# Chernoff Bound and the Refined Large Deviation Approximation for Connection Admission Control in CDMA Systems

Yeong Min Jang ***Regular member***

ABSTRACT

This paper proposes a transient (predictive) connection admission control (CAC) scheme using the transient quality of service (QoS) measure for CDMA cellular systems with bursty On-Off sources. We need an approximate and bounded approach for real-time CAC applications. We derive the transient outage probability as the QoS measure using the Chernoff bound and the refined large deviation approximation. Numerical results show that the predictive CAC is a promising approach for the multicell CDMA systems.

## I. Introduction

The increasing popularity of wireless communication services together with the limited amount of the available radio spectrum calls for highly efficient usage of resources. CDMA (Code Division Multiple Access) has been the standard for the second and third-generation mobile telecommunication systems. Since CDMA capacity is interference limited, any reduction in interference converts directly and linearly into an increase in capacity. Blocking in CDMA systems will be defined to occur when the interference level, due to other user activity, reaches a predetermined level. Furthermore, any spatial isolation through use of multi-beamed or multi-sectored antennas, which reduces interference, provides a proportional increase in capacity.

CDMA utilizes the effect of statistical multiplexing without the complex radio channel allocation. The number of simultaneous users occupying a base station (BS) must be limited to satisfy the QoS requirement. CAC thus play an important role in CDMA system because it directly controls the number of users by accepting or rejecting users in uplink. A stringent limit on the number of simultaneous users is determined by the total interference created by the admitted users exceeding a threshold. The CAC is necessary when there are various types of traffic with various QoS requirements and when the system operates in the vicinity of its full capacity. Practically, interference-based CAC is mostly used for CDMA systems[1]. Although there are many QoS requirements, e.g., outage probability, frame loss ratio, delay, jitter, and etc., in this paper, outage probability at the uplink is chosen as the QoS measure.

Since the duration of any connection in a real-world network is always finite and real-world traffic is typically bursty, the performance of the connection is in fact dominated by the *transient* behavior of the underlying traffic process carried by the connection. Control schemes based on a *transient* analysis of outage probability can achieve a high link utilization. The operating CDMA systems may not reach to the steady-state. So, we need a *transient* approach for CAC. For traffic modeled as a superposition of On-Off sources, many approaches for evaluating the outage probability have been proposed[1,2]. Since the connection setup time is constrained, these rather complex models are not practical. indeed, one needs a fast algorithm that decides "on the fly" whether a new connection may be accepted or not. Generally, however, queuing analyses, such

논문번호 : 010302-1024, 접수일자 : 2001년 10월 24일
※ This work was supported by a grant from Duksung Women's University Research Fund 2001.

as bufferless fluid-flow models, provide only steady-state results [7] due to the complexity of modeling transient behavior. Large deviation theory, for example, provides a methodology for estimating probabilities of rare events[6]. The CAC problem has been extensively studied for CDMA networks[1,2]. Control schemes using fluid-flow models require the burdensome computation of a matrix inversion. We focus on the performance evaluation of a transient bufferless fluid-flow model with On-Off traffic and develop a Chernoff bound and the refined large deviation approach for a real-time CAC.

This paper is organized as follows. In Section 2, we describe the model assumptions, the traffic model, and the proposed CAC algorithm. The uplink capacity is calculated in Section 3. Section 4 derives the transient outage probability based on Chernoff bound and the refined large deviation approximation. An application for CAC is presented in Section 5. Section 6 presents the numerical results. The paper is concluded in Section 7.

## Ⅱ. Model Description

### 2.1. Model Assumptions

For simplicity, we assume that all cells are equally loaded with the same number of connections per cell and sector which are uniformly distributed over each sector. Assuming a perfect power control, the received power at the base station transceiver system (BTS) would be the same for each mobile terminal (MT). We also assume that the number of MTs and active connections at each cell are same. Suppose that N independent On-Off sources (connections) are connected to the CDMA uplink. A VP (virtual path) can be modeled as a single server system with uplink capacity. A QoS predictor at the BSC predicts outage probability for a *time t* (=100 ms) ahead. We assume that the uplink distance between the MTs in the reference cell and the BSC is same. In this work, we assume that arrival processes are stationary. We use a statistical bufferless fluid model to predict outage probability at *time t*, (i.e., the round-trip propagation delay) based on the traffic statistical behavior and the number of active sources at *time 0*.

### 2.2. Traffic Model

A voice source is active when the talker is actually speaking, and it is idle when the speaker is silent. We assume that simple Markovian models i.e., models based on the binary On-Off model, are sufficient to capture the short range correlation for telephone speech and video telephone. Each source is modeled as an On-Off source. We assume that a series of ATM cells arrive in the form of a continuous stream of bits and use a fluid model. We also assume that the "OFF" and "ON" periods for sources are both exponentially distributed with parameters $\lambda$ and $\mu$, respectively. The transition rate from ON to OFF is $\mu$, and from OFF to ON is $\lambda$. Hence the average length of the "ON" and "OFF" periods is $\frac{1}{\mu}$ and $\frac{1}{\lambda}$, respectively. In this traffic model, when a source is "ON", it generates ATM cells with a constant interarrival time, $\frac{1}{Rp}$ seconds/bit. When the source is "OFF", it does not generate any cells. In reality, there may be multiple rates to indicate activity/inactivity as is done in IS-95 and IMT-2000 systems. However, the On-Off case considered here is sufficient to provide insight into system functioning.

### 2.3. Proposed CAC Algorithm

We next provide an outline of the predictive CAC algorithm. A flow chart of the CAC algorithm is depicted in Fig. 1.

The CAC function is network specific. Each user who requests a connection is subject to CAC. This decision is based on the traffic descriptors of the new source such as the peak and average transmission rates and the average burst length. The connection is admitted into the radio access network if the QoS requirements can be met for both the new connection and the existing connections. A predictive CAC algorithm

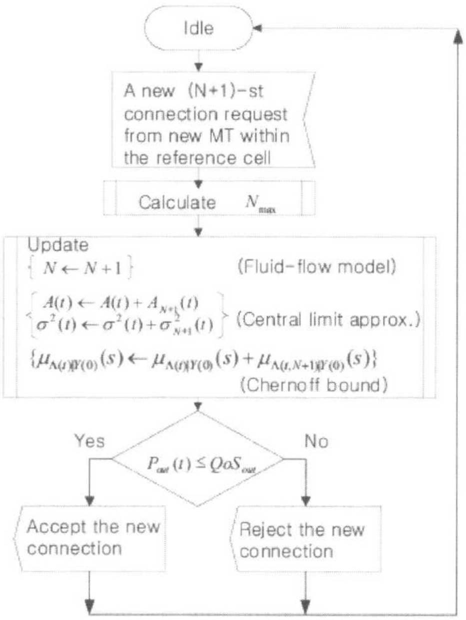


Fig. 1. Porposed CAC algorithm for a new (N+1)-st connection request.

is implemented in the BSC and executed for each radio cell. The algorithm is executed whenever we have a new connection request. The connection acceptance decision must be made within the required allowable connection set-up latency in accordance with the received connection set-up request. This means that the CAC must operate in real-time, even when various kinds of traffic are multiplexed.

## Ⅲ. Calculating the Uplink Capacity

### 3.1. Calculating the Mean Arrival Rate using Fluid-flow Model

Let $\Lambda(t)(=R_p Y(t))$ denote the aggregate arrival rate from $Y(t)$ active connections. According to a fluid model, outages occur at *time t* when $\Lambda(t)$ exceeds the uplink capacity. Taking into consideration the fact that each of N existing connections, given by an arbitrary initial condition Y(0)=*i*, we obtain the conditional moment generating function of $\Lambda(t)$, s $\geq$ 0 [3]:

$$G_{\Lambda(t)|Y(0)}(s) = E[\,e^{s\Lambda(t)} \mid Y(0)=i]$$

$$= [p(t)(e^{sR_p}-1)+1]^{N-i}\,[q(t)(e^{sR_p}-1)+1]^{i}. \quad (1)$$

p(t) and q(t) is defined as:

$$p(t) = \frac{\lambda}{\lambda+\mu}[1-e^{(\lambda+\mu)t}] \quad (2)$$

and by symmetry,

$$q(t) = \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t}, \quad (3)$$

where p(t) is the transition probability that a source is active at future *time t*, given the source is idle at *time 0*. q(t) is the transition probability that a source is active at future *time t*, given the source is active at *time 0*.

Thus, the conditional mean arrival rate is

$$A(t) = G'_{\Lambda(t)|Y(0)} = R_p[(N-i)p(t)+iq(t)]. \quad (4)$$

### 3.2. Calculating the Uplink Capacity from $(\frac{E_b}{I_o})req$

In a CDMA system, the number of connections within each cell is critically dependent upon the interference received at the BTS. In a multicell CDMA system, received interference consists of interference from MTs within the same cell and, in addition, interference due to MTs within surrounding cells. The number of available virtual trunks, or uplink capacity is determined by the total interferences, On-Off traffic parameters $(\lambda, \mu, R_p)$, and the number of active connections at time *0* (*i*). The outage probability is defined as the probability of the bit error rate (BER) exceeding a certain threshold $10^{-3}$ required for acceptable performance. As stated in [3], with an efficient modem, a powerful convolutional code, and two-antenna diversity, adequate performance (BER < $10^{-3}$) is achieved on the uplink with $(\frac{E_b}{I_o})req$ > 7dB. When a connection arrives, it gets blocked if it does not find an idle virtual trunk, which is a radio channel resource. We obtain a distribution on the total interference, both from other users in the reference cell, and from other-cell users on the desired user's uplink transmission. Let W(Hz) be the spreading bandwidth, $(\frac{E_b}{I_o})_{req}$ be the required target value, $R_p$ be the peak transmission rate, and $f$ represents the other cells interference as a fraction of the interference from the reference cell. The estimate of total other-cell interference as being from 33% to 42% of the total power received from MTs in the same cell. Recall that $N$ denotes the total number of connections in the reference cell. $S$ denotes the total power received from each MTs at the BTS. A perfect power control at each BTS ensures that the total power received from each MTs within that cell is limited to $S$. $N_0$ stands for the background thermal noise spectral density. Including the transient voice activity factor $(p(t), q(t))$, the number of active connections at time $0$ ($i$), and taking into account the inference from adjacent cells ($f$), the maximum number of connections, $N_{\max}$, which may be supported within any particular cell can be expressed as:

$$N_{\max} = \frac{W}{P(t)(f+1)\ R_p\,(\frac{E_b}{I_0})_{req}} + \frac{1}{f+1} - \frac{i(q(t)-p(t))}{p(t)} - \frac{N_0 W}{p(t)(f+1)S}, \quad (5)$$

where $N_{\max} \geq i$.

The capacity of the CDMA system in terms of maximum cell loading can be found by finding the maximum $N_{\max}$ such that for the required QoS (BER< $10^{-3}$or $(\frac{E_b}{I_o})_{req}$ > 7 dB) for adequate performance of the modem and decoder for digital voice transmission. To obtain outage probability as a function of $N_{\max}$, we used a bufferless fluid-flow model [4]. Unfortunately, control schemes using fluid-flow models require the burdensome computation of a matrix inversion. The exact calculation of outage probability for each connection set-up stage is not possible for real-time CAC because the computational complexity increases rapidly with an increase in the number of existing connections. So we will use a bound and approximation model in next section.

## IV. Transient Outage Probability

### 4.1. Chernoff Bound

Let us consider a bufferless queuing system fed by $N$ On-Off fluid stochastic processes, where the state of the arrival process represents the arrival rate of workload (power or traffic) to the system, that is, the amount of workload arrived per time unit. We will use a statistical bufferless fluid-flow model to predict the probability that outage occurs at *time t* based on the traffic statistical behavior at *time 0*. Equivalently, we can express the capacity of the CDMA uplink, as the maximum number of simultaneously active connections, $\lceil N_{\max} \rceil - 1$, that can be supported on the uplink. $\lceil x \rceil$ is the smallest integer that is greater than or equal to a given number $x$. We assume that $N$ On-Off sources in each radio cell share the capacity, $N_{\max}\ R_p$ of a uplink of CDMA system. According to a fluid model, outages occurs at *time t* when $\Lambda(t)$ exceeds the link capacity $N_{\max}\ R_p$. To estimate small $P_{out}(t)$, which is several standard deviations away from the mean, other approximations such as the Chernoff bound and the large deviations theory can be used. For small tail probabilities, the central limit approximation can underestimate the outage probability[4,5]. In this subsection, we provide an upper bound for $P_{out}(t)$. To this end, for any random process $\Lambda(t)$ and constant $N_{\max}\ R_p$, let

$$\mu_{\Lambda(t)|Y(0)}(s) = \ln G_{\Lambda(t)|Y(0)}(s) = \ln\ E[e^{s\Lambda(t)} | Y(0) = i]. \quad (6)$$

Note that $\mu_{\Lambda(t)|Y(0)}(s)$ is the logarithm of the conditional moment generating function for $\Lambda(t)$. $\mu_{\Lambda(t,N+1)|Y_{N+1}(0)=1}(s)$ is the logarithm of the conditional moment generating function for $\Lambda(t, N+1)$, associated with a new connection request. The Chernoff bound gives an upper bound for $P(\Lambda(t) > N_{max} R_p | Y(0) = i)$. Assume that $E\{\lambda(t)\}$ exists and $N_{max} R_p \geq E\{\lambda(t)\} \geq -\infty$. Then the supremum in Eq. (7) is obtained within values $s^* \geq 0$, and minimizing the right hand side of Eq. (7) with respect to $s$ yields the Chernoff upper bound

$$\begin{aligned} P_{out}(t) &= P(\Lambda(t) > N_{max} R_p | Y(0) = i) \\ &\leq e^{-s^* N_{max} R_p} G_{\Lambda(t)|Y(0)}(s^*) \\ &= e^{-s^* C + \mu_{\Lambda(t)|Y(0)}(s^*)}, \end{aligned}$$

for $s^* \geq 0$ and $N_{max} R_p \geq E\{\Lambda(t)\}$. (7)

where $s^*$ is the unique solution to

$$\begin{aligned} \mu'_{\Lambda(t)|Y(0)}(s^*) &= \frac{\frac{d}{ds^*} G_{\Lambda(t)|Y(0)}(s^*)}{G_{\Lambda(t)|Y(0)}(s^*)} \\ &= \frac{(N-i)p(t) R_p e^{s^* R_p}}{p(t)(e^{s^* R_p} - 1) + 1} + \frac{iq(t) R_p e^{s^* R_p}}{q(t)(e^{s^* R_p} - 1) + 1} \\ &= N_{max} R_p. \end{aligned} \quad (8)$$

The Chernoff bound can be used to estimate the number of connections that is needed to satisfy a given the conditional outage probability bound, $QoS_{out}$ at the uplink, i.e. $P(\Lambda(t) > N_{max} R_p | Y(0) = i) \leq QoS_{out}$. The Chernoff bound gives the admission control condition

$$e^{-s^* C + \mu_{\Lambda(t)|Y(0)}(s^*)} \leq QoS_{out}. \quad (9)$$

### 4.2. Refined Large Deviation Approximation

The upper bound given by Eq. (9) is often in the correct order of magnitude. But the bound can be converted to an accurate approximation by applying the theory of large deviation and the central limit theorem. The shifted distribution can be approximated very accurately by a normal distribution around its own mean [6].

Suppose that $\Lambda_{s^*}(t)$ is approximately normally distributed with conditional mean $\mu'_{\Lambda(t)|Y(0)}(s^*) (= N_{max} R_p)$ and conditional variance $\mu''_{\Lambda(t)|Y(0)}(s^*)$, The Q(u) is upper bounded by $\frac{e^{\frac{-u^2}{2}}}{u\sqrt{2\pi}}$. Thus,

$$P_{out}(t) \approx \frac{1}{s^* \sqrt{2\pi \mu''_{\Lambda(t)|Y(0)}(s^*)}} e^{-s^* N_{max} R_p + \mu_{\Lambda(t)|Y(0)}(s^*)}, \quad (10)$$

where

$$\begin{aligned} &\mu''_{\Lambda(t)|Y(0)}(s^*) \\ &= \frac{(N-i)p(t) R_p^2 e^{s^* R_p}[2p(t) e^{s^* R_p} - p(t) + 1]}{(p(t)(e^{s^* R_p} - 1) + 1)^2} \\ &\quad + \frac{iq(t) R_p^2 e^{s^* R_p}[2q(t) e^{s^* R_p} - q(t) + 1]}{(q(t)(e^{s^* R_p} - 1) + 1)^2}. \end{aligned} \quad (11)$$

Note that the correcting factor of the exponential term considerably improves accuracy by applying the central limit theorem. A refined large deviation approximation for $P_{out}(t)$ gives the following admission control condition:

$$P_{out}(t) \approx \frac{1}{s^* \sqrt{2\pi \mu''_{\Lambda(t)|Y(0)}(s^*)}} e^{-s^* N_{max} R_p + \mu_{\Lambda(t)|Y(0)}(s^*)} \leq QoS_{out}. \quad (12)$$

## V. Application to CAC

Using this admission rule Eqs. (9) and (12), a new (N+1)-st connection is established. We update $N \leftarrow N+1$. We then admit the new connection if, and only if, the condition in Eq. (13) is met.

$$P_{out}(t) = \sum_{k=\lceil N_{max} \rceil}^{N+1} P_k(t) \leq QoS_{out}. \quad (13)$$

Let $N^*$ be the optimal number of connections that can be supported in a cell such that the probability of outage i.e., the number of simultaneously bursting connections, exceeds $\lceil N_{max} \rceil - 1$ with probability less than $QoS_{out}$. When a subscriber requests to establish a connection to the BSC, the BSC radio resource management (RRM) function which executes the bandwidth allocation will predict the outage probability. Then the RRM function will deny or accept the connections. Based on the QoS requirement, the optimal number of connections, $N^*$, will be computed as follows:

$$N^* = \max\{N | P_{out}(t) = \sum_{k=\lceil N_{max} \rceil}^{N} P_k(t) \leq QoS_{out}\}. \quad (14)$$

To satisfy the QoS, the RRM function at BSC would accept any new connections as long as $N^*$ > N, the number of current connections. Using Eq. (12), the admission control scheme operates as follows. The logarithms of conditional moment generating functions, $\mu_{\Lambda(t)|Y(0)}(s)$, are first calculated off line for all connections. Now, suppose a new connection N+1 is requested. We update $\mu_{\Lambda(t)|Y(0)}(s) \leftarrow \mu_{\Lambda(t)|Y(0)}(s) + \mu_{\Lambda(t,N+1)|Y_{N+1}(0)=1}(s)$ and find the $s^*$ that satisfies Eq. (11). Finally, we admit the connection if and only if Eq. (12) is satisfied.

## Ⅵ. Numerical Results

Some numerical results have been generated. As an example, consider the following system: t=0.1 sec, $R_p$=9.6kbps, $\lambda$=0.5, $\mu$=0.833, $(\frac{E_b}{I_o})_{req}$ =7 dB, $f$=0.33, $W$=1.25 Mbps (for IS-95), $S$=-60 dBm, and $N_o$=-166 dBm/Hz. The linearity or nonlinearity of the queuing system implies that the total output response at *time t* is a function of several inputs (connections) during t and an initial condition at *time 0*. For the purpose of illustration, in this paper, we assume an average connection duration of 180 sec and the prediction time of 0.1 sec. Then, the mean interarrival time of connections is 3.6 sec for 50 On-Off connections. So we may assume that no new/completed connections (sources) occur during the prediction time (0.1 sec).

We find the maximum number of connections in a radio cell according to the number of initial active connections using Eq. (5). The more the number of initial active connections, the less the maximum number of connections. The parameters of On-Off sources also affect the $N_{max}$ [4,5]. To compare transient outage against outage in steady state, we need to compute the steady-state limits of the outage measures. The transient approaches are more complex than the steady-state approaches. The outage measures in steady state are defined by $P_{out}(\infty)$. We compare the transient outage probability with the steady-state outage probability ($P_{out}(\infty)$) (which is independent of the number of active connections at *time 0*). At 0.1 sec, the round trip delay and processing delay in a CDMA system, we observe differences in the results obtained as a function of different initial conditions (the number of active connections at time 0 ). We, therefore, conclude that the computation of the transient $N_{max}$ is more accurate and can lead to different values of the steady-state outage probability, and consequently to different CAC decisions.

We assume the following network and radio channel parameters: t= 0.1 sec, $R_p$=9.6 kbps, $\lambda$ =0.5, $\mu$=0.833, $QoS_{out}$ = $10^{-2}$, $(\frac{E_b}{I_o})_{req}$=7dB, $f$=0.33, $W$=1.25Mbps, $S$=-60dBm, and $N_o$ =-166dBm/Hz. But we also may apply for WCDMA (for example, $(\frac{E_b}{I_o})_{req}$=5dB, $W$=5.12 Mcps and $R_p$=74.3kbps). But, in this paper, we only focus on IS-95 system. To satisfy the QoS, we have to decrease the optimal number of connections, $N^*$. We also obtain the performance of the proposed CAC algorithm to obtain the optimal number of connections, $N^*$ in Fig. 2. Whenever the number of active connection at *time 0* are increased, the transient outage probability using the central limit approximation is more close to that using fluid-flow model[4,5]. For small tail probabilities, for example Y(0)=0, the central limit approximation can underestimate the outage probability. The Chernoff bounds give us the upper bound of exact fluid and refined large deviation approximation (LDA) approaches. The refined LDA results give us the optimistic results for the exact fluid results. So we can conclude that exact fluid analysis and the refined LDA method are very close irrespective of the initial conditions. For the larger number of active sources, central limit approximation may be used because we can avoid the solution for Eq. (8), which has been used for the refined LDA and Chernoff bound approaches. The nonlinear equation, Eq. (8), can be solved by the standard Newton-bisection method. Whenever we increase the initial number of active sources, the $s^*$ value is decreased.

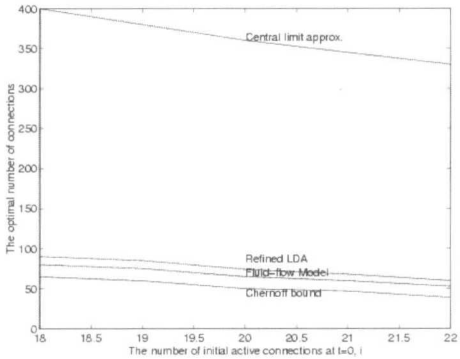


Fig. 2. The optimal number of connections, $N^*$, for different values of initial number of active connections using $P_{out}(0.1)$

Fig. 2 shows the importance of considering the transient outage probability, the quantity we use to make CAC decisions. We, therefore, conclude that the computation of the transient outage probability is more accurate and consequently to different CAC decisions.

## Ⅶ. Conclusions

We have developed approximation and bound rules using the transient QoS measures for admitting On-Off sources to the uplink. An advantage of the transient scheme is that the network may be more fully utilized because network states are almost always in the transient-state, instead of the steady-state. The central limit approximation slightly underestimated the transient outage probability. We have derived the transient outage probability using a fluid model[4], central limit approximation[5], Chernoff bound, and the refined large deviation approximations. We have found that the refined large deviation approximation is quite accurate for both transient state and steady state. The proposed predictive (transient) scheme is an excellent candidate for real-time CAC.

## References

[1] Y. Ishikawa and N. Umeda, "Capacity design and performance of call admission control in cellular CDMA systems," *IEEE JSAC,* Oct. 1997.

[2] W. Matragi and S. Nanda, "Capacity analysis of an integrated voice and data CDMA system," *IEEE VTC*, 1999.

[3] K. S. Gilhousen and et al., "On On the capacity of a cellular CDMA system," *IEEE Tran. on Vehicular Tech.,* May 1991.

[4] Y. M. Jang and J. Ahn, "A connection admission control using transient outage probability in CDMA systems," *IEEE VTC* 2000, Sept. 2000.

[5] Y. M. Jang, "Central limit approximation approach for connection admission control for broadband satellite systems," *IEE Electronics Letters,* Feb., 2000.

[6] J. Y. Hui, *Switching and Traffic Theory for Integrated Broadband Networks*, Boston, MA; Kluwer, 1990.

[7] M. Reisslein and K. W. Ross, "Call admission for prerecorded sources with packet loss," *IEEE JSACs*, Aug. 1997.

장 영 민(張暎民)

1985년: 경북대학교 전자공학과 학사

1987년: 경북대학교 전자공학과 석사

1999년: University of Massachusetts 컴퓨터과학과 박사

1987년~2000년: 한국전자통신연구원 무선방송기술연구소 선임연구원

2000년~현재: 덕성여자대학교 컴퓨터과학부 교수

2001년~현재: 정보통신부 IT 표준 전문가(4세대 이동통신 분야)

<주관심 분야> 4G Mobile Systems(PHY Layer, MAC, RLC, RRC, RRM, Mobile IP), WLAN, Bluetooth, and Satellite Networks.