

Analytical Model for Admission Control in Next Generation Wireless Systems

H. Birkan Yilmaz
Dept of Computer Engineering
Bogazici University
Bebek, Istanbul, Turkey 34342
Email: yilmhuse@boun.edu.tr

Tuna Tugcu
Dept of Computer Engineering
Bogazici University
Bebek, Istanbul, Turkey 34342
Email: tugcu@boun.edu.tr

Feodor Vainstein
Georgia Tech Savannah
School of Electrical & Computer Eng.
Georgia Institute of Technology
Savannah, Georgia 31407
Email: fvainste@gtsav.gatech.edu

Abstract—The heterogeneous structure of *Next Generation Wireless Systems (NGWS)* makes admission control very complex. Accessibility of the subsystems at the time of connection or handoff request, availability of resources in the subsystems, user preferences, and connection class need to be considered in admission control. In this paper, we first give a general connection admission control algorithm. We also propose the first analytical model in the literature for admission control in NGWS. We point out the major challenges in modeling for NGWS and propose a neat solution to calculate state probabilities in a reasonable way even as the state space proliferates.

I. INTRODUCTION

Next Generation Wireless Systems (NGWS) merge multiple subsystems with different access technologies to provide high bandwidth access anytime, anywhere. Though existing PCS, satellite, and WLAN systems are also designed with these objectives in mind, they fail to satisfy all of the requirements *simultaneously* due to constraints like global coverage, indoor/outdoor communications, and frequent handoffs. Unfortunately, it is not likely that emerging new technologies will be able to provide continuous coverage, either. Therefore, NGWS will utilize multiple wireless systems and the new technologies to come as *subsystems* in order to provide high bandwidth access everywhere. PCS, WLAN, satellite, and new wireless systems like 4G Mobile are examples of the subsystems in NGWS. NGWS will have the following basic properties:

- *end-to-end* packet-based service, including the air interface,
- support for voice, multimedia, and data traffic with QoS provisioning,
- backbone traffic of NGWS carried over the Internet.

In [1], [2], a cell design methodologies for the optimal design of multitier cellular systems are proposed. The use of intelligent networks for mobility and internetworking is discussed in [3]. In [4], Zeng *et al.* discuss emerging technological trends in PCS systems. Key issues about IMT-2000 are discussed from the perspective of the service provider in [5]. Integration of cordless and cellular systems in 3G systems is discussed in [6].

In the literature, there are several proposals for the NGWS architecture under different names including 4G, All-IP, Beyond-3G, and Beyond-UMTS. In [7] and [8], different

NGWS architectures are described with a focus on the possible driving forces for the deployment of these networks. NTT DoCoMo proposes an NGWS architecture, which is an extension of current 3G architecture [9]. A next generation wireless communication architecture that is comprised of old and new wireless communication standards has been presented in [10]. Telefonica's NGWS architecture is composed of WLANs, cellular networks, personal area networks, and distribution networks organized in a layered structure [11]. In the Wine Glass Project, [12], WLANs and UMTS are merged into a next generation wireless network. Furthermore, Siemens [13] adopts 3GPP's IP Based Multimedia System (IMS) specifications [14] and defines its own next generation wireless network architecture. Integrating multiple subsystems into one NGWS brings many challenges ranging from interworking among inherently different wireless subsystems to QoS provisioning [12], [15]. The selection of the subsystem for connection establishment and handoff is a key factor in the performance of NGWS. The Global Mobile Broadband System (GMBS) described in [16] aims to include GPRS, UMTS, WLAN, and satellite-based systems. Although certain characteristics of these proposals are common, they are independent of each other in terms of architecture and operation.

In this paper, firstly we give a general connection admission control scheme for NGWS. Our scheme is universal in the sense that it does not depend on the number and type of subsystems included in NGWS. Then, we present an analytical model for the admission control scheme. This is the first analytical model for NGWS in the literature to the best of our knowledge. The complexity of NGWS also creates significant challenges, especially in terms of the size of the state space. We demonstrate these challenges, and develop a neat solution that calculates state probabilities for critical and subcritical states. We provide some preliminary results from our tests. Our aim in this paper is to develop an analytical model for NGWS and demonstrate how the challenges in the size of the state space can be overcome rather than evaluate the performance of a specific algorithm.

The remainder of this paper is organized as follows. In Section II, the NGWS network architecture is defined. The connection admission control scheme is detailed in Section III. The analytical model of connection admission control is

given in Section IV. The numerical results are explained in Section V. Finally, Section VI concludes this paper.

II. NETWORK ARCHITECTURE

NGWS will be composed of multiple subsystems of different types. Each subsystem will serve as an access network to the users. Since the service areas of the subsystems overlap, the mobile terminals, MT , will have access to multiple subsystems simultaneously. WLANs, PCS, satellite systems, and their future variations together with new wireless systems like 4G Mobile are candidates as subsystems. Our NGWS architecture is flexible and allows all or a subset of these subsystems to be a part of NGWS.

We define NGWS as the tuple

$$\mathcal{NGWS} = (\mathcal{S}, HR) \quad (1)$$

where $\mathcal{S} = \{wl, pcs, sa, 4g, \dots\}$ is the set of subsystems, and HR is the global home register for all subsystems $s \in \mathcal{S}$. In Equation 1, wl , pcs , sa , and $4g$ correspond to WLAN, PCS, Satellite, and 4G Mobile subsystems, respectively. Note that the set \mathcal{S} can be expanded as needed to include other types of wireless subsystems. Such additions will not affect our admission control scheme.

Each subsystem has its own cellular infrastructure. For each subsystem, numerous access nodes are deployed to cover the service area, sometimes partially. The access node is the base station in PCS subsystem, access point in WLAN, transponder in satellite subsystem, etc. We denote the i^{th} access node of subsystems s as b_i^s . For a given access node b_i^s , we define its cell as the set of locations from which it is possible to communicate with that access node. Thus, each subsystem s splits the service area into cells. Clearly, there is a one-to-one correspondence between the cells and access nodes. We use the term access node for the *device* that provides access for MT s and cell for the set of locations served by the same access node. The size, shape, and location of the cells depend on the location and power of the access node, and the terrain. Therefore, the cellular layouts of the subsystems differ.

We state above that NGWS will provide various types of services such as voice, video, and multiple types of data. We denote the set of connection classes as

$$\mathcal{C} = \{\text{voice, video, low bandwidth data, high bandwidth data, } \dots\}.$$

Each type of connection class has different requirements ranging from bandwidth to end-to-end latency. For the sake of simplicity, we consider only the bandwidth requirement and assume that reasonable values for the other requirements can be achieved if enough bandwidth is provided. We denote the bandwidth requirement of a class k connection as $bw(k)$.

The overlaid structure of NGWS implies that MT has access to multiple subsystems simultaneously. MT knows the set of subsystems it can access by scanning the pilot signals from the access nodes. The selection of the subsystem s for connection establishment depends on several factors:

- **Subsystem accessibility:** The pilot signal of subsystem s must be strong enough for communication.
- **Resource availability:** The load of access node b_i^s , l_i^s , must not exceed the capacity of b_i^s , c_i^s , if the request is admitted.
- **Service class and user preferences:** The user is able to indicate which subsystem he¹ prefers for each connection class. We denote the probability that MT prefers subsystem s for a connection of class k as $p(k, s)$ where $k \in \mathcal{C}$.

In NGWS, there is a *home register* HR that serves all subsystems in mobility and connection management. The home register resides outside the subsystems, as a node in the Internet (Figure 1). The function of HR is to store static and dynamic information about all registered users. HR has a different interface for each type of subsystem and acts as the home agent for WLAN networks, as the HLR for the PCS networks, etc.

An example configuration of NGWS consisting of satellite, PCS, and WLAN subsystems is depicted in Figure 1. The service area is covered with overlapping cells of different subsystems. The coverage area of the subsystems may be discontinuous, as in the case of WLAN. Thus, the set of subsystems to which a mobile terminal can access at a given moment varies. Each subsystem has its own local register, LR , and the backbone traffic between the subsystems is carried over the Internet. The global HR serves all subsystems.

In current wireless systems, the central location register is queried for the user information with the user's specific address as the index. The mobile phone number in PCS systems, and the home address in Mobile IPv6 are used for addressing the user in the corresponding systems. In NGWS, the information about a user can be retrieved from the home register by using the user's *next generation user address* ($NGUA$) as the index. $NGUA$ is the only address visible to the human users, independent of the subsystem in which the mobile terminal resides. It is the responsibility of HR to translate the $NGUA$ specified by the caller party to the address (e.g., phone number, home IP address, etc.) in the relative subsystem.

III. NEXT GENERATION CONNECTION ADMISSION CONTROL (NGCAC) SCHEME

In a wireless network, it is the connection admission control scheme that decides whether connection and handoff requests will be accepted. For new connection requests, NGWS must select the appropriate subsystem for connection establishment. Furthermore, the mobility of a user may require changes in the use of wireless resources, resulting in a handoff attempt for the user. It is the duty of the *Next Generation Connection Admission Control* ($NGCAC$) scheme to manage the connection requests and handoff attempts in a way that maximizes network utilization, minimizes outage, and distributes the load between subsystems.

¹Throughout this paper, "he" should be read as "he or she", "his" as "his or her", etc.

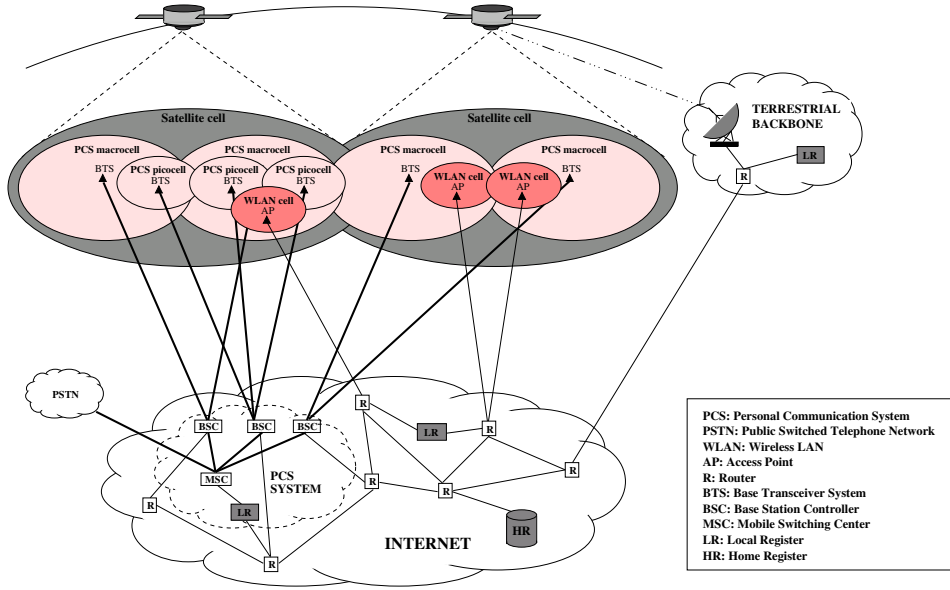


Fig. 1. An example interconnection of subsystems in NGWS

The connection admission control scheme is triggered in three cases:

- Outgoing connection request: When a user initiates a connection request.
- Incoming connection request: When a remote (mobile or fixed) user initiates a connection request destined at the mobile user.
- Handoff request: When a mobile user with an active connection crosses a cell boundary.

A. Problem Statement and Proposed Solution

Since MT has access to multiple subsystems simultaneously, NGWS must select one of the subsystems for connection. Among the accessible subsystems, one subsystem that can accommodate the connection request will be selected, subject to connection class and user preferences.

We define the vicinity of b_i^s as

$$\{b_j^t \mid t \neq s \text{ and } b_i^s \cap b_j^t \neq \emptyset\},$$

and call all access nodes in the vicinity as neighbors. Each access node b_i^s periodically transmits its load information l_i^s to all of its neighbors, and also keeps record of their loads. We denote the recorded value of l_j^t at access node b_i^s as l_j^t . Since the load information is exchanged between only a few access nodes in the vicinity, and the information exchange is performed over abundant wired links, this overhead is negligible. l_j^t may not be exactly up-to-date, but since load in a cell does not fluctuate wildly, l_j^t will be reasonably close to l_j^t . We denote the new load of b_i^s if request rq is accepted as $\hat{l}_i^s(rq)$. We also denote the load of b_j^t , based on the recorded value l_j^t , if rq is accepted as $\hat{l}_j^t(rq)$.

With each connection or handoff request rq , we associate an *ordered list* of accessible access nodes in which ordering criteria is the user's preferences for the connection class of rq . We denote the ordered list of access nodes specified in request rq as $\mathcal{L}_{ac}(rq)$. For outgoing connection setup and handoff requests, MT sends the request to the first access node, b_i^s , in $\mathcal{L}_{ac}(rq)$. However, for incoming connections the initiator (caller) is a remote node that is not aware of the subsystems accessible by MT , availability of the resources in the subsystems, and user preferences for MT . Furthermore, the paging process, which precedes connection establishment, need not be done through the subsystem over which the connection will be established. Therefore, we propose that in the paging reply message, MT specifies $\mathcal{L}_{ac}(rq)$ to be used in connection admission. Then, connection establishment is performed over the first access node, b_i^s , in $\mathcal{L}_{ac}(rq)$. Since $\mathcal{L}_{ac}(rq)$ contains the identifiers of a few access nodes, its overhead is negligible.

When b_i^s receives request rq , either directly from MT or from a remote node, it checks if the request can be accommodated. If $\hat{l}_i^s(rq)$ remains below capacity c_i^s , b_i^s accepts the request, establishes the connection, and makes the necessary resource allocations. On the other hand, if $\hat{l}_i^s(rq)$ exceeds c_i^s , b_i^s contacts, *on behalf of* MT , all access nodes b_j^t in $\mathcal{L}_{ac}(rq)$ for which $\hat{l}_j^t(rq) \leq c_j^t$. If there exists an access node b_j^t in $\mathcal{L}_{ac}(rq)$ that can accommodate request rq , the request will be transferred to b_j^t , and the connection will be established over that access node. On the other hand, if $\hat{l}_j^t(rq) > c_j^t$ for all b_j^t in $\mathcal{L}_{ac}(rq)$, request rq will be rejected. In this case, MT may either call off the request or revise the connection class (connection requirements) and resubmit it as a new request.

The algorithm for the proposed scheme is presented as a flowchart Figure 2.

An example scenario is depicted in Figure 3 with three subsystems, *sa*, *pcs*, and *wl*. Although the service area is covered by three subsystems, there are *holes* where no signal is received from one or more subsystems. Let's assume *MT*'s subsystem preference decreases in the order *wl*, *pcs*, and *sa*. *MT* initiates a connection at point *A* where it has access to all subsystems, and selects *wl* according to user preferences. As it moves on its trajectory, it encounters an intra-subsystem handoff between two WLAN cells at point *B*. At point *C*, *MT* enters a hole in the WLAN cell where access to *wl* is lost. Therefore, *MT* encounters an inter-subsystem handoff at point *C*, from *wl* to *pcs*, preferring *pcs* over *sa*. *MT* gets out of the hole at point *D*, but it will keep communicating over *pcs* until it reaches point *F*. At point *F*, *MT* leaves the PCS cell, so it will encounter another inter-subsystem handoff from *pcs* to *sa*. Finally at point *G*, *MT* successfully completes its connection.

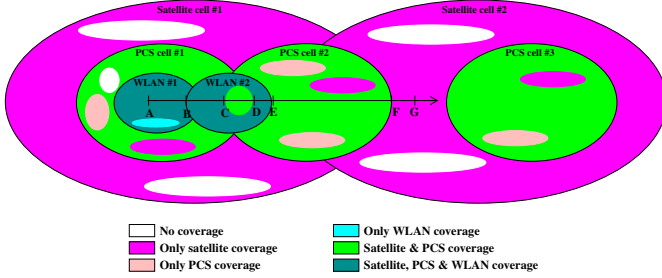


Fig. 3. Connection admission scenario

IV. ANALYTICAL MODEL OF NGCAC SCHEME

In the analysis of wireless systems, the service area is typically split into cells since the partitioning criteria is the access node that controls the area. However, in the case of NGWS, cellular granularity is too coarse to define a partition since there are multiple subsystems serving the same service area. Therefore, in our model, the service area is partitioned into smaller regions we call physical areas.

A. System Definition

Let \mathcal{B} be the set of all access nodes in all subsystems in the service area Ω . We denote the set of all subsets of \mathcal{B} as $\mathcal{A} = 2^{\mathcal{B}}$. We define $\tilde{b} : \Omega \rightarrow \mathcal{A}$ such that $\tilde{b}(x)$ is the set of all access nodes reachable from location x . This mapping introduces an equivalence relation \sim on Ω such that $x_1 \sim x_2$ if $\tilde{b}(x_1) = \tilde{b}(x_2)$.

We call the elements of \mathcal{A} as *areas*. For every area $a \in \mathcal{A}$, we define a physical area $\bar{a} \subset \Omega$ as

$$\bar{a} = \{x \in \Omega \mid \tilde{b}(x) = a\}.$$

It is clear that physical areas are the blocks of the partition induced by " \sim ". Furthermore, since there exist some $a \in \mathcal{A}$

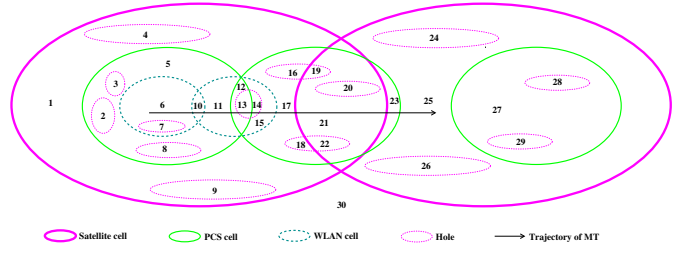


Fig. 4. An example partitioning of the service area

for which $\bar{a} = \emptyset$, it can be shown that if $\bar{a}_1 = \bar{a}_2 \neq \emptyset$, then $a_1 = a_2$.

We find it more convenient to work with areas rather than with physical areas since it reduces the complexity of the description of the model. We call the areas for which corresponding physical areas are non-empty as *effective areas*. For obvious reasons, we are interested only in the effective areas. The upper bound on the number of effective areas is given by $\binom{|\mathcal{B}|}{K}$ where K is the maximum number of access nodes reachable from one location.² The partitioning of the service area and the relationship between areas and cells according to the coverage in Figure 3 is shown in Figure 4. The number of area crossings on the trajectory of *MT* provides a reasonable explanation for not initiating a handoff procedure to the most preferred access node at every area boundary.

For each area a_i , we have the following:

- $n_{a_i}(t)$: Number of users in area a_i at time t .
- $V_{a_j, a_i}^k(t)$: Migration rate of a class k connection from area a_j to a_i at time t . Hence, the number of class k migrating from a_j to a_i during the time period τ is equal to

$$\int_t^{t+\tau} n_{a_j} \cdot V_{a_j, a_i}^k(t) dt.$$

- $r_{a_i}^k(t)$: Connection generation rate of class k connections in area a_i at time t .
- $f(a_i, k, t)$: Connection profile of class k connections in area a_i at time t , i.e., the distribution of class k such as 0.60 for voice, 0.10 for multimedia, 0.30 for data.
- $p(a_i, k, s, t)$: Probability that a user in area a_i with a class k connection prefers subsystem SS^s at time t .

We denote the number of active connections of class k communication with access node b in area a as $x_a^k(b)$. The state of the system is any particular distribution of values taken by the variables $x_a^k(b)$. Thus, the state of the system is the

²From this point on, we will use the term area in the sense of effective areas.

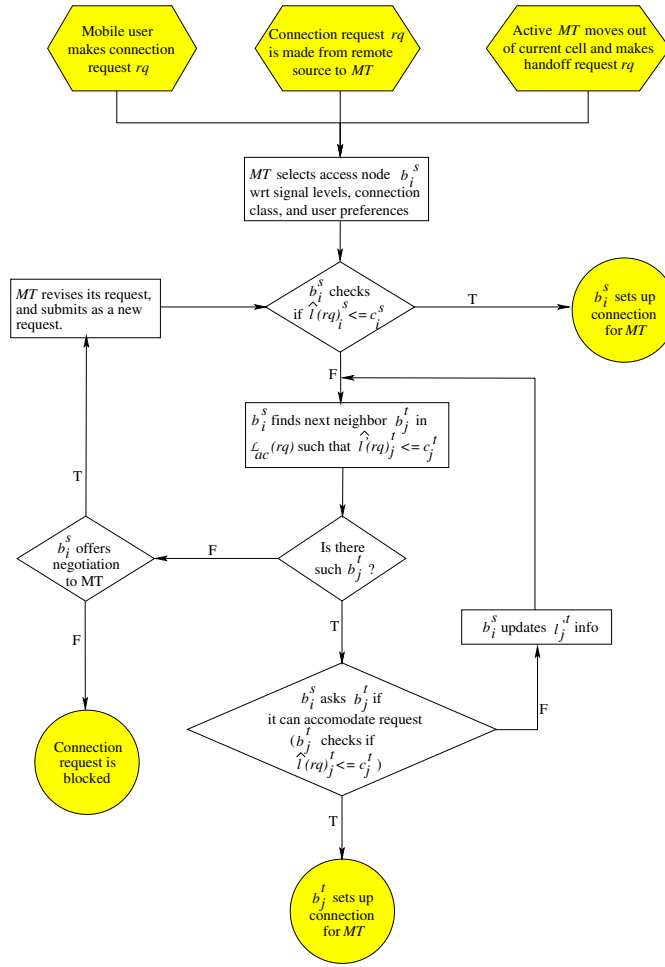


Fig. 2. The algorithm of the connection admission control scheme

tuple

$$g = \begin{bmatrix} x_1^1(1), & x_1^1(2), & \dots, & x_1^1(|\mathcal{B}|) \\ x_2^1(1), & x_2^1(2), & \dots, & x_2^1(|\mathcal{B}|) \\ \dots & & & \\ x_1^{|\mathcal{C}|}(1), & x_1^{|\mathcal{C}|}(2), & \dots, & x_1^{|\mathcal{C}|}(|\mathcal{B}|) \\ x_2^{|\mathcal{C}|}(1), & x_2^{|\mathcal{C}|}(2), & \dots, & x_2^{|\mathcal{C}|}(|\mathcal{B}|) \\ \dots & & & \\ x_2^{|\mathcal{C}|}(1), & x_2^{|\mathcal{C}|}(2), & \dots, & x_2^{|\mathcal{C}|}(|\mathcal{B}|) \\ \vdots & & & \\ \vdots & & & \\ x_{|\mathcal{A}|}^1(1), & x_{|\mathcal{A}|}^1(2), & \dots, & x_{|\mathcal{A}|}^1(|\mathcal{B}|) \\ \dots & & & \\ x_{|\mathcal{A}|}^{|\mathcal{C}|}(1), & x_{|\mathcal{A}|}^{|\mathcal{C}|}(2), & \dots, & x_{|\mathcal{A}|}^{|\mathcal{C}|}(|\mathcal{B}|) \end{bmatrix}. \quad (2)$$

The state can also be represented as a mapping

$$g : \mathcal{X} \rightarrow \mathbb{N} \cup \{0\} \quad (3)$$

where $\mathcal{X} = \{x_a^k(b)\}$. Thus, a state is any particular distribution of values taken by individual $x_a^k(b)$. We denote the set of all states as \mathcal{G} .

Although each variable $x_a^k(b)$ can assume values up to the capacity of corresponding access node b_i^s , we have

$$\sum_{a \in \mathcal{A}} \sum_{k \in \mathcal{C}} w(k) \cdot x_a^k \leq c_i^s$$

since the same wireless resources are shared by all areas covered by the same cell. Therefore, many states in the state space are never visited.

B. Elementary Events

There are various elementary events that cause the system to switch from one state to another. Keeping in mind that a state of the system is a mapping as given by Eq. 3, each elementary event causes a change in the mapping at one or two points. For example, let's assume that the current state of the system is defined by the mapping g_i , and access node b accepts a new connection request of class k from area a . This new connection causes $g_i(x_a^k(b))$ to be incremented by one, defining a new mapping g_j . In Figure 5, an example in which $x_5^1(2)$ is incremented due to a new connection. Two mappings, g_i and g_j are the same except at the point marked with an arrow.

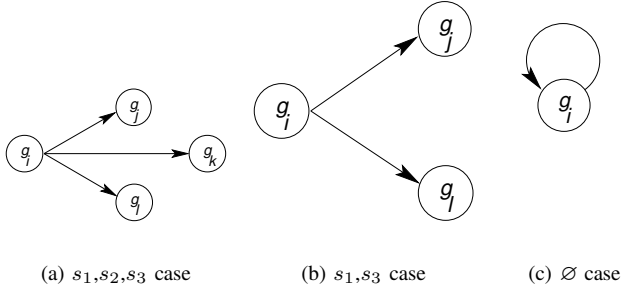


Fig. 6. Outgoing arcs for state g_i in the transition graph

C. Transition Graphs

We introduce *transition graphs* to explain how the next state is found if the present state is given. There is an individual graph $\Gamma(e, s)$ for every elementary event e and every subsystem s . All transition graphs have the same vertex set \mathcal{G} , the set of all states. The set of arcs, \mathcal{V} , depends on e and s . We define the set of graphs associated with event e as $\Gamma_e : \mathcal{G}, \mathcal{V}, e, \mathcal{S}$. The number of states adjacent to a given state is of the order of $|\mathcal{X}|$.

We explain the transition graphs with an example in which subsystems s_1 , s_2 , and s_3 are accessible in area a . The received signal levels from the access nodes of these subsystems will be different at every point in a . We denote the probability that the received signal from the access node of subsystem s_n is the strongest for any randomly selected user in area a as q_n . We examine the transition graph for the *outgoing new connection* event. Depending on the state, we have the $\{s_1, s_2, s_3\}$, $\{s_1, s_2\}$, $\{s_1, s_3\}$, $\{s_2, s_3\}$, $\{s_1\}$, $\{s_2\}$, $\{s_3\}$, and \emptyset cases where $\{x, y, z\}$ represents the case that corresponding cells of the subsystems x , y , and z have enough resources, and \emptyset represents the case that none of the subsystems has enough resources. The outgoing arcs in the transition diagrams that correspond to the $\{s_1, s_2, s_3\}$, $\{s_1, s_3\}$, and \emptyset cases are depicted in Figure 6(a-c), respectively. In Figure 6(a) for the $\{s_1, s_2, s_3\}$ case, the flow out of g_i due to outgoing new connections is split into three according to the ratios q_1 , q_2 , and q_3 toward the states g_j , g_k , and g_l . However, in Figure 6(b) for the $\{s_1, s_3\}$ case, the flow out of g_i is split into two according to the ratios $\frac{q_1}{q_1+q_3}$ and $\frac{q_3}{q_1+q_3}$ toward the states g_j and g_l .

D. Transition Probabilities

Transitions between states occur when *MT* initiates connections, moves in the service area, and hangs up. Due to the complexity of the NGCAC scheme, we examine each case separately.

1) New Connection Events:

a) Outgoing New Connection:

- Direct Outgoing New Connection:

The probability that a direct outgoing new connection

attempt of class k occurs for access node b is

$$P_{no}(b, k, \Delta t) = \Delta t \cdot \sum_{a \in \mathcal{A}(b)} \{r_a(t) \cdot f(a, k, t) \cdot n_a(t) \cdot p(a, k, s(b), t)\} + o(\Delta t) \quad (4)$$

where $\mathcal{A}(b)$ is the set of areas constituting the cell of access node b , $r_a(t)$ is the connection generation rate of all connection classes in area a at time t , and $s(b)$ is the subsystem to which access node b belongs.

- Indirect Outgoing New Connection:

The probability that an indirect outgoing new connection attempt of class k occurs for access node b is

$$\tilde{P}_{no}(b, k, \Delta t) = \Delta t \cdot \sum_{a \in \mathcal{A}(b)} \{r_a(t) \cdot f(a, k, t) \cdot n_a(t) \cdot \alpha(b)\} + o(\Delta t) \quad (5)$$

where

$$\alpha(b) = \frac{\sum_{u=1}^{|\mathcal{B}|} \left(\sum_{\substack{v=1 \\ v \neq b}}^u R^k(v) \cdot p_v \right)}{\sum_{w=1}^{|\mathcal{B}|} p_w - \sum_{v=1}^u p_v} \quad (6)$$

represents the total probability that *MT* prefers other access nodes serving the same area, but those access nodes cannot accommodate the request due to lack of resources, and $R^k(v)$ is the probability that access node v rejects a request of class k , i.e.,

$$R^k(v) = P \left(\left\{ \sum_{k \in \mathcal{C}} \sum_{a \in \mathcal{A}(b)} x_a^k(v) \right\} + bw(k) > C(v) \right)$$

where $C(v)$ is the capacity allocated to access node v .

b) Incoming New Connection:

- Direct Incoming New Connection:

The probability that a direct outgoing new connection attempt of class k occurs for access node b is

$$P_{ni}(b, k, \Delta t) = \Delta t \cdot \sum_{a \in \mathcal{A}(b)} \{r_a^k(t) \cdot n_a(t) \cdot p(a, k, s(b), t)\} + o(\Delta t) \quad (7)$$

- Indirect Incoming New Connection:

The probability that a direct outgoing new connection attempt of class k occurs for access node b is

$$\tilde{P}_{ni}(b, k, \Delta t) = \Delta t \cdot \sum_{a \in \mathcal{A}(b)} \{r_a^k(t) \cdot n_a(t) \cdot \alpha(b)\} + o(\Delta t) \quad (8)$$

Thus, the probability that a new connection attempt of class k occurs for access node b is

$$\begin{aligned} P_{new}(b, k, \Delta t) &= P_{no}(b, k, \Delta t) + \tilde{P}_{no}(b, k, \Delta t) \\ &+ P_{ni}(b, k, \Delta t) + \tilde{P}_{ni}(b, k, \Delta t) \\ &+ o(\Delta t). \end{aligned} \quad (9)$$

2) Migration Events:

a) *Intra-cell Movement:* The probability that a mobile that has an active connection of class k over access node b moves from one area to another in the same cell without changing its access node is

$$\begin{aligned} P_{intra}(b, k, \Delta t) &= \Delta t \cdot \sum_{a_j \in \mathcal{A}(b)} \sum_{\substack{a_i \in \mathcal{A}(b) \\ a_i \neq a_j}} \\ &V_{a_j, a_i}^k(t) \cdot x_{a_j}^k(b) + o(\Delta t). \end{aligned} \quad (10)$$

b) *Intra-subsystem Handoff:* The probability that access node b receives a handoff request of class k from another cell in the same subsystem, $s(b)$, is

$$\begin{aligned} P_{intra}(b, k, \Delta t) &= \Delta t \cdot \sum_{a_j \in \mathcal{A}} \sum_{a_i \in \mathcal{A}(b)} \sum_{\substack{c \in \mathcal{B} \\ s(c) = s(b) \\ c \neq b}} \\ &V_{a_j, a_i}^k(t) \cdot x_{a_j}^k(c) + o(\Delta t). \end{aligned} \quad (11)$$

c) Inter-subsystem Handoff:

• Direct Inter-subsystem Handoff:

The probability that access node b receives a direct handoff request of class k from another cell of another subsystem, $s(b)$, is

$$\begin{aligned} P_{inter}(b, k, \Delta t) &= \Delta t \cdot \sum_{a_j \in \mathcal{A}} \sum_{a_i \in \mathcal{A}(b)} \sum_{\substack{c \in \mathcal{B} \\ s(c) \neq s(b)}} \\ &V_{a_j, a_i}^k(t) \cdot x_{a_j}^k(c) \cdot p(a_j, k, s(b)) \\ &+ o(\Delta t). \end{aligned} \quad (12)$$

• Indirect Inter-subsystem Handoff:

The probability that access node b receives an indirect handoff request of class k from another cell of another subsystem, $s(b)$, is

$$\begin{aligned} \tilde{P}_{inter}(b, k, \Delta t) &= \Delta t \cdot \sum_{a_j \in \mathcal{A}} \sum_{a_i \in \mathcal{A}(b)} \sum_{\substack{c \in \mathcal{B} \\ s(c) \neq s(b)}} \\ &V_{a_j, a_i}^k(t) \cdot x_{a_i}^k(c) \cdot \alpha(b) \\ &+ o(\Delta t). \end{aligned} \quad (13)$$

Thus, the probability that a handoff attempt of class k occurs for access node b is

$$\begin{aligned} P_{handoff}(b, k, \Delta t) &= P_{intra}(b, k, \Delta t) \\ &+ P_{inter}(b, k, \Delta t) \\ &+ \tilde{P}_{inter}(b, k, \Delta t) \\ &+ o(\Delta t). \end{aligned} \quad (14)$$

3) *Hangup Event:* The probability that MT terminates the connection voluntarily is

$$\begin{aligned} P_{hangup}(b, k, \Delta t) &= \Delta t \cdot \sum_{a \in \mathcal{A}(b)} h_a^k(t) \cdot n_a(t) \\ &\cdot p(a, k, s(b), t) + o(\Delta t) \end{aligned} \quad (15)$$

where $h_a^k(t)$ is the rate at which class k connections in area a hangup at time t .

V. NUMERICAL RESULTS

A. Calculating State Probabilities

Calculating the state probabilities for such a complex system analytically is a challenge on its own. The state probabilities in our dynamical system can be calculated using simulations.

We denote the probability that the system is in state g_i at time t as $\mathbf{P}_i(t)$. Initially, the system starts in state g_0 in which there are no ongoing connections. Thus, $\mathbf{P}_0(0)$, the probability of being in state g_0 at time 0, is equal to 1.0 while $\mathbf{P}_i(0) = 0.0$ for all other states g_i . As time elapses, the elementary events mentioned in Subsection IV-B will occur, carrying the system from state to state. This will cause the state probabilities to change in time.

Though we do not know exactly in which state the system is at a given point in time, we can calculate the probability of being in each state g_i at time t , literally $\mathbf{P}_i(t)$. However, during a short time interval Δt , the system may switch to/from g_i due to the elementary events. Therefore, $\mathbf{P}_i(t + \Delta t)$ will be different from $\mathbf{P}_i(t)$. The change in the probability of being in state g_i during Δt , $\mathbf{P}'_i(t)$, is given by

$$\mathbf{P}'_i(t) = \mathbf{P}_i(t) \cdot \left(- \sum_{j \neq i} \lambda_{ij} \right) + \sum_{j \neq i} \mathbf{P}_j(t) \cdot \lambda_{ji} \quad (16)$$

where λ_{ij} is the rate of flow from state g_i to g_j . λ_{ij} values are calculated using Eqs. 4-15 by considering which state transitions are triggered by each elementary event.

B. Challenges in Calculating State Probabilities

The analytical model presented above provides a tool to calculate state probabilities. However, the size of the state space constitutes a major challenge. Any combination of values for the tuple given in Eq. 2 constitutes a different state as long as the cell capacities are not exceeded. To make the problem more tangible, we consider the simple cellular layout in Figure 7, where each cell is assigned only 4 channels. We also assume only one connection class for the sake of simplicity.

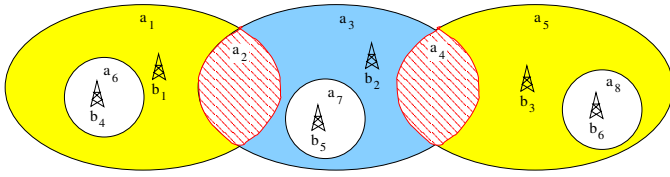


Fig. 7. Cellular layout

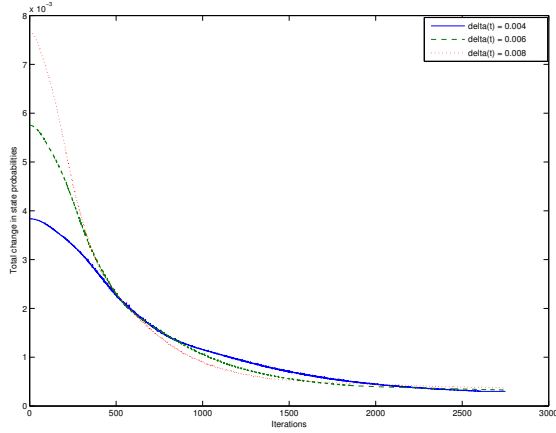


Fig. 8. Convergence of state probabilities

The reader should note that some of the users in a_2 may communicate over b_1 while others communicate over b_2 . The sum of the variables $x_{a_1}^1(b_1)$, $x_{a_1}^2(b_1)$, and $x_{a_1}^3(b_1)$ should not exceed 4, the capacity of the cell. When we consider the other variables similarly, we find that there are 201 different combinations for each access node. Since we have 6 access nodes, the number of states reaches $201^6 = 65944160601201$. However, some of these states are not effectively possible. For example, the value of the variable $x_{a_3}^1(b_1)$ should always be 0. When we exclude such states and consider only *effective states*, the number of states reduces down to 8962362486784. Since it is practically impossible to evaluate a model with so many states, we need to find an intelligent way to consider only states that are relevant for modelling purposes. One way to achieve this goal is to note that most of these states represent the case that the system is far away from trouble, i.e., no requests will be denied since there are enough resources. In other words, the system designer is interested only in states that represent cases where one or more cells are operating close to the capacity. We consider only the *critical states* (no channels available in some cells), *subcritical states* (only 1 channel available in some cells), and *subsubcritical states* (only 2 channels available in some cells). Thus, we reduce the number of states 156800, which is reasonable though still large. *Using this approach, we are able to find the probability that the system goes to the critical state, given that it is operating in the range close to full capacity.*

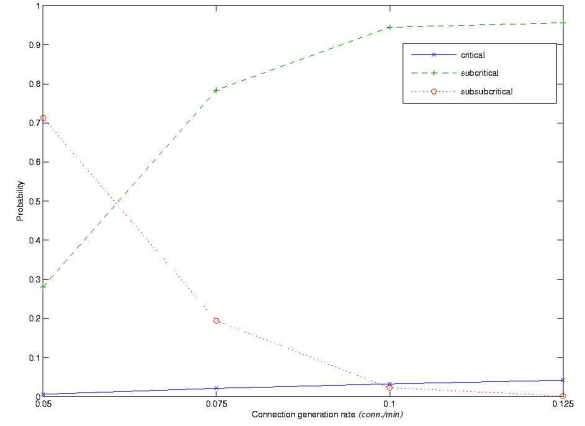


Fig. 9. Effect of connection generation rate

C. Convergence and State Probabilities

We first analyze how the system converges to stability. In each iteration, most of the state probabilities change, but as the number of iterations increase the change in state probabilities should decrease, and the system should converge. In Figure 8, the vertical axis represents the sum of the absolute values of the change in state probability for every state, for different time increments. For larger time increments, the state probabilities change faster in the beginning. However, as it is apparent from the figure, the system converges around 2000 – 2500 iterations for all cases. We have used 2000 iterations and a time increment of 0.004 s in the rest of the tests.

In Figure 9, we analyze the effect of connection generation rate on state probabilities. Each one of the three curves in the figure represents the sum of state probabilities of many states. We vary connection generation rate from 0.05 to 0.125 *conn/min* while hangup rate is fixed at 0.05 *conn/min*. As the connection generation rate increases, we observe that the systems shifts from subsubcritical states to critical and subcritical states. Thus, the analytical model provides a tool for the system designer to understand at what load the system goes to critical states.

Figure 10 represents the state probabilities for critical, subcritical, and subsubcritical under different loads. In case of lighter load, the system is mostly in subsubcritical states. As the load increases, the system spends most of its time in subcritical and critical states. If we analyze the change in state probabilities individually for each type of state, we observe that for critical and subcritical states the probability increases as the load increases. For the subsubcritical states, the probability decreases as the load increases, diminishing to almost zero for a connection generation rate of 0.125 *conn/sec*. This difference results from the fact that there is a shift towards critical states.

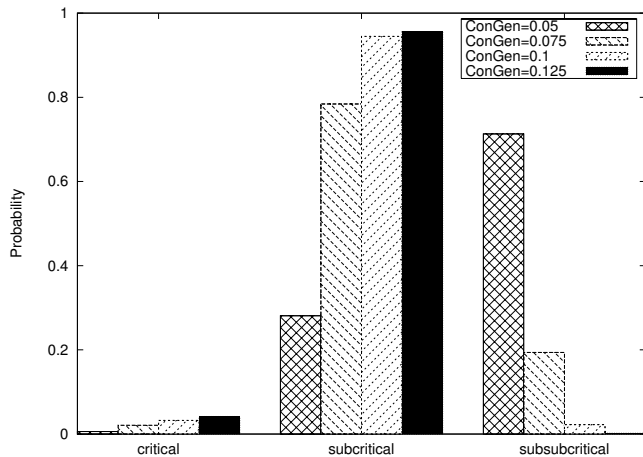


Fig. 10. Critical, subcritical, and subsubcritical state probabilities

VI. CONCLUSIONS AND FUTURE WORK

NGWS will be composed of multiple subsystems. The selection of the appropriate subsystem for connection setup is a crucial issue for overall system performance. In this paper, after defining the network architecture, we proposed a novel connection admission control scheme. The proposed NGCAC scheme considers the accessibility of the subsystems and the availability of the resources in those subsystems in addition to connection class of the connection request and the user's preferences. We also provided an analytical model of the proposed scheme. We pointed out major challenges in analytically modeling NGWS and provided a workaround. As future work, we will consider more complex scenarios and evaluate the effects of more parameters on the performance of NGWS.

ACKNOWLEDGMENT

This research is partially supported by Scientific and Technical Research Council of Turkey (TUBITAK) under grant number 104E032 and Bogazici University Research Fund under grant number BAP04S104.

REFERENCES

- [1] A. Ganz, C. M. Krishna, D. Tang, and Z. J. Haas, "On optimal design of multitier wireless cellular systems," *IEEE Communications Magazine*, vol. 35, pp. 88-93, February 1997.
- [2] X. Lagrange, "Multitier Cell Design," *IEEE Communications Magazine*, vol. 35, pp. 60-64, August 1997.
- [3] N. Faggion and T. Hua, "Personal Communications Services Through the Evolution of Fixed and Mobile Communications and the Intelligent Network Concept," *IEEE Network Magazine*, pp. 11-18, July 1998.
- [4] M. Zeng, A. Annamalai, and V. K. Bhargava, "Recent Advances in Cellular Wireless Communications," *IEEE Communications Magazine*, pp. 128-138, September 1999.
- [5] K. Buchanan, R. Fudge, D. McFarlane, T. Philips, A. Sasaki, and H. Xia, "IMT-2000: Service Provider's Perspective," *IEEE Personal Communications Magazine*, pp. 8-13, August 1997.
- [6] S. G. Niri and R. Tafazolli, "Cordless-cellular Network Integration for the 3rd Generation Personal Communication Systems," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 1, pp. 402-408, 1998.

- [7] A. Bria, F. Gessler, O. Queseth, R. Stridh, M. Unbehaun, J. Wu, J. Zander, and M. Flament, "4th-Generation Wireless Infrastructures: Scenarios and Research Challenges," *IEEE Personal Communications Magazine*, 2001.
- [8] M. Flament, F. Gessler, F. Lagergren, O. Queseth, R. Stridh, M. Unbehaun, J. Wu, and J. Zander, "Telecom Scenarios for the 4th Generation Wireless Infrastructures," in *Proceedings of the PCC Workshop*, (Stockholm), 1998.
- [9] NTT-DoCoMo, "Outline of Fourth-Generation Mobile Communications," in http://www.nttdocomo.co.jp/corporate/rd/new_e/4gen01_e.html, 2002.
- [10] R. Berezdivin, R. Breinig, and R. Topp, "Next Generation Wireless Communications Concepts and Technologies," *IEEE Communications Magazine*, vol. 40, pp. 108-116, March 2002.
- [11] J. Jimenez, "Towards the 4G," in <http://research.ac.upc.es/conferencies/ITCSS/jimenez.pdf>.
- [12] "WINEGLASS - Wireless IP NETwork as a Generic platform for Location Aware Service Support," in <http://domo-bili.csel.it/WineGlass>, 2002.
- [13] W. W. Lu, "Compact Multidimensional Broadband Wireless: The Convergence of Wireless Mobile and Access," *IEEE Communications Magazine*, pp. 119-123, November 2000.
- [14] 3rd Generation Partnership Project, "IP Based Multimedia Services Framework Report," 3GPP Technical Specification 23.228 V5.3.0.
- [15] T. Robles, A. Kadelka, H. Velayos, A. Lappetelainen, A. Kessler, H. Li, D. Mandato, J. Ojala, and B. Wegmann, "QoS Support for an All-IP System Beyond 3G," *IEEE Communications Magazine*, vol. 39, pp. 64-72, August 2001.
- [16] P. Conforto, C. Tocci, G. Losquadro, A. Spazio, R. Sheriff, M. Chan, and Y. Hu, "Ubiquitous Internet in an Integrated Satellite-Terrestrial Environment: The SUITED Solution," *IEEE Communications Magazine*, vol. 40, pp. 98-107, January 2002.