Computer Communications 31 (2008) 2428–2434

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Dual-threshold admission control for non-real-time traffic in wireless data networks

Show-Shiow Tzeng *, Hsin-Yi Lu

Department of Optoelectronics and Communication Engineering, National Kaohsiung Normal University, No. 62, Shenjhong Road, Kaohsiung 802, Taiwan, ROC

ARTICLE INFO

Article history:
Received 6 July 2007
Received in revised form 10 March 2008
Accepted 10 March 2008
Available online 17 March 2008

Keywords:
Admission control
Non-real-time services
Wireless data networks
Quality-of-service
Radio resource management

ABSTRACT

Non-real-time services are an important category of network services in future wireless networks. When mobile users access non-real-time services, mobile users are usually concerned with the total time to complete their data transfer. In addition, system providers expect that channel utilization should be as high as possible because radio channels are limited. Therefore, it is crucial to design a system that satisfies the total completion time requirement of mobile users while maximizes channel utilization. In this paper, we design an optimal call admission control scheme that uses two thresholds; one threshold is used to determine whether or not to accept a new call arrival into a cell, and the other threshold is used to limit the total number of calls in a cell. An analysis is developed to evaluate the performance of the proposed call admission control scheme. Numerical results show that the proposed scheme not only satisfies mobile users' quality-of-service requirements (in terms of the total completion time and hand-off dropping probability) but also maximizes channel utilization.

1. Introduction

Future wireless networks are expected to provide various non-real-time services, such as file transfer, web browsing, data retrieval, etc. When mobile users access non-real-time services, mobile users usually care about the total time to complete their data transfer. Therefore, from the perspective of mobile users, a wireless system with good quality-of-services is a system that is designed such that the total completion time for transferring non-real-time data is as less as possible.

Two major factors determine the total completion time to transfer a non-real-time data; one is the size of the non-real-time data, and the other is the data rate to transfer the non-real-time data. Given a fixed size of non-real-time data, higher data transfer rate will produce shorter transfer time. One possible way to increase data transfer rate is to employ more radio channels in transferring the data. However, system providers usually expect to serve mobile users as many as possible to fully utilize limited radio channels, which results in that fewer channels are allocated to a mobile user. This phenomenon of fewer channels further leads to lower data transfer rate and longer transfer time. Therefore, there is a trade-off between the total transfer time and channel utilization in a wireless system. From the viewpoint of system designers, it is crucial to design a system that satisfies the completion time requirement of mobile users while maximizes channel utilization in a system. A number of papers have studied radio channel allocation for various data, including non-real-time data [1–4]. The channel allocation schemes in [1–4] attempt to effectively utilize limited radio channels. A number of papers [1,2] also take into account the hand-off dropping probability or forced termination probability, both of which are important quality-of-service metrics. The hand-off dropping probability is the probability that a mobile user is dropped when the mobile user attempts to hand-off; the forced termination probability is the probability that a mobile user is forced to terminate its connection during its connection period. All the papers [1–4] allocate a fixed number of channels (or bandwidth) to a mobile user regardless of system load; that is, a mobile user employs a fixed number of channels during the mobile user's connection period, and the number of channels allocated to the mobile user is constant no matter whether the system load is in a light situation or heavy situation. Instead of using constant bandwidth, a number of papers [5–9] employ dynamic number of channels to transfer non-real-time data. That is, when a system load decreases, a mobile user can use more channels to increase data transfer rate and shorten data transfer time. When the system load increases, a mobile user will employ fewer channels to accommodate more calls instead of rejecting calls. All the aforementioned papers [1–9] do not consider the total completion time that a mobile user can accept at a call level.

As mentioned earlier, mobile users usually concern the total completion time for transferring data. It is feasible for wireless systems to dynamically allocate channels to a mobile user according to system load, as long as the total completion time satisfies the mobile user's requirement. In such a system that allocates dynamic number of channels to a mobile user, the mobile user will obtain more channels in a light load than in a heavy load. Since system providers also consider pricing and other system requirements,

* Corresponding author. Tel.: +886 7 7172930 7715.
E-mail address: sstzeng@nknucc.nknu.edu.tw (S.-S. Tzeng).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.03.010

the maximum number of channels allocated to a mobile user is usually below a limitation and the minimum number of channels allocated to a mobile user is usually above a certain value. Although previous literatures in [5–9] allow wireless systems to dynamically allocate channels to a mobile user, these literatures do not discuss how to keep the total completion time within a certain level that mobile users can accept at a connection level. In other words, when the existing schemes in [5–9] are directly applied into non-real-time traffic in wireless systems, the mean of the total completion time may be intolerable long for mobile users. To guarantee the mean of the total completion time below a certain level, we require a novel call admission control for non-real-time data in wireless data networks.

To keep the mean of the total completion time under a certain level, the first step is to quantify the transfer time that mobile users can accept in degree. This paper employs a measure, called stretch factor, to quantify the total completion time. The concept of the stretch factor has been introduced and used in [10–15]. Similar to the concept of stretch factor in [10–15], we define the stretch factor for transferring non-real-time data as the ratio of the total time in transferring the data on a loaded system which allocates minimum number of channels to a mobile user to the total time in transferring the data on a light-load system which allocates the maximum number of channels to a mobile user. The stretch factor is a useful metric to measure the total completion time that mobile users require.

Based on the stretch factor, this paper proposes a novel call admission control for non-real-time data in wireless data networks. The objective of the proposed admission scheme aims to satisfy the total completion time requirement of mobile users while maximizes channel utilization. The proposed admission control scheme employs two thresholds, one of which is used to determine whether or not to accept a new call arrival into a cell, and the other of which is used to limit the total number of calls in a cell. A numerical analysis is developed to find the optimal values for the two thresholds in the system. Numerical results show that the proposed call admission control scheme provides good quality-of-service in terms of the total completion time and hand-off dropping probability while fully utilizes limited radio channels.

The rest of this paper is organized as follows. Section 2 describes the system model. Then, call admission and channel allocation procedures are explained in Section 3. Section 4 describes the performance analysis of the proposed admission control scheme and threshold selection procedure. Subsequently, the numerical results are described in Section 5. Finally, some concluding remarks are presented in Section 6.

2. Description of system model

In this paper, we consider a two-dimensional space that consists of hexagonal cells. Fixed channel assignment is used to allocate a fixed set of radio channels to each cell [16]. Various network services, such as real-time services, non-real-time services, etc., can be provided in wireless systems. In this paper, we discuss call admission control on non-real-time traffic; therefore, we focus on non-real-time services. The total number of channels allocated to non-real-time services in a cell is denoted by C. A mobile user in a cell can employ different number of channels to access non-real-time services; besides, the mobile user can adjust their channel requirement according to the traffic load in the cell. The minimum channel requirement of mobile users is denoted by $c_{\min}$; the maximum channel requirement of mobile users is denoted by $c_{\max}$. The possible channel requirement of mobile users ranges from $c_{\min}$ to $c_{\max}$.

New mobile users arrive at a cell according to a Poisson process with mean rate λ_n. All cells have the same arrival rate. In the previous literatures [17–21], new arrivals on a call level are usually assumed to be Poisson arrival. On the contrary, the literature [21] describes that the arrival of data packet on a packet level is not Poisson arrival and exhibits the property of long-range dependence. However, this paper focuses on call admission control on a call level. Therefore, it is reasonable to use Poisson arrival to model new call arrivals. When a mobile user arrives at a cell, a call admission control procedure is initiated to determine whether or not to accept the mobile user; once the mobile user is accepted, a channel allocation procedure is further involved to allocate channels between accepted mobile users. The probability that new mobile users are blocked is called new call blocking probability. The new call blocking probability of mobile users is denoted by P_b. The call admission control and channel allocation procedures used in this paper will be described in Section 3 in detail.

Once a mobile user is admitted into a cell, the number of radio channels allocated to the mobile user is in the range between $c_{\min}$ and $c_{\max}$. When an admitted mobile user uses the maximum number of channels $c_{\max}$ to transfer data during its entire connection period, the lifetime that the mobile user experiences is referred to as minimum lifetime and is assumed to be exponentially distributed with mean $1/\mu_n$. The duration that a mobile user sojourns in a cell is exponentially distributed with mean $1/\mu_h$. When a hand-off occurs, the mobile user hand-offs to each of the adjacent cells with equal probability. The probability that hand-off users are dropped is called hand-off dropping probability. The hand-off dropping probability of mobile users is denoted by P_d. When a mobile user departs from a cell, the channels used by the mobile user will be released and a channel allocation procedure stated in Section 3 is involved to re-allocate channels between those mobile users still sojourning in the cell.

3. Call admission and channel allocation procedures

In this section, we first present call admission procedures for mobile users. Then, we introduce a procedure that allocates channels between accepted mobile users.

Mobile users are differentiated into new calls and hand-off calls. Different admission procedures are initiated when new calls and hand-off calls arrive at a cell. When a new mobile user arrives in a cell, a new call admission control procedure is initiated. If the number of active mobile users in the cell is greater than or equal to a threshold t, then the new call is blocked; otherwise, the new call is admitted. Since the threshold t is used to determine whether or not a new call can perform a call setup to establish a new connection, we refer to the threshold t as "call-setup threshold." When a mobile user moves from one cell to an adjacent cell, a hand-off call admission procedure is initiated to maintain the mobile user's communication. The hand-off procedure deals with the hand-off call according to the number of mobile users in the adjacent cell, which can be classified into two cases. In the case that the number of mobile users in the adjacent cell is less than the maximum number of mobile users that can be accommodated in a cell, N, the hand-off call continues its communication. Otherwise, the hand-off call is forced to terminate. Since the N denotes the maximum number of mobile users that can be simultaneously served in a cell, we refer to the N as "capacity threshold." Obviously, it is meaningless that the values of capacity threshold and call-setup threshold are less than or equal to zero. In addition, the possible value of capacity threshold is less than or equal to $\lfloor C/c_{\min} \rfloor$. To achieve new call admission control, the value of call-setup threshold should be less than or equal to the value of capacity threshold. It is crucial to determine the values of the capacity threshold and call-setup threshold such that the