Random Access Design for UMTS Air-Interface Evolution

Amitava Ghosh, Rapeepat Ratasuk, Igor Filipovich, Jun Tan, Weimin Xiao
Networks and Enterprise Business, Motorola
1501 West Shure Drive, Arlington Heights, IL 60004, USA

***Abstract* — Comprehensive long term evolution of the Universal Mobile Telecommunications System (UMTS) specifications is currently ongoing to provide significant improvement over the current release. Important goals for the evolved system include significantly improved system capacity and coverage, low latency, reduced operating costs, multi-antenna support, flexible bandwidth operations and seamless integration with existing systems. To ensure low latency, users must be able to establish a connection to the network quickly. This paper provides a preliminary design and procedure for the random access channel used to establish a connection when the mobile is not yet time-synchronized to the network in the uplink.**

I. Introduction

With the emergence of packet-based mobile broadband systems such as 802.16e, it is evident that a comprehensive long term evolution (LTE) of UMTS is required to remain competitive. As a result, work has begun on Evolved UMTS Terrestrial Radio Access (E-UTRA) aimed at commercial deployment around the 2010 timeframe. Long term goals for the system include support for high peak data rates (100 Mbps downlink and 50 Mbps uplink), low latency (10ms round-trip delay, 100ms control plane delay), improved system capacity and coverage, reduced operating costs, multi-antenna support, efficient support for packet data transmission, flexible bandwidth operations (up to 20 MHz) and seamless integration with existing systems [1]. To reach these goals, a new design for the air interface has been adopted with Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink and Single-Carrier Frequency Division Multiple Access (SC-FDMA) in the uplink.

When the mobile is not time-synchronized to the base station in the uplink, it must use a contention-based random access channel to access the network. It may use this channel to request initial access, initiate handoff procedure, and transition from idle to connected state. To ensure low latency, random access procedure must be designed such that the control plane latency requirement of less than 100 ms is achieved. This paper provides a preliminary physical layer design and procedure for random access in E-UTRA.

II. E-UTRA Random Access

Figure 1 illustrates non-synchronized random access structure. In the figure, random access occupies a bandwidth of BW_{RA} = 1.08 MHz and its length is a multiple of 1ms sub-frame. Multiple frequency regions may be defined within one access period in order to provide sufficient random access opportunities. In addition, this access period occurs at a fixed timing from the system broadcast frame. Although the timing is known, the number, location, and periodicity of the channels are signaled on the broadcast channel. In addition, these parameters may be configurable based on random access load. Note that data transmission may also be scheduled in the random access regions at the discretion of the scheduler.

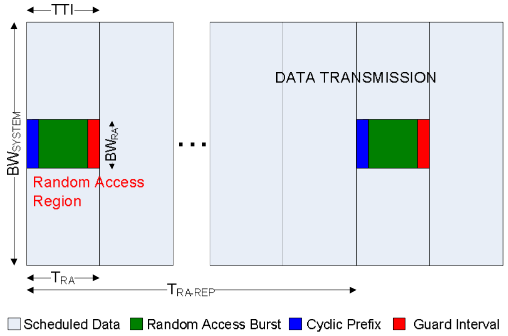

Figure 1. Non-synchronized random access.

Within each random access region, multiple users may access the channel simultaneously by selecting different preambles. The preambles must exhibit good detection performance and robustness to interference as well as provide accurate timing estimation. This is because in E-UTRA uplink transmissions must be synchronized in order to prevent interference. In [8], it was shown that Ep/No (energy per sequence over noise) of approximately 16-18 dB is required to achieve 1% false alarm and missed detection probabilities. In addition, random access must be possible from the cell edge where the operating C/I may be very poor. As a result, to meet coverage requirement, it was determined that only the preamble (and no message) can be transmitted in the contention channel. From coverage perspective, it was determined that 64 different preambles can be supported per random access region.

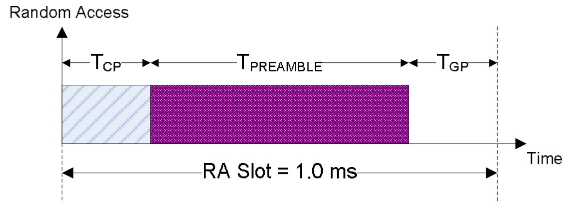

Figure 2. Non-synchronized random access burst.

Figure 2 illustrates the baseline 1.0ms random access burst with cyclic prefix (T_{CP}) of 100µs, preamble length of 800µs, and guard time (T_{GP}) of 100µs. A cyclic prefix is added to aid in frequency domain processing in order to reduce detection complexity. A guard interval is required to prevent interference to other transmissions arising from timing misalignment. This timing misalignment between mobiles in the cell is dependent on the cell size and the 100µs guard interval corresponds to a supportable cell radius of approximately 14 km. However, in E-UTRA, random access

1550-2252/$25.00 ©2007 IEEE

must be designed to support large cells of size up to 100 km. In this case, repetition is used to extend the random access burst with appropriate adjustments to the cyclic prefix and guard period.

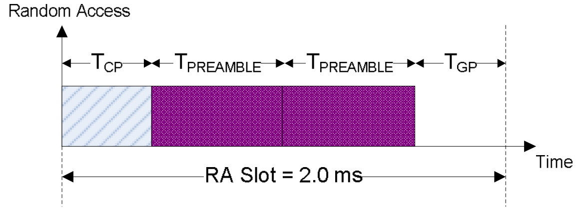

Figure 3. Extended random access burst for large cells.

For instance, to support a 25 km cell, an extended 2ms random access burst is deployed as shown in Figure 3 where the 800µs preamble is repeated twice with the cyclic prefix length and guard interval each extended to 200µs.

III. RANDOM ACCESS PROCEDURE

A. Access Procedure

Figure 4 provides the random access procedure for initial access. Four different messages are exchanged as part of the random access procedure and contention resolution. They are (1) random access preamble, (2) random access response, (3) RRC connection request, and (4) RRC contention resolution. First, a mobile randomly selects a random channel. Next, it selects a random preamble from the set of available preambles. The mobile then determines the initial power setting using open-loop power control and transmits the preamble.

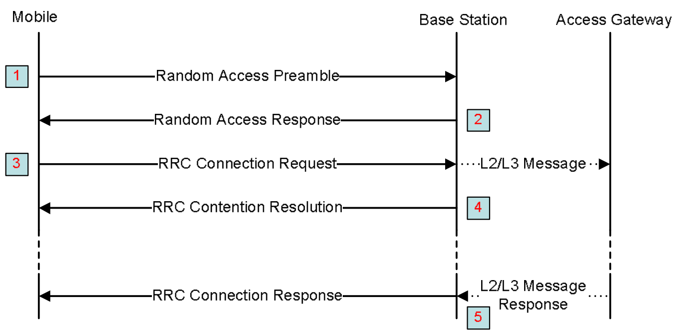

Figure 4. Random access procedure.

Upon reception of the random access preamble, the base station sends a response in a semi-synchronous manner in order to allow for some scheduling flexibility and load balancing. This response is sent using a combination of L1/L2 control and downlink shared data channels. The L1/L2 control channel points to the location within the shared data channel where the actual random access response is contained. Note that multiple responses may be multiplexed into the shared data channel. Each random access response contains an uplink scheduling grant for RRC connection request transmission, timing advance information, and an assignment of a temporary C-RNTI (Cell Radio Network Temporary Identifier). Timing advance information is used by the mobile to time-synchronized its uplink transmission. Note that H-ARQ is not used for transmission of the random access response due to possible contention (i.e. more than one mobiles selecting the same preamble sequence).

In the third step, the mobile transmits its RRC connection request in the uplink using its temporary C-RNTI. Included in this message is the mobile identifier and whether this mobile has already been assigned a C-RNTI from its previous network access. This message is of dynamic size and H-ARQ can be used to ensure it is successfully received at the base station. The temporary C-RNTI will serve as its identity for contention resolution purposes which the base station would echo in the fourth message. This would serve as an early indication if a collision occurred during the previous transmission and allow the mobile to reinitiate random access procedure as soon as possible. Otherwise, the mobile will have to wait until the fifth message before contention is resolved. This may incur significant delay since the RRC connection request response has to come from the Access Gateway.

B. Random Access Load and Overhead

Non-synchronized random access will be used for initial access, handoffs to a non-synchronized cell, and scheduling requests. In general, random access load will be dominated by handoff requests. For random access load approximation, a traffic model representing busy hour traffic statistics was provided in [10]. In this model, two different traffic types were considered – real-time service such as VoIP and non real-time service such as web browsing. For real-time service, an average of 1 call per hour per user is assumed with the call duration of 90 seconds. For non real-time traffic, an average of 2 calls per hour per user is assumed with duration of 300 seconds. To estimate the number of handoffs, a mobile is assumed to change serving cell every 20 seconds. Using this traffic model, Figure 5 illustrates the probability of collision as a function of the number of camped mobiles within the cell for different random access overheads.

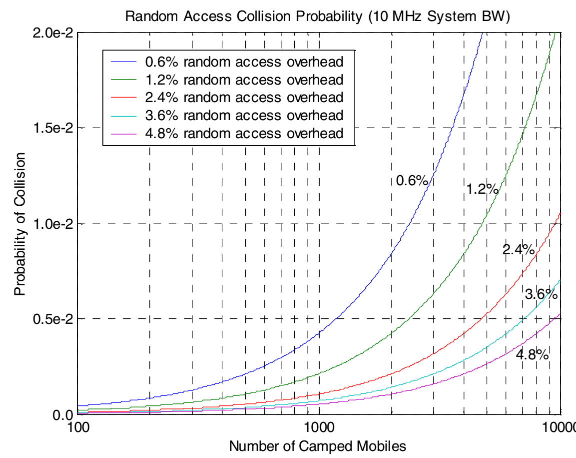

Figure 5. Random access collision probability.

From the figure, it is seen that amount of random access overhead required is naturally dependent on the target collision probability and cell load. For instance, at 5e-3 collision probability and 7000 camped mobiles, 3.6% of the system bandwidth is required to support random access. This translates to three random access channels for each 10ms radio frame. If the collision probability can be relaxed to 1e-2, then the overhead is correspondingly reduced to 2.4%. It should be noted that the traffic model in [10] is somewhat simplistic as only two traffic types are considered. In some situations the actual random access load may be significantly higher. Nonetheless, it may be inferred from the analysis that random access overhead of 3-6% should be sufficient to handle most deployment scenarios.

C. Latency Analysis

A key requirement of E-UTRA is to limit control plane latency to less than 100 ms. From Figure 4, it can be seen that random access failure due to for example poor radio conditions or missed detection can happen in each of the four messages. In addition, contention resolution is not performed until the fourth message. In [11], it was shown that a minimum of 12 ms is required to complete the RRC contention resolution step and that a large percentage of users is able to complete the random access procedure in this time frame. However, 30-50 ms access time may be required for a small percentage of the mobiles (less than 5%) due to contention. Nonetheless, from the analysis shown in [11] it may be concluded that control plane latency of less than 100 ms is feasible with this procedure.

D. Simultaneous Data Transmission

Since low latency is an important requirement, random access channels must occur with sufficient frequency (e.g. every 10 ms or so) which may result in high bandwidth overhead especially for system with smaller bandwidth. For 10 MHz system, it is seen that random access overhead of 3-6% may be needed. In [8], it was shown that received sequence energy over noise of 16-18 dB is required to meet 1% false alarm and missed detection probabilities. For the 1ms preamble structure, this energy is spread over 800µs resulting in very low preamble SINR requirement per sub-carrier. Thus, it should be possible to transmit data simultaneously on some or all of the resource blocks used for non-synchronized random access. This is especially beneficial when systems are temporarily experiencing low random access loads (e.g. in the middle of the night). Naturally, the decision to schedule data users in those resource blocks lies with the base station and other factors such as possible interference should also be taken into consideration. In practice, the base station may restrict data transmission only to low QoS traffic such as best effort services and then for example using very low data rates to minimize possible interference. As a result, it is being proposed that data transmission may also be scheduled in the random access region at the discretion of the scheduler.

E. Preamble Message Content

Because random access load is expected to be small, it should be possible to convey some implicit information to the base station while still maintaining low overall random access collision probability. In this case, approximate knowledge of channel condition (e.g. good or bad channel) can substantially reduce the amount of resource required for the preamble response. This information may be used as a guide for downlink power control or link adaptation and possibly for uplink resource assignment. Without any channel knowledge, fixed downlink transmission power to reach the desired cell coverage must be used. This is substantial as the downlink C/I may be as low as -4 dB at the 95% cell coverage for systems with 1732 meter inter-site distance and 20 dB penetration loss [6]. At this operating point, the spectral efficiency is approximately 0.3 b/s/Hz [7]. Thus, to convey an exemplar 40-bit random access response will require approximately 130 sub-carriers, which is a substantial overhead.

Table 1. Average control channel power saving.

No of bits used to convey channel condition	Average Power Saving (dB)
0	0.0
1	4.5
2	6.7

Table 1 lists potential saving in downlink control channel power allocation if some rough channel quality information is available assuming uniform distribution of users. From the table, it is seen that substantial control channel overhead saving (in power or bandwidth) can be achieved with knowledge of channel quality information. Although the addition of this information will correspondingly reduce the number of random preambles, the effect on collision probability can be minimized through careful partitioning of the user space.

F. Dedicated Preambles

In general, UE randomly selects a preamble from the available set for transmission with possible contention. However, several proposals have been made to reserve a subset of preambles for dedicated purposes such as handover, uplink synchronization, and feedback. Depending on the load and access pattern, this option may be more efficient than using random preambles. In this case, preambles are reserved and allocated to UEs so that they may be use in a contention-free manner. For example, during handover the target base station can allocate a specific preamble to a user making the transition so that the handover process can be performed in a contention-free manner. For some applications, it may be sufficient to simply reserve the preambles without having to allocate them to users. For instance, indicative feedback regarding broadcast service quality can be provided using just one common preamble because user-specific information is not required (i.e. the networks only need to know that some users are experiencing poor radio conditions, but not which users). In this case, responses from multiple users are transmitted using a common preamble. Some energy aggregation may be performed to further gauge the approximately number of responses. Naturally, with the use of dedicated preambles, capacity on the non-synchronized random access channel is correspondingly reduced and more time-frequency regions may be needed to maintain low collision probability for other random access users.

IV. PHYSICAL LAYER DESIGN

A. *Preamble Sequence Design*

Naturally, preamble waveforms for random access should have good detection probability while maintaining low false alarm rate, allow accurate timing estimation, and low power de-rating. Two promising sequences were analyzed - Zadoff-Chu with Zero Correlation Zone (ZC-ZCZ) [3] and Generalized Chirp-Like with Zero Correlation Zone (GCL-ZCZ) [4]. A sequence set with zero correlation zone property has zero periodic cross-correlation for a contiguous set of delays. Both sequences belong to a family of Constant Amplitude Zero Auto-Correlation (CAZAC) sequences. The constant amplitude results in low peak-to-average power ratio in the transmitter. This is especially important in the uplink where peak-to-average power ratio must be kept low due to power amplifier limitations.

The Zadoff-Chu sequence of length N is given by the expression

$$g_p(n) = \begin{cases} e^{-j\frac{2\pi}{M}\frac{1}{2}pn^2} & N \text{ is even} \\ e^{-j\frac{2\pi}{M}\frac{1}{2}pn(n+1)} & N \text{ is odd} \end{cases}$$

where p, the sequence index, is relatively prime to N (i.e. the only common divisor for p and N is 1). For a fixed p, the Zadoff-Chu sequence has ideal periodic auto-correlation property (i.e. the periodic auto-correlation is zero for all time shift other than zero). For different p, Zadoff-Chu sequences are not orthogonal, but exhibit low constant cross-correlation of $1/\sqrt{N}$ regardless of time shift. If the sequence length N is selected as a prime number, there are $N-1$ different available sequences. The zero-correlation zone for the Zadoff-Chu sequence is generated using cyclic shift version of the base carrier sequence. Note that each zero-correlation zone must be large enough to accommodate the maximum timing misalignment between mobiles in the cell, which is dependent on the cell size. Thus, the number of zero correlation zones that can be generated per each sequence index p is based on the sequence length N and the cell size. The maximum available number of sequences available in the system is then $(N-1)\times L$ where L is the number of zones per sequence index.

The Generalized Chirp-Like (GCL) sequence of length N is defined as

$$c(n) = g_p(n)b(n \bmod m), \quad n = 0,1,\ldots,N-1$$

with the sequence length N satisfying the relationship $N=sm^2$ where s and m are positive integers. Note that length constraint of the GCL sequence can be relaxed to $N=tm$ where t and m are positive integers [5]. The carrier sequence $g_p(n)$ is the Zadoff-Chu sequence of length N. As before, the sequence index p must be a relative prime to N. Due to the restriction on the sequence length, the number of available GCL groups is generally less than that for Zadoff-Chu. To provide a set of orthogonal GCL sequences, a common Zadoff-Chu sequence is modulated by m different orthogonal sequences $\{b_i(k)\}$, $i=0,\ldots,m-1$. An example of orthogonal modulation is the Hadamard sequence. The periodic cross-correlation between any two GCL sequence with the same Zadoff-Chu carrier is zero for all delay not equal to an integer multiple of sm. As a result, the resulting sequence has a zero correlation zone of length sm-1.

Since in general multiple ZC-ZCZ or GCL-ZCZ sequence sets may be needed to generate the required number of preambles, sequence performance will be affected by the cross-correlation performance between groups. For the ZC-ZCZ, the periodic cross-correlation between any two sequences is constant. For the GCL-ZCZ, the cross-correlation depends on selected indices. Figure 6 illustrates the normalized average cross-correlation at zero delay between the sequences when more than one Zadoff-Chu carrier sequences are used. From the figure, it is seen that when the number of groups is small, Zadoff-Chu and GCL have similar cross-correlation performance. However, as the number of groups increases, the cross-correlation of the GCL sequence is significantly worse than that of the Zadoff-Chu.

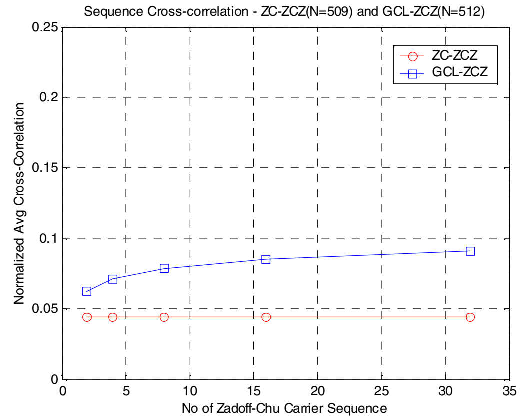

Figure 6. Normalized average cross-correlation between ZC-ZCZ and GCL-ZCZ.

Based on the analysis, sequence comparison may be summarized as follows –

- Both Zadoff-Chu and GCL can be designed to have the same number of zero-correlation zones based on cell range. In general, multiple root sequences may be needed within each cell which will introduce interference between sequences from different roots.
- The number of available GCL root sequences is less than that for Zadoff-Chu. Therefore, preamble planning with GCL will be more complicated.
- As the number the root sequences deployed within a cell increases, GCL cross-correlation between groups worsen. On the other hand, it remains the same for Zadoff-Chu.

As a result, Zadoff-Chu sequence with cyclic shift was selected as the preamble for E-UTRA. For the baseline preamble length of 800µs, this corresponds to a sequence of length 863 samples.

B. *Sequence Planning*

An important aspect of system design is sequence planning for random access. Based on the chosen sequence design, 862 different Zadoff-Chu sequences are available. For each sequence, a number of possible cyclic shifts are available based on the cell radius to be supported. However, the number of possible cyclic shifts should be restricted to a power of two so that 64 preambles can be exactly constructed from the appropriate set of Zadoff-Chu root sequences. Naturally, only one value of cyclic shift is used within a cell based on the supported cell size. Figure 7 illustrates the number of root Zadoff-Chu sequences required to provide 64 preambles based on the maximum cell radius.

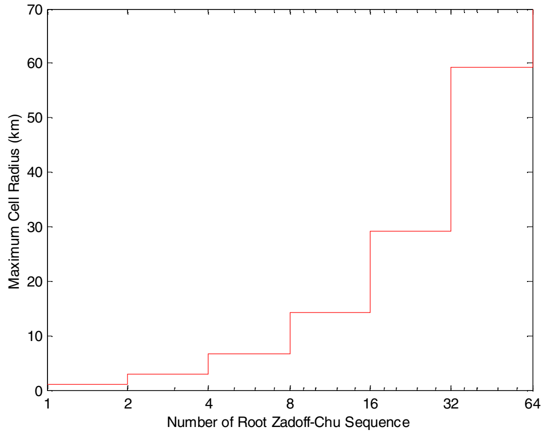

Figure 7. Number of root Zadoff-Chu sequences required.

From the figure, it is seen that in most cases sequence planning should not be a concern due to the large number of available Zadoff-Chu root sequences. For instance, even for 30 km cells, a sequence reuse factor of 53 is provided. Additional frequency reuse is possible by deploying random access regions in different frequency regions among cells.

C. Physical Layer Access Procedure

From the physical layer perspective, the L1 random access procedure encompasses successful transmission of the random access preamble and response. Open-loop power control with ramping is used in the preamble transmission in order to reduce interference. The following steps are required for the L1 random access procedure:

1. Prior to initiation of the non-synchronized physical random-access procedure, Layer 1 shall receive the following information from the higher layers - random access channel parameters, preamble format for the cell, number of root Zadoff-Chu sequences and sequence indices, power ramping step size, and the maximum number of preamble retransmissions.

2. A random access channel is randomly selected from the available non-synchronized random access channels. A preamble sequence is then randomly selected from the available preamble set based on the implicit message to be transmitted. The random function shall be such that each of the allowed selections is chosen with equal probability.

3. The initial preamble transmission power level (which is set by the RRC layer) is determined using an open loop power control procedure. The transmission counter is set to the maximum number of preamble retransmissions. Preamble transmission then occurs using the selected random access channel, preamble sequence, and preamble transmission power.

4. If no response corresponding to the transmitted preamble sequence is detected then another random access channel and preamble are randomly selected. If the maximum transmission power and the maximum number of retransmissions have not been reached, then preamble retransmission occurs. Otherwise the L1 status is passed to the higher layers and the physical random access procedure is terminated.

5. If a response corresponding to the transmitted preamble sequence is detected, then the L1 status is passed to the higher layers and the procedure is terminated.

V. CONCLUSIONS

In this paper, a preliminary look at the design of the random access channel and procedure for E-UTRA is provided. From the analysis, it is apparent that the E-UTRA requirements of low latency and overhead can be achieved. Specifically, the design provides a flexible approach where access regions can be added as random access load grows and where sequence planning can be tailored based on supported cell size. Random access overhead of 3-6% is seen to be sufficient for most system deployment scenarios. In addition, sequence planning should not be a concern due to the large number of available Zadoff-Chu root sequences. Finally, control plane latency of less than 100 ms is feasible with this random access procedure.

REFERENCES

[1] 3GPP TR 25.913, Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), v.7.3.0, March 2006.

[2] 3GPP TR 25.814, Physical Layer Aspects for Evolved UTRA, v.2.0.0, June 2006.

[3] D. C. Chu, "Polyphase codes with good periodic correlation properties," IEEE Trans. Inform. Theory, vol. 18, pp531-532, July 1972.

[4] B. M. Popovic, "Generalized Chirp-Like Polyphase Sequences with Optimum Correlation Properties," *IEEE Trans Inform. Theory*, vol. 38, no. 4, pp1406-1490, July 1992.

[5] R1-062128, "Generalized Construction of ZCZ-GCL Random Access Preambles," Huawei, RAN1#46, Tallinn, Estonia, Aug. 2006.

[6] R1-060009, "E-UTRA Downlink Control Channel Design and Performance and TP," Motorola, RAN1 LTE Ad Hoc, Helsinki, Finland, Jan. 2006.

[7] R1-061164, "Downlink Control Channel Modulation and Coding," Motorola, RAN1#45, Shanghai, China, May 2006.

[8] R1-061166, "Random Access Payload Size," Motorola, RAN1#45, Shanghai, May 2006.

[9] R2-062477, "Contention Resolution and Initial C-RNTI Allocation," Motorola, RAN2#54, Tallinn, Estonia, Aug. 2006.

[10] R2-062160, "RACH Contention and Retry Cases," NTT DoCoMo, RAN2#54, Tallinn, Estonia, Aug. 2006.

[11] R2-070268, "RACH Budget and Delay Analysis," NTT DoCoMo, RAN2#54, Tallinn, Estonia, Aug. 2006.

Note – 3GPP documents may be downloaded from ftp://ftp.3gpp.org