher. LEDA Manual. Technical Report, Max-Planck-Institut für Informatik, 1993.

[7] Kevin Brown & Suresh Singh, "M-TCP: TCP for Mobile Cellular Networks", ACM SIGCOMM, Vol. 27, No. 5, October '97, pp19-43.

[8] O.T.W. Yu and V.C.M. Leung, "Adaptive resource allocation for prioritised call admission over an ATM-based wireless PCN", IEEE J. Selected Areas of Communication, vol. 15, pp 1208-1225, Sept. 1997.