Simple admission control schemes supporting QoS in wireless multimedia networks

Jae Young Lee and Saewoong Bahk

The authors present a simple cell-oriented adaptive admission control scheme for keeping the handoff dropping probability below a target value and compare its performance with that of a representative mobile-oriented adaptive admission control scheme.

Introduction: To maintain the handoff dropping probability within a specified level (thus providing a probabilistic QoS guarantee), various admission control schemes have been developed [1–4]. The authors of [1, 3] assumed an exponentially distributed cell residence time, which may not be true [5]. Otherwise, they are based on user mobility information, which is impractical and called mobile-oriented. Here we propose a cell-oriented adaptive admission control scheme based on the handoff dropping events in each cell. We also compare our proposed scheme (PROP) with the representative mobile-oriented scheme in [4] (CS98).

System model and admission control: We consider a cellular infrastructure with a cell i having capacity C(i). The service model accommodates multiple classes of traffic (e.g. voice and video). Let BU denote the bandwidth unit.

Admission control test: To give priority to handoff calls, we adopt the well-known reservation method, i.e. at cell i, a new call is accepted if C(i) + B ≤ T(i), while a handoff call is accepted if C(i) + B ≤ T(i). The allocated bandwidth of cell i is B, i.e. the required bandwidth of the call, and T(i) ≤ C(i) is the admission threshold of cell i. C(i) – T(i) can be interpreted as the reserved bandwidth for handoff calls at cell i.

Adaptive control algorithm: There might exist an optimal steady-state admission threshold T_{adm}(i) at each cell i for a specific traffic load and user mobility [Note 1]. Here we use the term ‘optimal’ in the sense of maximising (minimising) utilisation (P_U) while keeping P_D below a target value P_{QoS}. The problem is how to adjust T as close as possible to, but not over, T_{adm}. First, we describe an adaptive algorithm to adjust T based on monitored handoff drops at each cell. The algorithm executed by the base station of each cell in a distributed manner is shown below.

1. S_P = 1, P_{QoS} = 0.
2. S_D = 0, P_{QoS} = 0; L_HD = 0; L_SM = 0; T = T_{min}.
3. WHILE (time increases)
   * IF(receive decrease_T message) THEN
   * IF(T > avg_T of adjacent cells) THEN
     T = max(T – d, T_{min}).
   * IF(receive increase_T message) THEN
   * IF(T < avg_T of adjacent cells and QoS_state == IN) THEN
     T = min(T + d, T_{max}).
10. IF(a mobile handoff into the current cell) THEN
    S_D = S_D + 1; L_HD = L_HD + 1.
12. IF(it is dropped) THEN
    S_D = S_D + 1; L_SM = L_SM + 1.
13. IF(L_HD > 1) THEN
    L_P = L_P + S_P.
16. T = max(T - d, T_{min}).
* send decrease_T messages to adjacent BSs;
* IF(S_D < S_P) THEN
   * IF(S_D < S_P) THEN
     T = min(T + d, T_{min}).
* send increase_T messages to adjacent BSs;
   * IF(S_D < S_P) THEN
     L_P = L_P + S_P.
* IF(S_D > S_P) THEN
   * IF(S_D < S_P) THEN
     L_SM = L_SM + 1.

The algorithm consists of two parts. First, the lines without * adjust T to keep the P_D in each cell below P_{QoS}. The main idea is to monitor handoff dropping events over both the short term and long term. The short-term SP is given by the number of handoff attempts [1/P_{QoS}], not by the length of time. The counts for the short-term handoff attempts S_D and handoff drops S_{HD} are reset to 0 at the start of each period. The long-term period L_P is determined by handoff attempts as S_P ☓ max(L_{HD}, 1) where L_{HD} is the count of the long-term handoff drops. The counts for the long-term handoff attempts L_D and handoff drops L_{QoS} are reset to 0 at the start of each long-term period. The objective of long-term monitoring is to keep P_D below P_{QoS}, whereas the short-term monitoring is used to maximise utilisation. Note that the dropping probability for each long-term period is kept below the target value:

\[ P_D = \frac{L_{HD}}{L_P} = \frac{L_{HD}}{S_P \times \text{max}(L_{HD}, 1)} \leq \frac{1}{S_P} \approx P_{QoS} \] (1)

Secondly, the lines with * denote the balancing between cells. When a BS dynamically adjusts T regardless of the state of its adjacent cells, a significant imbalance of thresholds among adjacent cells can occur, which can adversely impact performance [5]. The main idea to overcome this problem is to balance the thresholds. When the BS of cell i decreases T, it sends decrease_T messages to the BSs of cell i’s adjacent cells (A). The threshold of an adjacent cell j (\( j \in A \)) thus also decreases if it is higher than the average T of A. Similarly, when T is increased, the threshold of certain adjacent cells also increases. The latter condition in line 8 is included so that T should be increased only if the long-term QoS is satisfied (i.e. QoS_state is IN).

Fig. 1 Probability comparison of PROP and CS98

* a Probabilities
  b Utilisation
  PROP, F_1 = 1.0
  ----------- PROP, F_1 = 0.5
  ----------- CS98, F_1 = 1.0
  ----------- CS98, F_1 = 0.5

Fig. 2 Complexity comparison of PROP and CS98

* a Computational complexity
  b Signalling messages
  PROP
  ----------- CS98

Comparative performance evaluation: Simulations were performed on 19 hexagonal cells with wrap-around structures. The arrival process of new call requests was Poisson with rate \( \lambda \) (calls/s/cell). A new call was either for voice (1 BU) or video (4 BU) with probability \( F_1 \) and \( 1 - F_1 \), respectively. The velocity of a mobile was randomly selected from [80, 120] (km/h) and its direction of movement was also randomly selected. The cell duration was exponentially distributed with mean \( 1/120 \) s. The capacity of each cell was \( C = 100 \) BUs, and the cell diameter was 1 km. The other simulation parameters were \( T_{min} = T_{max} = 100 \) BUs, \( T_{max} = 0 \) (BU), \( P_{QoS} = 0.01 \) and step size \( d = 1F_1 + 4(1 - F_1) \). The offered load per cell, \( L \), was calculated by \( L = \lambda(1 + d + (1 - F_1)) \lambda^{-1} / C \).

We first compared the performance of PROP with that of CS98. Fig. 1 shows \( P_D \) and the utilisation against offered load with high mobility for \( F_1 = 1.0 \) and 0.5. Both satisfy the QoS constraint of keeping \( P_D \) below 0.01. They show higher utilisation for \( F_1 = 0.5 \) than for \( F_1 = 1.0 \), since the more video calls existed, the more bandwidth was needed. The utilisation in PROP was higher than that in CS98 for each value of \( F_1 \) [Note 2].

We then compared the complexity of the two schemes. First, we compared the computational complexity for an admission decision. The complexity of CS98 with respect to an admission decision depends on the number of operations per admission decision:

\[ \text{Complexity} = (1F_1 + 4(1 - F_1)) \lambda^{-1} / C \]

Note 1: We will drop the index i for notational simplicity when the reference is clear.

Note 2: We also simulated the low mobility of [40, 60] (km/h) range and time-varying traffic/mobility cases. They showed similar tendencies, and are omitted due to space constraints.

ELECTRONICS LETTERS 24th May 2001 Vol. 37 No. 11