1

Basic Theory

Long-Term Evolution (LTE) is a new standard for radio access technology introduced by 3GPP. Its objective is to accommodate today's increasing demands for high-speed data services such as conversational voice, video and online gaming. LTE has the capability of moving towards fourth generation wireless systems. For consumers, LTE offers better, faster and packet-sized multimedia services. High-speed data over cellular networks will enable a rich suite of mobile multimedia services. Mobile phones and handheld devices are new media centers, with access to music, photos, games, video and a host of connectivity options (Table 1.1).

Benefits of LTE	
– High data rate	– Lower power requirement
– Enhanced mobility	– Lower deployment cost
– Low latency	– Greater Spectral Efficiency
– Multiple user support	– Flexibility of services
– Enhanced security	

Table 1.1. *Advantages of LTE technologies*

Emerging broadband wireless technologies (Figure 1.1), include Worldwide Interoperability for Microwave Access (WiMAX/IEEE 802.16), LTE, Long-Term Evolution-Advanced (LTE-A), High-Speed Packet Access (HSPA), HSPA+ and Cloud Radio Access Networks (C-RAN). WiMAX, LTE and

LTE-A are all intended to offer ubiquitous broadband at multiple megabits per second. Carriers across the USA and around the world are competing to build the next-generation of data networks to meet the ever increasing demand for fast Internet connections.

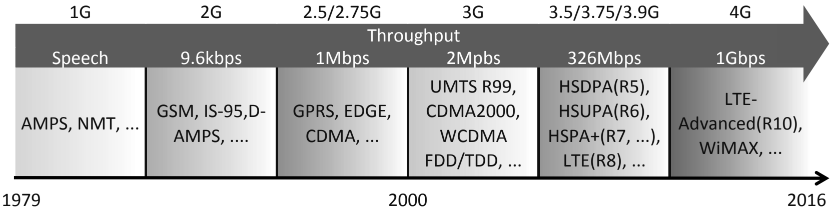

Figure 1.1. *Evolution of broadband wireless technologies*

1.1. LTE overview

Figure 1.2 depicts the basic architecture of LTE. The radio network architecture proposed by the 3GPP LTE consists of evolved NodeB (eNodeB). eNodeB provides a link between the user equipment and core network. As shown in Figure 1.2, eNodeB is connected to the core network (by means of MME/S-GW) via the S1 interface, and each eNodeB is interconnected via the X2 interface. The eNodeB is responsible for the majority of the radio resource management (RRM) functions such as packet scheduling. Both Mobility Management Entity (MME) and Serving Gateway (S-GW) are part of the core networks. The MME is responsible for paging and User Equipment (UE) mobility in idle mode within the network, while the S-GW node is responsible for routing user data packets and handling other user requests, for example handover.

LTE uses orthogonal frequency division multiple access (OFDMA) as a radio interface. OFDMA divides the bandwidth into subcarriers and assigns them to the users depending on the current demand of service. Each subcarrier carries data at low rate, but using multiple subcarriers at once to provide high data rates [RAM 09].

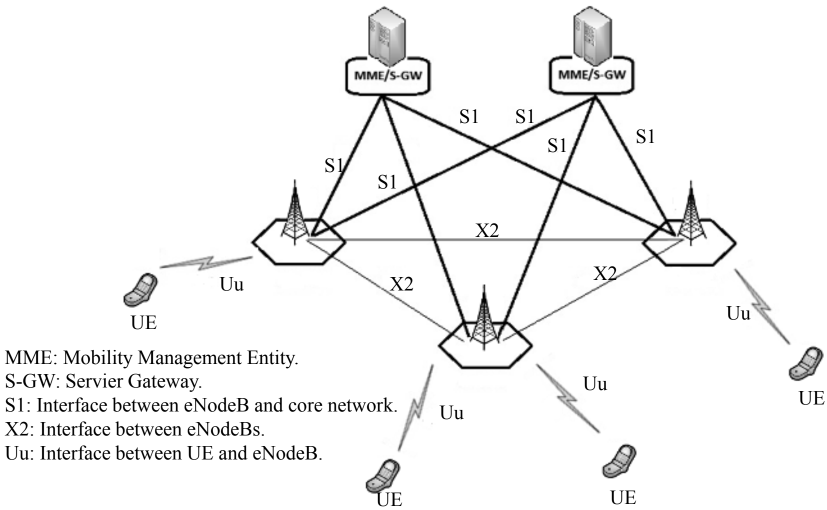

Figure 1.2. *LTE architecture*

Some advantages of OFDM: (1) OFDM uses the multiple carrier transmission technique. As a result, the symbol time becomes substantially larger than the channel delay spread. Hence, the effect of inter-symbol interference (ISI) is reduced significantly. In other words, against the multipath interference (frequency selective fading), the OFDM provides high robustness with less complexity; (2) OFDM uses Fast Fourier Transform (FFT) processing that allows us a low-complexity implementation; (3) OFDM offers complete freedom to the scheduler by using the OFDMA; and (4) OFDM provides the spectrum flexibility that helps smooth evolution from all the existing radio access technologies towards LTE.

However, there is a drawback to OFDM. The instantaneous transmit radio frequency (RF) power can change drastically within a single OFDM symbol. This can lead to high peak to average power ratio (PAPR) and requires costly transmission equipment consuming high power. Instead of OFDM, single carrier–frequency division multiple access (SC-FDMA) can be used at uplink to cope with a high PARP problem. SC-FDMA is also called Discrete Fourier Transform-spread Orthogonal Frequency-Division Multiplexing (DFT-spread OFDM) modulation. The main difference between SC-FDMA and OFDM is that SC-FDMA applies both FFT and inverse FFT (IFFT) onto both the transmitter side and the receiver side. In OFDM, on the other hand, FFT is applied to the receiver side, but IFFT to the transmitter side. Because of its inherent single carrier structure, SC-FDMA has lower PAPR, thereby decreasing the power consumption in the user terminal [MAR 09] while offering same advantages as OFDMA.

Figures 1.3 and 1.4 depict LTE's resource blocks (RBs) concept. RBs are the radio resources available for users. RBs are defined by frequency domain (Figure 1.3) and time domain (Figure 1.4). In the frequency domain (Figure 1.3), one RB is a collection of 12 contiguous subcarriers each consisting of 180 kHz bandwidth.

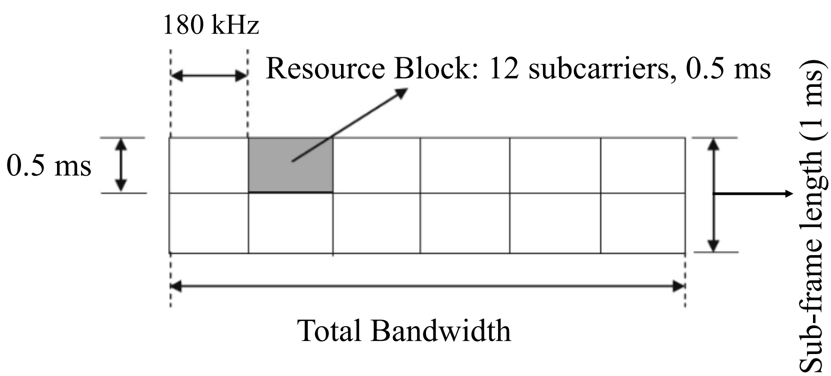

Figure 1.3. *Resource block grid structure in the frequency domain*

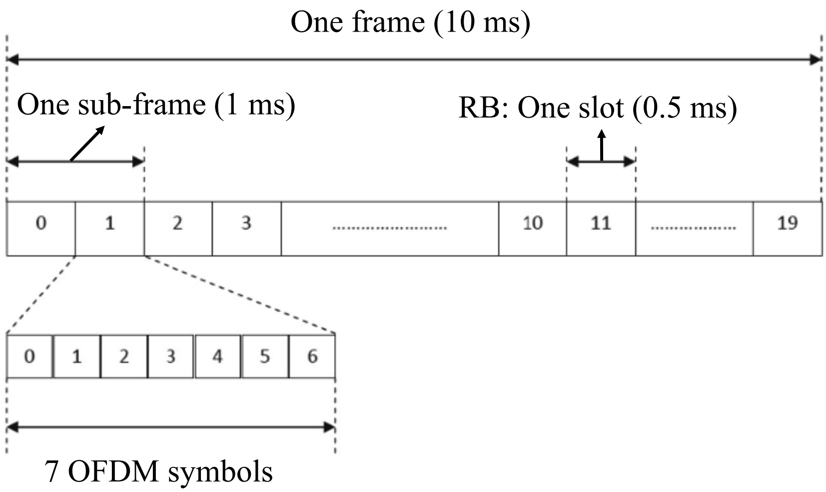

Figure 1.4. *Frame structure in the time domain*

In the time domain (Figure 1.4), each RB is defined as a 0.5 ms time slot [ZYR 07, DON 10]. Each time slot carries seven OFDM symbols. Two consecutive time domain RBs make one sub-frame, and the duration of one sub-frame is called the transmission time interval (TTI). Since one time slot is 0.5 ms, each sub-frame or TTI which contains two time slots has a duration of 1 ms (i.e. 0.5 ms time slot x 2). Ten sub-frames make one frame. Thus, one frame in LTE consists of 10 ms. On every TTI, each user reports its channel condition to its corresponding eNodeB. The report includes the received signal-to-noise ratio (SNR) of each subcarrier at the user side. These feedback reports also contain other radio parameter statuses perceived by the user, such as the channel quality indicator (CQI), rank indicator and user buffer status.

1.2. Scheduling in LTE

Scheduling is a process of allocating the physical resources among users. Since a wireless channel has time-varying behavior in comparison to a wired network, wireless and wired networks should have different schemes for resource allocation. Therefore, this section will discuss the characteristics of LTE scheduling.

The main aim of a scheduling algorithm is to maximize the overall system throughput while maintaining fairness, delay and packet loss rate within QoS requirements. In general, users are classified based on their traffic characteristics, such as real-time and non-real-time traffic. For real-time traffic (e.g. video, VoIP and gaming), scheduling must guarantee that QoS requirements are satisfied. The packet loss rate and delay play vital role in user experience. The real-time traffic packet must arrive at a user within the certain delay threshold. Otherwise, the packet is to be considered lost or discarded.

1.2.1. *Quality of Service parameters*

The scheduling decisions can be made on the basis of the following parameters:

– guaranteed bit rate (GBR) bearer guarantees a minimum bit rate for particular services such as multimedia services (e.g. VoIP and video);

– non-guaranteed bit rate (non-GBR) bearer does not require guaranteeing bit rate for particular services such as best effort services (e.g. File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP)).

The main quality of service (QoS) parameters at the bearer level (i.e. per bearer or per bearer aggregate) are QCI, allocation and retention priority (ARP), GBR and Aggregate Maximum Bit Rate (AMBR), which are defined as:

– *QoS class identifier (QCI)*: It is used to provide the information on how to perform packet-forwarding treatment (e.g. scheduling weights, admission thresholds, queue management thresholds and link layer protocol configuration) by using the specific node parameters. The operator preconfigures these parameters at the eNodeB. It has a number range from 1 to 9 based on the priority of various traffic services. The mapping of QCI weights onto QCI standardized characteristics is given in Table 1.2.

– *Allocation and retention priority (ARP)*: Its aim is to make decisions of call admission acceptance on the basis of available resources at the servicing eNodeB. In the case of exceptional resource limitations (e.g. at handover), ARP also plays an important role by considering the requested service to be dropped.

– *GBR*: GBR is the expected bit rate provided by a GBR bearer.

QCI	Resource type	Priority	Packet delay budget (ms)	Example services
1	GBR	2	100	Conversational voice
2		4	150	Conversational video (live streaming)
3		5	300	Non-conversational video (buffered streaming)
4		3	50	Real-time gaming
5	Non-GBR	1	100	IMS signaling
6		7	100	Voice, video (live streaming), interactive gaming
7		6	300	Video (buffered streaming) TCP-based (WWW, email), Chat, FTP, P2P file sharing
8		8	300	
9		9	300	

Table 1.2. *QCI values and their associated parameters*

LTE specifies a number of standardized QCI values with standardized characteristics, which are preconfigured for the network elements. This ensures multivendor deployments and roaming. The set of standardized QCIs and their characteristics is presented in Table 1.2. The QCI table

[PED 09] specifies values for the priority handling, acceptable delay budget and types of services for each QCI value.

– *QCI*: The QCI index identifies a set of QoS attributes (priority, packet delay and packet error loss rate).

– *Resource type:* The type of bearer indicating either a GBR bearer or non-GBR.

– *Priority:* A smaller number means higher priority at the scheduler.

– *Packet delay budget:* The maximum allowable packet delay for each bearer service.

– *Packet error loss rate:* The acceptable packet loss rate.

1.2.2. *Channel quality indicator*

To perform frequency-domain scheduling, the base station (BS), which is also called the eNodeB in LTE parlance, ideally needs to know the instantaneous channel state information (CSI) for all subcarriers for all users (UEs) in the cell. This channel information needs to be fed back to the BS by each user. In [GES 04], every user sends CSI for a subcarrier only if the subcarrier's channel gain is above a certain threshold. In [CHO 06], each user only indicates which n subcarriers have the best gains, and what their gains are. In [SAN 07], a 1 bit feedback scheme is shown to be asymptotically optimal in terms of capacity. Even more drastic feedback reduction techniques are resorted into a practical system such as LTE, where CSI is quantized into a 4 bits' value called channel quality indicator (CQI). The 4 bits' values of CQI report is generated by the user on the basis of received Signal to Interference-plus-Noise Ratio (SINR) on the downlink and feedback to the corresponding eNodeB.

A suitable modulation and coding scheme (MCS) is chosen on the basis of the channel condition reported by the user to meet the QoS requirement. In the case of a sub-band CQI reporting scheme, if the CQI value of a specific sub-band becomes lower than compared to previous value, then the eNodeB will choose a lower order MCS for that sub-band. In addition, if the CQI value of a specific sub-band becomes higher than compared to previous value, then the eNodeB will choose a higher order MCS for that sub-band [DON 10].

There are two methods of CQI feedback reporting; periodic feedback and aperiodic feedback. In the periodic feedback method, a user periodically sends a CQI report to the eNodeB on every TTI. This method only allows wideband and user selective feedbacks. In aperiodic feedback method, the eNodeB first asks for a CQI report from a specific user. In response, the user sends the feedback to the eNodeB [DON 10].

The list of different modulation and coding schemes on the basis of CQI index is presented in the Table 1.3. The values are based on [36. 10].

CQI index	Modulation	Effective coding rate = $\frac{c_r}{e_r}$ * 1,024	Spectral efficiency = $\frac{R_b}{B}$
0	out of range		
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

Table 1.3. *4 bits CQI values*

The parameters of CQI (Table 1.3) are defined as follows

– *CQI index:* This identifies the set of CQI attributes (modulation, effective coding rate and spectral efficiency).

– *Modulation:* This defines the type of modulation scheme being used with the corresponding CQI index.

– *Effective coding rate:* This is the ratio of code block bits (c_r) to rate matching bits (e_r) by multiplication of 1,024.

– *Spectral efficiency:* This is the ratio of bit rate (R_b) to the bandwidth of the channel (B).

1.2.3. *Buffer state and resource allocation history*

Buffer State refers to how much data are available for the user to be scheduled by eNodeB (Figure 1.5). The scheduler must take into account the time duration of the data queued in the buffer at eNodeB. Packet delay and dropping rate can be minimized by giving higher priority to the users who have data in queues for longer periods. The priority also depends on the basis of other QoS parameters such as QCI [RAM 09].

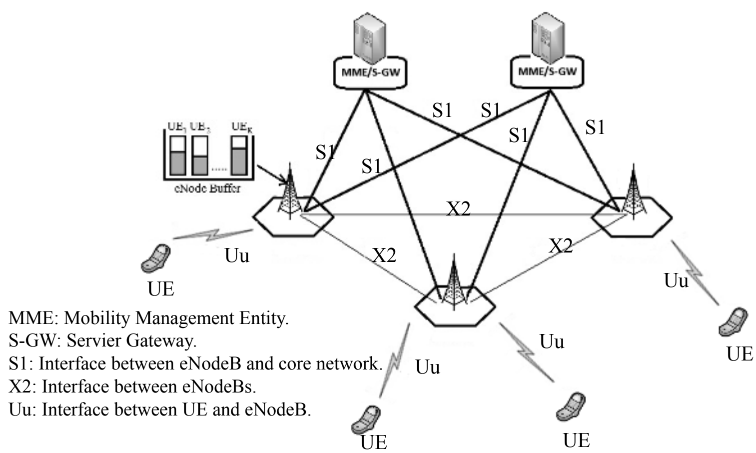

Figure 1.5. *Generalized packet scheduling model 3GPP LTE system*

Another important factor for scheduling is the resource allocation history of users. Consider a situation in which a user has higher priority in the previous sub-frame but its priority may be lower in the current sub-frame. The more general approach is to update the average data rate of users on every TTI and try to maintain average data rate of users that are currently being served [SAD 09].

1.3. LTE Traffic measurements

In this section, we present an empirical measurement of LTE network traffic. Using real LTE equipment in an actual application in daily life (Skype Voice, Skype Video, YouTube and Live FootBall Match), we measured the inter-arrival time (IAT) of packets to characterize network behavior of LTE (Figures 1.6–1.9). To the best of our knowledge, this is the first report on the empirical research describing network traffic behavior of LTE/LTE-Advanced.

Network testing and performance measurement are important for assessing, maintaining and trouble-shooting network services. Accurate and realistic measurement, as a result, assures provisioning of the acceptable level of QoS and user experience (Quality of Experience). Especially, for designing an efficient DRX mechanism in LTE/LTE-Advanced, the characteristics of real LTE network traffic behavior need to be taken into account. Therefore, proper performance assessment is a necessary step towards development and deployment of LTE/LTE-Advanced.

Typically, modeling and simulation are used to access such a complex system as a network behavior. Although modeling and simulation are robust and time/cost-effective tools, they cannot replace empirical testing. Empirical data from actual measurement are valuable as they provide reference values, which would also serve as a standard for designing/

developing modeling and simulation tools for more complex network architecture.

1.3.1. *Testing environment*

For data collection, we were provided with the Samsung™ GT-B3730 USB modem launched by TeliaSonera®. It is the 4G/3G/2G USB modem used for both LTE and UMTS connectivity and all the raw data were collected via this device. Samsung GT-B3730 switches automatically between 3G and 2G, but for LTE a reconnection is required to switch between 3G and 4G.

This USB modem supports LTE (2,600 MHz), 3G HSPA/UMTS (2,100 MHz) and 2G/Enhanced Data rates for GSM Evolution (EDGE) (900/1,800 MHz) with the peak speed up to 100/50 Mbps, 17/5.76 Mbps and 296/107 Kbps DL/UL, respectively. The device supports Windows, Ubuntu, Linux and Mac operating systems.

We collected the measurements for both LTE and UMTS by connecting the device to a Dell desktop computer having an Intel Core 2 Duo 1.6 GHz processor and 1 GB RAM and running different kinds of applications using the Internet. Wireshark Network Protocol Analyzer version 1.6.7 was used to capture network traces for Skype Voice and Video calls, live Web streaming and YouTube streaming.

For Skype data, version 5.10.0.114 was installed on the desktop computer and the other user, having a Dell Core i3 laptop with 2 GB RAM and running the same Skype version, was called and the data were recorded. The collection of data was made on the desktop PC while the streaming data were gathered by running different kinds of streams through the Web browser on the same PC. The observations for the applications were gathered by running them one by one and all other programs and software were stopped during the data collection, which use Internet for communication.

Figures 1.6–1.9 show empirical traffic measurement of live Skype Voice call (Figure 1.6), Skype Video call (Figure 1.7), Post Video (Figure 1.8) and Live Video Streaming (Figure 1.9), respectively.

The data were analyzed through non-parametric distribution for each case since none of the data set was normally distributed. In all the graphs, the x-axis represents IAT in logarithm, (IAT(log(N)), while the y-axis is the probability distribution function (PDF), which is the ratio of the relative frequency to the bin size or the class interval. Furthermore, the traffic data in the graphs contain both UL and DL.

1.3.2. *VoIP preliminary capacity*

There are seven visible peaks in Figure 1.6. This indicates that most of the data streams follow the IATs pointed to by these peaks; peak IAT of 0.017, 0.050, 0.98, 1.85, 2.84, 10.2 and 19.4 ms. It is also indicative of the presence of seven different streams of traffic associated with Skype Voice samples, for example signaling, data and ACKs.

Among these seven peaks, two solely represent the UL traffic with IAT peaks of 0.017 and 0.050 ms. The remaining peaks signify the combination of both UL and DL traffics with IAT peaks of 0.98, 1.85, 2.84, 10.2 and 19.4 ms.

Based on the cumulative distribution function, the median of the IAT is at 8.5 ms and 70% of the traffic arrived between 2.6 and 22.3 ms. In logarithmic scale, traffic distribution of the live Skype Voice call is slightly skewed to the right and concentrated in a relatively short span.

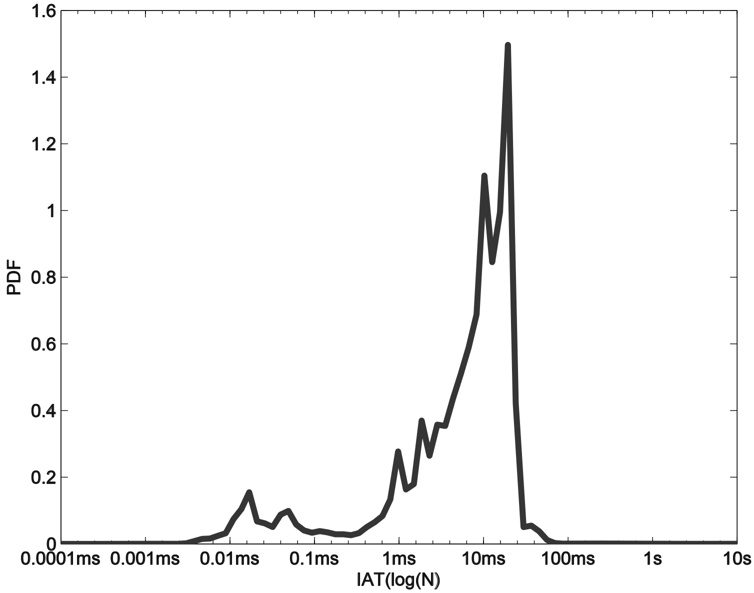

Figure 1.6. *LTE Voice conversation*

1.3.3. *Video conversation preliminary capacity*

There are five visible peaks in Figure 1.7. This indicates that most of the data streams follow the IATs pointed to by these peaks; peak IAT of 0.11, 0.35, 0.90, 1.94 and 10.7 ms. It is also indicative of the presence of five different streams of traffic associated with Skype Video samples, for example signaling, data and ACKs.

Among these five peaks, two represent the UL traffic with IAT peaks of 0.35 and 1.94 ms. The remaining peaks signify the DL traffic with IAT peaks of 0.11, 0.90 and 10.7 ms.

Based on the cumulative distribution function, the median of the IAT is at 1.1 ms and 90% of the traffic arrived between 0.078 and 15.7 ms. In other words, the data traffic of the live Skype Video call was clustered in a short window of IAT.

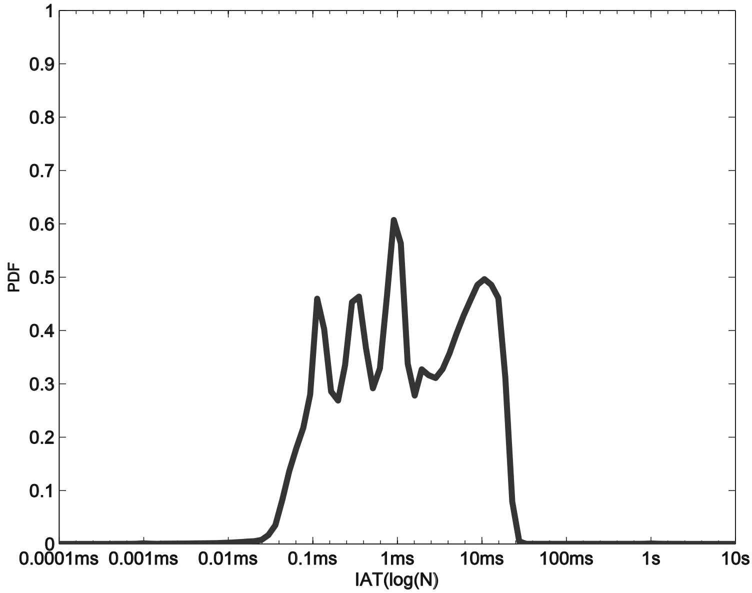

Figure 1.7. *LTE Video conversation*

1.3.4. *Post video and live video preliminary capacity*

There are 11 visible peaks in Figure 1.8. This indicates that most of the data streams follow the IATs pointed to by these peaks; peak IAT of approximately 0.008, 0.012, 0.017, 0.03, 0.08, 0.14, 0.91, 1.61, 8.87, 33.5 and 71.5 ms.

Among these 11 peaks, two solely represent the UL traffic with IAT peaks of 0.012 and 0.08 ms and two of them are solely the DL traffic with the peaks of 0.017 and 0.14 ms.

The remaining peaks signify the combination of both UL and DL traffics with IAT peaks of 0.008, 0.03, 0.91, 1.61, 8.87, 33.5 and 71.5 ms.

Based on the cumulative distribution function, the median of the IAT is 0.027 ms and 70% of the traffic arrived between

0.008 and 0.07 ms. Given that majority of the data traffic of Post Video arrived between 2.6 ms and 22.3 ms with the median of 8.5 ms, data traffic arrival of the Post video is concentrated in relatively early IAT. In a logarithmic scale, traffic distribution of the Post Video is slightly skewed to the left and concentrated in a relatively short span.

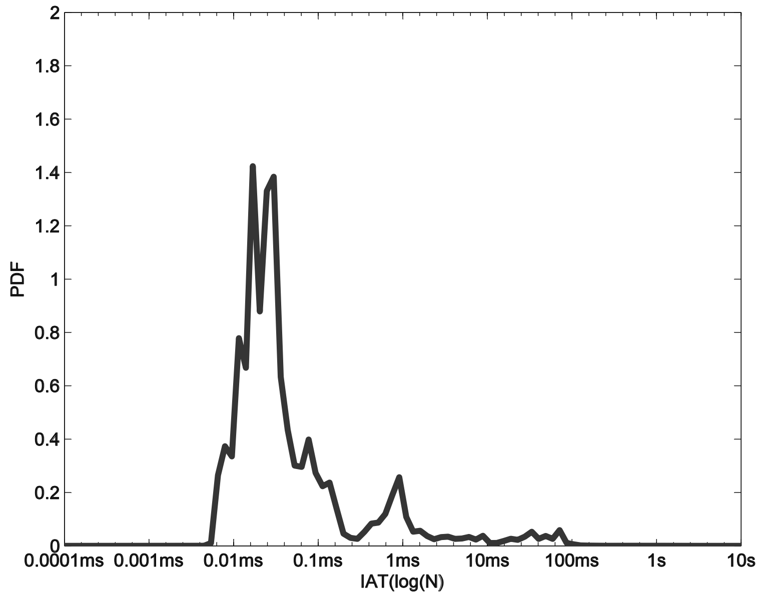

Figure 1.8. *LTE Post Video*

There are nine visible peaks in Figure 1.9. This indicates that most of the data streams follow the IATs pointed to by these peaks; peak IAT of approximately 0.004, 0.008, 0.031, 0.045, 1.13, 5.15, 9.10, 50.2 and 129.6 ms. It is also indicative of the presence of nine distinctive streams of traffic associated with Live Video streaming samples, for example signaling, data and ACKs.

Among these nine peaks, three solely represent the UL traffic with IAT peaks of 0.004, 0.031 and 0.045 ms. The remaining peaks signify the combination of both UL and DL traffics with IAT peaks of 0.008, 1.13, 5.15, 9.10, 50.2 and 129.6 ms.

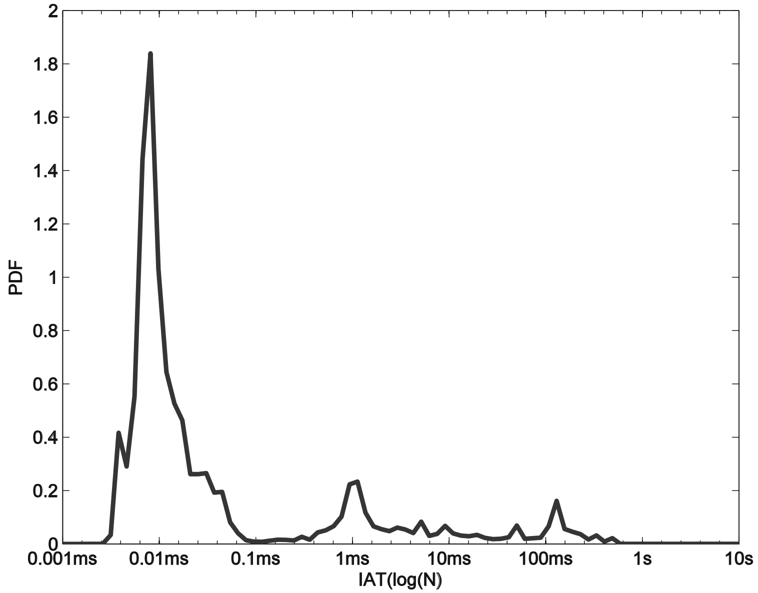

Figure 1.9. *LTE Realtime Video*

Based on the cumulative distribution function, the median of the IAT is 0.009 ms and 75% of the traffic arrived between 0 and 0.029 ms. The next 20% scattered through IAT of approximately 0.03–44 ms. This indicates that traffic distribution of the Live Video streaming is largely skewed to the left in a logarithmic scale. In other words, majority of the Live Video streaming traffic arrives as soon as IAT reaches 0.029 ms.

1.3.5. *Summary on the LTE Traffic measurements*

The raw data in units of time were quantized by means of PDFs for analysis and visualization of the data. The PDF graphs initially generated through Matlab simulations were expressed in discrete time, which then were converted into continuous time by applying non-parametric distribution fit. This way, information about the packet sequence of the traffic was obtained. Using such information, we can classify IATs into different sets or groups of time assuming a particular Gaussian distribution corresponding to the each significant peak of IATs. However, the peak analysis was beyond the scope of this chapter due to the huge volume of empirical data.

An in-depth peak analysis is going to be the next step of our future work. The analysis of individual peaks will specify the number and type of traffic sources (e.g. control or data traffic) composing each peak. For further advanced analysis, a Gaussian peak can be assigned to each peak based on the Gaussian distribution parameters, i.e. mean and standard deviation, to map the detailed traffic behavior. Subsequently, the results can be utilized for protocol modifications for mathematical modeling. Since mathematical modeling of the data traffic is effective in analysis of the specific parts of the data in a distributed manner, the knowledge obtained through the process is especially useful for designing an efficient DRX mechanism in LTE/LTE-Advanced.

1.4. User equipment power saving in LTE

1.4.1. *DRX cycle*

Discontinuous Reception (DRX) is a power-saving technique in UE. Simply put, it switches off the receiver and puts the device into power-saving mode when it is not in use.

Turning off the receiver and switching between active and sleep modes, *per se*, is not a new idea in telecommunication systems. The basic concept of DRX has already been applied to 2nd generation (2G) systems, for example. Global System for Mobile communications (GSM). LTE and LTE-Advanced specification have adopted DRX at the link level [BO 10]. The main difference between LTE DRX and previous versions is that the UE is allowed to enter a sleep state even when the traffic buffer is not completely empty. LTE DRX also employed the three-state model that consists of wake, light sleep and deep sleep, whereas old version only has only two states: wake and sleep. These new features of LTE DRX are intended to improve power saving significantly, but at the same time they also create new challenges regarding delay and other QoS issues. We present these trade-offs in Chapter 2.

The basic DRX mechanism works as follows. Each UE is assigned a periodic wake (ON) period to sense packets through PDCCH. At other times, UE turns its receiver off and goes into power-saving mode. Only if packets are sensed, UE remains ON so that the packets will be transferred. Otherwise, UE goes back to sleep mode until the next periodic wake period to check the packet arrival. This way, power consumption status and battery life at the UE should be improved. Note that, in DRX, downlink data transfer happens only during awake (ON) time. Therefore, the trade-off between power saving and network traffic flow (thereby provisioning of QoS) is an important factor when applying the DRX cycle efficiently.

Functionality of DRX is managed by the Radio Resource Control (RRC) reference. In RRC_CONNECTED state, RRC controls the sleep/wake scheduling of each UE by configuring the following parameters: DRX inactivity timer, DRX short cycle, DRX long cycle and DRX short cycle timer. The details of the parameters are explained in Figures 1.10–1.14 and Table 1.4.

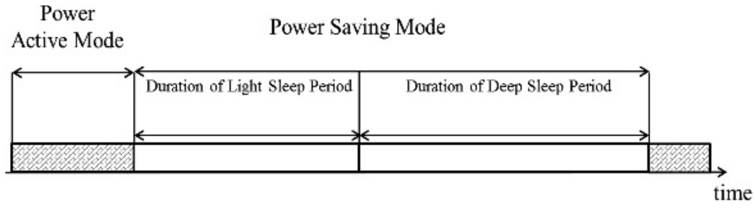

Figure 1.10. *General 3GPP LTE DRX model*

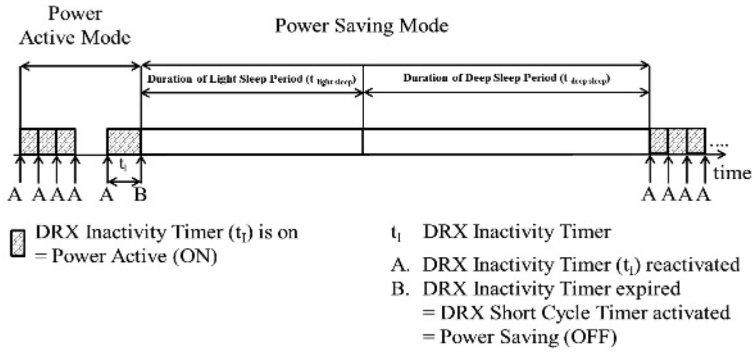

Figure 1.11. *General 3GPP LTE DRX model for t_I*

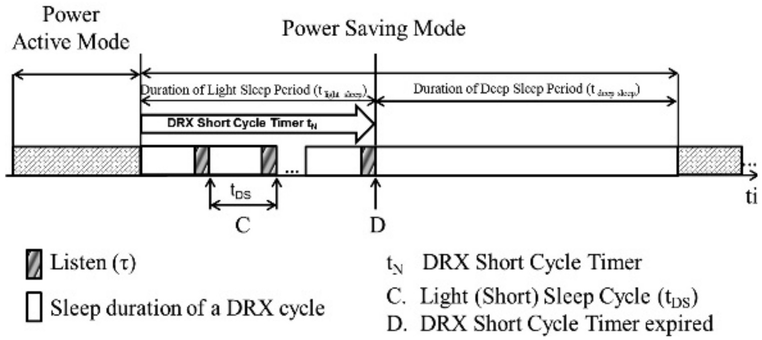

Figure 1.12. *General 3GPP LTE DRX model for light sleep*

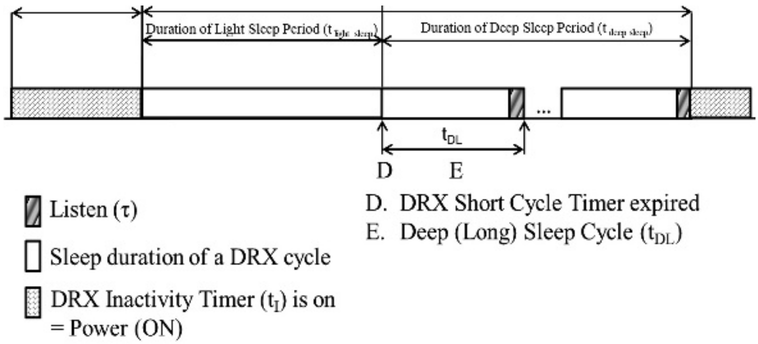

Figure 1.13. *General 3GPP LTE DRX model for deep sleep*

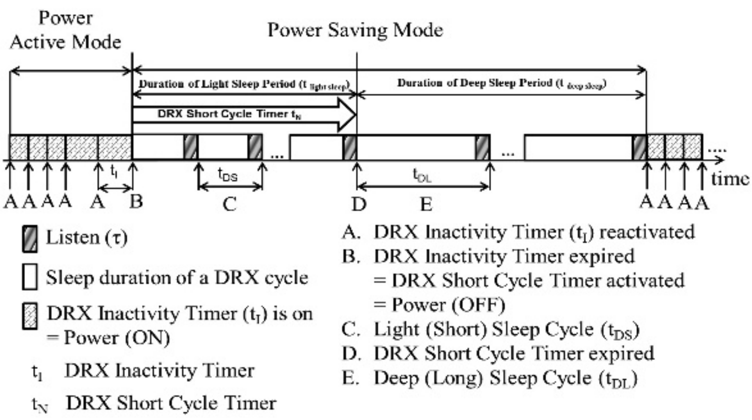

Figure 1.14. *Complete general 3GPP LTE DRX model*

The basics of LTE's power-saving mechanism in DRX are depicted in Figures 1.10–1.14.

Figure 1.10 depicts a simplified architecture of the DRX. There are two distinctive modes in the DRX: power active mode and power-saving mode. Power-saving mode can be either light sleep mode or deep sleep mode.

Values	Explanation
DRX inactivity timer (t_I)	It defines how long the UE must remain awake (ON) when UE find out that there are no packets scheduled after periodic monitoring of PDCCH. Note that, in DRX mechanism, when DRX inactivity timer (t_I) is on (sometimes referred to as *inactive period*, confusingly), then UE is awake (ON).
Light sleep mode	It consists of a series of light (short) sleep cycles (C in Figure 1.12).
Light sleep cycle (t_{DS})	One light (short) sleep cycle consists of a short sleep-duration (i.e. power OFF) plus a listen-period (τ).
Listen-period (τ)	It is periodic awake for detecting any data activities, that is UE monitoring PDCCH for UL grant reception. Once data activity is detected, UE wakes up and DRX inactivity timer (t_I) is activated. While the duration of τ is generally consistent between the long and short sleep cycles, the sleep duration differs between the two cycles.
DRX short cycle timer (t_N)	It defines how many light (short) sleep cycles should be repeated before it transit into deep sleep mode.
Deep sleep mode	It consists of a series of deep (long) sleep cycles (E in 1.13). Similarly to the light sleep mode, if data activities are detected during listen-period (τ), UE wakes up and DRX inactivity timer (t_I) is reactivated, which is a transition from deep sleep mode to power active mode.
Deep (long) sleep cycle (t_{DL})	One deep (long) sleep cycle consists of a long sleep-duration (i.e. power OFF) plus a listen-period (τ). DRX long sleep cycle (t_{DL}) or DRX short sleep cycle (t_{DS}) specifies the duration of sleep for the respective cycle, and the values are fixed. As the term indicates, a t_{DL} is longer than a t_{DS}. Since there is no specification for a timer to define the duration of deep (long) sleep mode, t_{DL} repeats until UE finally detects a downlink transmission during listen-period (τ).

Table 1.4. *Summary of the DRX values*

Figure 1.11 is a closer look of the power active mode. DRX inactivity timer (t_I) plays an important role in power active mode. t_I determines how long the UE must remain awake (ON) when the UE finds out that there are no packets scheduled after a periodic monitoring of PDCCH. There are two possible scenarios in the power active mode. One is "Power ON duration of a DRX cycle due to traffic arriving", which means that packets are arriving and the UE is processing them. In this case, t_I will be reactivated (A) for another round, and UE stays awake (ON). The other is "Power ON duration of a DRX cycle but no traffic", which means that it is in power active mode since t_I is still active, but there are no packets arriving before the t_I has expired. In this case, t_I will be eventually expired (B in Figure 1.11) after a certain period of time, and subsequently DRX short cycle timer is activated (B in Figure 1.11) and the UE transits into light sleep mode.

Note that, in the DRX mechanism, when DRX inactivity timer (t_I) is on (sometimes referred to as inactive period, confusingly), then UE is awake (ON).

Figure 1.12 shows the details of the power saving mode. What is unique about DRX in 4G LTE/LTE-Advanced is its two-state power-saving mode: light sleep and deep sleep. Light sleep mode consists of a series of light (short) sleep cycles (C in Figure 1.12). One light (short) sleep cycle consists of a short sleep-duration (i.e. power OFF) plus a listen-period (τ). τ is the period awake for detecting any data activities, that is UE monitoring PDCCH for UL grant reception. Once data activity is detected, UE wakes up and t_I kicks in. In other words, it transits from light sleep mode to power active mode.

If no data activities are detected during τ, then another DRX sort cycle follows. It continues until the DRX short cycle timer expires and transits into deep sleep mode (D in Figure 1.12). The short DRX cycle timer defines how many light (short) sleep cycles should be repeated before it transits into deep sleep mode.

Table 1.4 summarizes key elements of LTE DRX mechanisms discussed above.

Deep sleep mode consists of a series of deep (long) sleep cycles (E in Figure 1.13). One deep (long) sleep cycle consists of a long sleep-duration (i.e. power OFF) plus a listen-period (τ). DRX long cycle (t_{DL}) or DRX short cycle (t_{DS}) specifies the duration of sleep for the respective cycle, and the values are fixed. As the term indicates, a t_{DL} is longer than a t_{DS}. While the duration of τ is generally consistent between the long and short sleep cycles, the sleep duration differs between the two cycles. Thus, in deep sleep mode, there are longer intervals from one τ to the next τ to detect data activities in UL/DL.

Similar to the light sleep mode, if data activities are detected during τ, UE wakes up and t_I is reactivated, which is a transition from deep sleep mode to power active mode. If no data activities are detected during τ, then another t_{DL} follows and UE stays sleeping. Since there is no specification for a timer to define the duration of deep sleep mode, t_{DL} repeats until UE finally detects a downlink transmission during τ.

Figure 1.14 shows the whole picture of the DRX architecture described above. Note that the duration length of the DRX cycle affects major QoS parameters, for example delays. A long DRX cycle can result in delay because of the longer intervals between the τ (listen) periods. It is due to the fact that the eNobeB processor will not transmit any packets to the UE during the sleep period (whether it is long sleep or short sleep). It means that packets have to wait until the next τ (listen) so that the UE will detect the packets and transit from power-saving mode to power active mode, and then finally start data transmission. A short DRX cycle, on the other hand, eliminates the chance of delay but is less effective for energy saving. Therefore, the trade-off between the energy efficiency and DRX sleep cycles is an important research topic.

1.5. Models for LTE power saving

To conduct a performance analysis of a DRX power-saving mechanism, two major power consumption reference models have been used as a generic prototype: the three-state generic model and the Nokia model. The reference models can be applied for analytical measurement or simulation. This chapter summarizes the characteristics of these models.

1.5.1. *3GPP power consumption model*

As detailed in the timing diagrams (Figures 1.10–1.14), the 3rd Generation Partnership Project (3GPP) DRX mechanism consists of three states: active, light sleep and deep sleep. The three-state generic model depicts the transition of the three states in a very simple manner as shown in the state transition diagram (Figure 1.15).

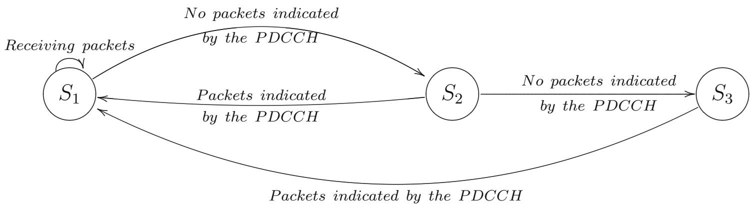

Figure 1.15. *A three-state semi-Markov process for DRX analysis*

– State 1 (S_1) indicates active state. In this state, UE is power active, and thus it includes a series of DRX inactivity timer activated periods, and corresponds to a whole packet call transmission. The state can either remain in S_1 or transit into another state, S_2, depending on the packet delivery status indicated by PDCCH.

– State 2 (S_2) indicates light sleep state. It includes a light sleep period that is entered from S_1. S_2 could either go back to S_1 or transit into another state, S_3, depending on the packet delivery status indicated by PDCCH.

– State 3 (S_3) indicates deep sleep state. It includes a deep sleep period that is entered from S_2. S_3 could either remain in S_3 or go back directly to S_1, depending on the packet delivery status indicated by PDCCH.

This reference model can be applied for statistical modeling or simulation. For example, Figure 1.15 illustrates a semi-Markov process for analytical study of the energy efficiency of the LTE DRX mechanism [ZHO 08]. An embedded Markov chain can be obtained if we view the semi-Markov process only at the times of state transitions. State transition probabilities can be expressed as $P_{i,j}$, where $i, j \in$ {1, 2, 3}

– $P_{1,1}$ refers to the probability of continuously receiving packets, and thus staying at S_1.

– $P_{1,2}$ refers to the probability of no packets indicated by PDCCH, and thus transiting into S_2.

– $P_{2,1}$ refers to the probability of packet arrival informed by PDCCH while in S_2, and thus going back to S_1.

– $P_{2,3}$ refers to the probability of no packets scheduled for a longer period while in S_2, and thus transiting into S_3.

– $P_{3,1}$ refers to the probability of packet arrival informed by PDCCH while in S_3, and thus going back to S_1.

The details of the power-saving performance analysis using statistical model will be discussed in Chapter 2.

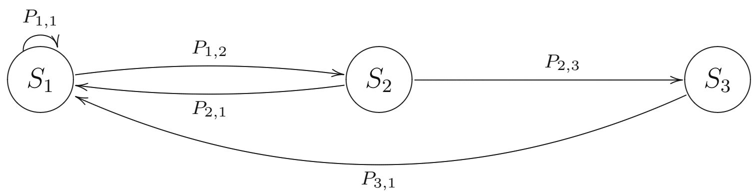

Figure 1.16. *A three-state semi-Markov process for DRX analysis*

1.5.2. *Characteristics of Nokia^{TM} power consumption model*

Nokia proposed a general state model for the UE power-saving (Figure 1.17). As Nokia states, it is a baseline model and needs further discussion within the 3GPP community for improvement and implementation [R20 07]. In the following section, we will list the characteristics of the

Nokia model. Some parts are similar to the general 3GPP model and others are Nokia specific.

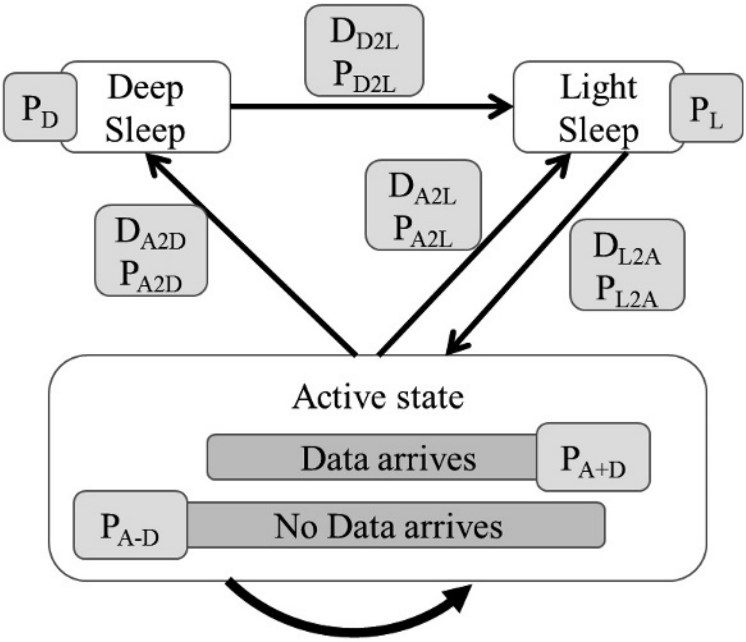

Figure 1.17. *Nokia LTE DRX model*

The Nokia model also utilizes three states of the UE: active, light sleep and deep sleep. In active states, the UE reads allocation information every TTI and is ready to transmit and receive upon scheduling. In deep sleep, DRX/DTX is longer and the UE should be able to save power more than in light sleep.

The approach is to identify the power consumption for each of the three states when in that state and obtain the average power consumption for each state.

For sleep states:

– P_D is the average power consumption at deep sleep.

– P_L is the average power consumption at light sleep.

The active state is divided into two different modes:

– P_{A+D} indicates that the UE is active and receive downlink data during that TTI, and thus needs to read the complete TTI information.

– P_{A-D} indicates that the UE is active, but does not receive downlink data during that TTI, and therefore there is no need to read the complete TTI information.

In addition, the transition time and its associated power consumption from one state to another are considered to be separate parameters. For example:

– D_{D2L} indicates the transition time from deep sleep to light sleep.

– P_{D2L} indicates the associated average power consumption during transition time D_{D2L}.

The same rule applies for other state transitions (D_{A2L} and P_{A2L}, D_{L2A} and P_{L2A}, D_{A2D} and P_{A2D}, and D_{A2L} and P_{A2L}).

There is an important assumption that the transitions between light sleep and deep sleep are not needed because fast transition directly from active to deep sleep is possible. (Thus there is no D_{L2D} and P_{L2D} or D_{D2L} and P_{D2L}.)

Another important feature of the Nokia model is that for the transition from deep sleep to active state, it has to go through the light sleep intermediate step, regardless of whether it is transition to the active with data transition or to the active without data transition.

Note that there will be a time resolution difference between the above state diagram and the actual simulator timing. Therefore, for simulating/modeling, all transition times above should be an integer number of sub-frames or TTIs, and then the associated power should be scaled accordingly. This way, the Nokia model may be adopted into existing link and system simulators used for LTE evaluation in a relatively straightforward manner.

1.6. Conclusion

The assessment of multi-user versus single-user is another issue on which Nokia model is focusing, since the trade-off between system/multi-user performance and single-user performance is a challenge for the DRX/DTX.

Implementation of the DRX/DTX concept should not sacrifice the system/multi-user performance, such as sufficient scheduling and adaptation flexibility. Otherwise, the DRX/DTX concept will not be attractive enough for multi-users, and the system cannot afford to provide UE power savings. At the same time, DRX/DTX concept should not sacrifice the system spectral efficiency. If it is impaired, then the loss directly returns to the individual users, consequently because UE needs to be awake for a longer period for receiving the same fixed amount of the data.

Thus, DRX/DTX concepts should be studied at system level, and the trade-offs between system/multi-user performance and single-user performance have to be taken into account, for example:

1) trade-off between single-user throughput and/or latency performance and the average UE power consumption for a fixed traffic model;

2) system capacity performance, such as the performance of several supported users and average throughput for some QoS criteria and;

3) the impact of associated control signaling errors. For these system-level assessments/evaluations of DRX concept, the use of statistical models is an effective strategy.

1.7. Bibliography

[36. 10] 36.213 G.T., "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures", vol. V9.3.0 (Release 9), available at www.3gpp.org.

[BO 10] BO H., HUI T., LAN C., *et al.*, "DRX-aware scheduling method for delay-sensitive traffic", *IEEE Communications Letters*, vol. 14, no. 12, pp. 1113–1115, December 2010.

[CHO 06] CHOI Y.-J., BAHK S., "Selective channel feedback mechanisms for wireless multichannel scheduling", *IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 289–300, June 2006.

[DON 10] DONTHI S., MEHTA N., "Performance analysis of subband-level channel quality indicator feedback scheme of LTE", *IEEE National Conference on Communications (NCC)*, January 2010.

[GES 04] GESBERT D., ALOUINI M.-S., "How much feedback is multi-user diversity really worth?", *IEEE International Conference on Communications (ICC)*, pp. 234–238, June 2004.

[MAR 09] MARTÍN-SACRISTÁN D., MONSERRAT J.F., CABREJAS-PEÑUELAS J., *et al.*, "On the way towards fourth-generation mobile: 3GPP LTE and LTE-advanced", *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1-10, March 2009.

[PED 09] PEDERSEN K., KOLDING T., FREDERIKSEN F., *et al.*, "An overview of downlink radio resource management for UTRAN long-term evolution", *IEEE Communications Magazine*, vol. 47, no. 7, pp. 86–93, July 2009.

[R20 07] R2-071285 N., "3GPP TS 36.300", vol. 3GPP TSG RAN WG2 LTE, March 2007, available at www.wireless-world-research.org.

[RAM 09] RAMLI H.A.M., BASUKALA R., SANDRASEGARAN K., *et al.*, "Performance of well known packet scheduling algorithms in the downlink 3GPP LTE system", *IEEE 9th Malaysia International Conference on Communications (MICC)*, pp. 815–820, December 2009.

[SAD 09] SADIQ B., MADAN R., SAMPATH A., "Downlink scheduling for multiclass traffic in LTE", *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 14, pp. 1–18, March 2009.

[SAN 07] SANAYEI S., NOSRATINIA A., "Opportunistic downlink transmission with limited feedback", *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4363–4372, 2007.

[ZHO 08] ZHOU L., XU H., TIAN H., *et al.*, "Performance analysis of power saving mechanism with adjustable DRX cycles in 3GPP LTE", *IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, September 2008.

[ZYR 07] ZYREN J., "Overview of the 3GPP long term evolution physical layer", available at ftp://ftp2.ie.netbsd.org/mirrors/download.sourceforge.net/pub/sourceforge/n/project/ns/ns3-lte/HARQ/Documents/State%20of%20the%20Art/LTE/314-69330.pdf, July 2007.