QoS-aware centralized radio resource management for TDMA-based heterogeneous application scenarios

Master Thesis

Ruta Tamasauskaite

Aalborg University
Computer Science

Abstract

A system that facilitates multiple heterogeneous devices is able to efficiently use the radio resources, has a larger capacity of supported devices and meets their diverse requirements. This study is based on the RTX A/S use-case to analyze requirements of the Quality of Service (QoS)-aware system and to provide the design approaches for its implementation. The radio resource management is centralized and executed by the base stations in order to keep the user equipment as simple as possible. In this thesis we provide multiple design choices that can be made and describe them in detail. We also compare and evaluate some of these decisions using the Uppaal modeling and verification tool. We aimed to create a realistic behavior templates of different device versions through the collaboration with the engineers at RTX A/S. The improvements are suggested starting from two ways of the Sleep mode implementation, dynamic frame size with QoS-aware scheduler maintaining profile-based priority queues, load balancing between the base stations controlling the handover and strategy based scheduling reducing transmission preparation synchronization issues. Finally, we propose a frame structure without a division to transmission and re-transmission sub-frames as conceptually each type of a media-streaming user device constantly transmits an updated data.

Contents

Glossary

ACK Acknowledgement. 10–15, 19, 20, 25, 28, 29, 36, 37

BCH Broadcast Channel. 10–15, 17–20, 24, 25, 27, 29, 36, 37

BS Base Station. 3–15, 17–25, 27, 29, 30, 35–37

CCH Control Channel. 6

CQI Channel Quality Indicator. 7, 11, 14, 21, 25, 29, 34–36

DECT Digital Enhanced Cordless Telecommunications. 4

DL Down-Link. 3, 6, 7, 9–15, 17–23, 25, 27–29, 36, 37

DLC Data Link Control. 5–7, 9, 22, 25

DSP Digital Sound Processor. 4

FDMA Frequency-Division Multiple Access. 5, 33, 34

FIFO First-In First-Out. 10

GCD Greatest Common Divisor. 15

GI Guard Interval. 4, 7–10, 13, 15, 23, 28

HqRTa High-Quality Real-Time Audio. 6

ISR Interrupt Service Routine. 6

LCM Least Common Multiple. 15

MAC Medium Access Control. 3, 5–7, 21, 22, 25

MCS Modulation and Coding Scheme. 7, 11, 14, 21, 35, 36

MCU Microcontroller Unit. 3, 5

NACK Negative Acknowledgement. 13, 20

PER Packet Error Rate. 5

PTT Push-To-Talk. 6

QCA Qualcomm Atheros. 3–5, 11, 15, 19, 22, 23

QoS Quality of Service. i, 3–5, 7–9, 14, 15, 17, 29, 31, 33–36

RACH Random Access Channel. 9–11, 13, 15, 17, 18, 36, 37

RF Radio Frequency. 3

RRM Radio Resource Manager. 27, 33

RTa Real-Time Audio. 6, 9, 10, 37

RTD Round Trip Delay. 5

RX Receive. 4, 5, 22

SOC System-on-Cip. 23

sRTa Sporadic Real-Time Audio. 6

STA Stochastic Timed Automata. 3

TCH Traffic Channel. 6, 8

TDMA Time-Division Multiple Access. 5, 33, 34

TX Transmit. 4, 5

UART Universal Asynchronous Receiver/Transmitter. 4, 5, 22, 23

UE User Equipment. 3–25, 27–30, 33–37

UL Up-Link. 3, 4, 6, 7, 9–16, 18, 20–23, 25, 27, 28, 35, 36

Chapter 1

Introduction

Wireless communication systems are crucial in every-day life and their importance keeps growing rapidly. The number of User Equipment (UE) is increasing constantly and the requirements for the service providers are highly diverse from one use case to another. The shortage of radio resources made an impact on the slowing down of technological advancement. Shared access to the limited frequency bands (either licensed or unlicensed), various types of user devices that must coexist and the power requirements that are especially stringent for portable devices - these are the main challenges in the telecommunication's area nowadays. Therefore, the relevance of a communication system comes with it being adaptive and dynamic, thus able to cope with the ever-changing environment.

In this thesis, we analyze the qualities of a QoS-aware system, propose design options for centralized radio resource management with consideration of different QoS requirements from multiple UE and evaluate the level of improvement. As a foundation we use the static system design provided by RTX A/S. The initial system and its enhancements are modeled using Stochastic Timed Automata (STA) and evaluated in multiple use cases.

1.1 Initial system analysis

The **hardware setup** of the system is inspired from the requirements given by RTX. The scheduling of Radio Frequency (RF) resource is solely executed on a Base Station (BS) which serves multiple UEs. Each BS and UEs are equipped with one Wi-Fi chip (Qualcomm Atheros (QCA)) for radio resource utilization. The chip contains a radio and a Microcontroller Unit (MCU) that handles Medium Access Control (MAC) received from a UE. Even though making use of multiple antennas would enable parallel communication mechanisms for the Up-Link (UL) and Down-Link (DL) transmissions, extending the capacity of the system, a lot of additional effort would be needed for interference avoidance and work-flow synchronization of the different radios. Thus, using only one chip per device is chosen for the cost-effectiveness and simplicity of the system. If a device is the UE, it also contains a sound card with the Digital Sound Processor (DSP) which is connected to the QCA through the Universal Asynchronous Receiver/Transmitter (UART) connection. Meanwhile, the BS has an internal processor for resource scheduling and routing, that is connected with the QCA as well. The hardware structure for a BS and a UE is shown in Figure 1.1.

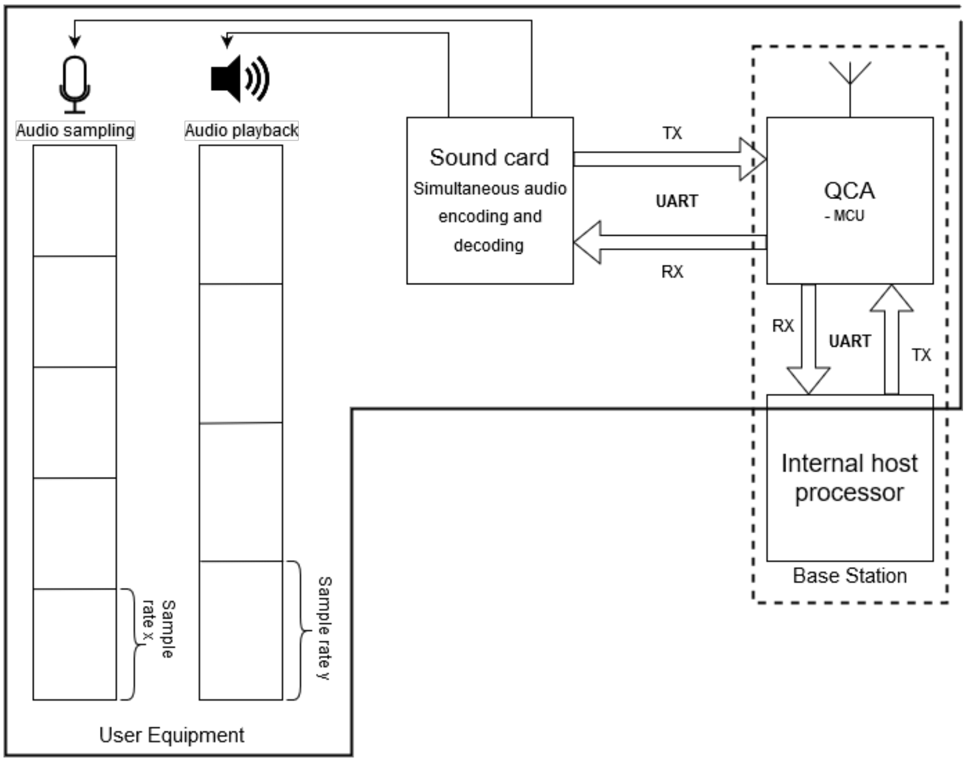

Figure 1.1: General hardware structure of devices in the system

As the special purpose hardware for audio processing allows sampling and playback processing simultaneous and independent of each other, there is only one bottleneck: the UART transmissions. While a UART is transferring the Receive (RX) or Transmit (TX) data, it is in a busy state and cannot operate new transfer requests of the same direction (RX or TX). As the UART speed is modifiable and there is always a Guard Interval (GI) between two concurrent one-way UL transmissions, it can be ensured through an appropriate GI and UART speed combination, that the UART will never be in a busy state when a new request arrives.

According to [5], Wi-Fi solutions in comparison with the Digital Enhanced Cordless Telecommunications (DECT) is an overall winner in providing wireless data service. With a QoS-aware data prioritization, it can also outperform DECT (which was primarily created for telephony) in the area of wireless telephony. As we have only one communication chip on a device, we will control access to shared radio resources by utilizing **Time-Division Multiple Access (TDMA)**. It uses a single frequency band for both TX and RX transmissions through the assignment of time slots. TDMA also brings simplicity (avoiding complex inter-user interference issues) while maintaining similar performance compared to Frequency-Division Multiple Access (FDMA) and being more suitable for higher bandwidths. Moreover, this type of duplexing is getting more and more favoured as the frequency spectrum gets more costly and scarce. Finally, the signal power balancing is not needed as only one device is transmitting in the channel at a time.[13]

For every frame, each UE requests certain **QoS** requirements. QoS can be determined through various parameters: the latency (can also be Round Trip Delay (RTD)) restrictions, data rate/throughput, energy consumption and reliability requirements which demand different time-bandwidth resources. Reliability is expressed through the Packet Error Rate (PER) and has a negative covariance relation with it.

The QoS criteria are selected based on the nature of transmitted data which depends on a profile. The UEs with the same QoS can be further ranked by the level of criticality through the role assignment and/or current status of the system (e.g., spatial distribution of the users).

The **data** exchanged between the BS and UEs is grouped into user and metadata. Metadata is the medium control data (referred to as MAC) that helps to maintain a connection between devices. User data is grouped into multimedia (audio/video) and Data Link Control (DLC) data which can be device metrics, notifications. Multimedia data is delay-sensitive and, depending on the exact behavioral profile, constantly or periodically streamed. We will mainly focus on audio transmission. Control data and DLC is not delay sensitive. Reliability is an important quality for these bulk data transfers while the higher latency values are allowed. The MAC control data is processed in the MCU within a QCA wireless network chip while DLC user data is transferred through the UART to the host processor to be processed, just like the multimedia data.

Depending on the required QoS, each UE uses different codecs for audio data encoding/decoding. The codecs identify sampling bit-rate. An audio sample is represented by a fixed number of bits and an audio packet may contain any number of the audio samples.

Sampling of the audio data is executed by the sound card in parallel with other processes within a UE . It can be the sample- or frame-based sampling where either a single sample is created and stored in the memory before taking another sample or, in case of the latter, several samples are collected and then stored in memory. Frame-based sampling has an inherent latency while the sample-by-sample approach provides minimal latency but consumes more CPU cycles due to more frequent Interrupt Service Routines (ISRs).

User profiles are defined with regards to the different bandwidth, latency, reliability requirements and nature of the data transmitted. The three main profiles that were scoped taking into account use cases from RTX are:

- Listen-Only
  UE with Listen-Only status does not get a dedicated Traffic Channel (TCH) for audio data. It receives broadcast DL channel and has a dedicated Control Channel (CCH) for connection maintenance and status information exchange.

- Sporadic Real-Time Audio (sRTa)
  UE of this profile transmits audio infrequently (e.g., in the Push-To-Talk (PTT) mode). The dedicated UL (and possibly DL) is assigned at the occurrence of certain events. The statistical model can be constituted to approximately identify user's behavior throughout the time. The remaining time when not transmitting audio, these users keep the connection with the BS in Listen-Only mode by observing broad-cast signals and periodically sending meta-data. The latency requirements are not critical for this profile. As an example, it can be limited to 70 ms.

- Real-Time Audio (RTa)
  The devices within this profile are immediately granted a dedicated connection. There are certain requirements for latency (e.g, 40 ms) and reliability, thus the priority is higher than for the devices with sporadic audio transmission behavior.

- High-Quality Real-Time Audio (HqRTa)
  Just as for the RTa profile, HqRTa UE gets a dedicated connection for audio transmission. The throughput and latency requirements can increase (e.g., to less than 20 ms or even 10 ms), for example, when these devices enter an area within a certain range from the BS.

The radio resource can be categorized into the fields of time and frequency. The time is divided into **frames** and each frame is further split into slots. As an example of a real-world application in the use case where the scheduling is static and all nodes are assigned with a dedicated UL slot, frame size can be 2 ms, while a single UL slot takes approximately 100 µs. The frame size is highly associated with the latency requirements of the active UEs in the system.

A **data packet** that is transmitted between a user device and a BS consists of four main parts: a header, MAC information, DLC user data and audio user data. The header data contains all necessary information to identify the burst, decrypt it, and other information regarding the transmitting device and, if it is a DL header of the BS, some information about other connected devices, channel quality indications and capacity limits. In the examined use case, MAC and DLC data share the space within a packet, and the MAC data has a priority over the DLC data. Finally, if it is a transmission from a BS or a streaming user device, additionally it will contain dedicated area for audio data and a checksum. Further decomposition of a packet is described in subsection 3.0.4.

Bandwidth assigned to channel and/or slot can be a variable unit ranging between 20 to 160 MHz. However, switching it is a considerably expensive operation and its introduced latency varies depending on the radio chip hardware choice. E.g., in the industrial case provided by RTX, the used Wi-Fi chips execute channel bandwidth switch in approximately 250 μs. Thus, we assume that it is static in the system design.

In [4], **modulation** is described as a process by which data is placed on the radio waves for transmission. It is achieved by varying the amplitude, phase, or frequency of a carrier wave using the base-band data-bearing signal.[11] The order of a scheme indicates how compactly the data is modulated onto the waves. The higher modulation scheme order is the higher data throughput can be obtained. However, higher schemes also require better signal conditions to work. Meanwhile, lower-order modulation schemes provide lower data throughput and require the lower quality of the connection.[4] Thus, modulation for data transmission is chosen after evaluating the Channel Quality Indicator (CQI). It also determines minimal values for the GIs needed for the configuration to function.

Error coding is an error correction using redundancy where data sequence is divided into blocks and amendment data is appended to each of the blocks. The proportion: $data + correction\_data/data$ determines a concept of a code rate.[10] E.g., the rate 1/2 code means that we have one correction bit for each data bit. Thus, throughput is reduced by 50%.

The **Modulation and Coding Schemes (MCSs)** are optimally combined pairs of modulation and transport block size.[2] The MCSs are classified into three families A, B and C that differ from each other by the basic payload size.[17] For the re-transmission part, the same or another MCS from the same family can be used. An UE is able to find out directly from a DL signal which MCS is used and adjust to it.[6]

GI is the time a radio must wait between transmissions to ensure that the receiver can determine the beginning and the end of the transmissions.[15] It is mostly relevant in the design of the BSs that are receiving multiple frequent UL bursts from the UEs.

Re-transmission is executed during each frame of the initial system and its structure is kept unchanged from the transmission sub-frame. It occupies half of the frame time (re-transmission sub-frame) and uses a secondary channel for transmission (changes the frequency to increase a chance of data reception). Overall, this approach provides high reliability and likeliness to meet QoS requirements, but it is inefficient as time is wasted for re-transmission of data that was received successfully at the first transmission attempt.

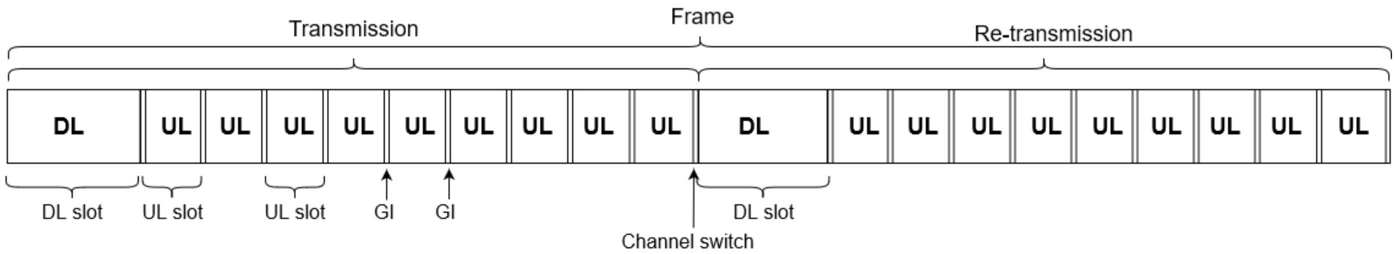

Figure 1.2: Frame structure of the initial system

Minimal requirements:
The wireless network must provide support for multiple QoSs, roaming capabilities to other BS (that facilitates hand-over support across the network), support for multi-cast, uni-cast and broad-cast types of transmissions.

1.2 Conclusion of system analysis

To simplify the design of the system and make it easier to analyze and improve by introducing changes to the scheduling of radio resources, we make the following assumptions:

- The bandwidth of channels is static.
- A UE that transmit audio through the TCH does not need to send additional metadata for connection maintenance. We assume the needed control data is embedded in the headers of audio packages sent.
- Piggybacking is used to send resource requests for subsequent transmissions.
- Switching to a secondary channel is assumed to happen once every frame - before the beginning of the re-transmission part.
- Minimal GI requirement is a static parameter of a specific system.

Chapter 2

Design

Problem solution design is introduced progressively starting from the initial system in section 2.1. The specification of it is provided by RTX A/S. Each further subsection of this chapter introduces new features that each brings additional efficiency and flexibility improvements to the overall system.

2.1 Initial system

The initial system design is simplistic but provides a high level of reliability. It does not support a concept of QoS, and only considers two possible profiles of a UE: RTa and Listen-Only. As a BS does not have knowledge about other BSs, there is no possibility of load balancing among them.

2.1.1 Frame structure

Each streaming (RTa profile) device connected to a BS has a single dedicated slot for data transmission (audio and/or control, depending on what a UE has to update at a given time). As the audio data sampling is frame-based and processes the same amount of audio bits, each audio slot has the static size. The Listen-Only UEs are not streaming audio, but receiving audio from other devices through the BS's broadcast DL transmission. These devices do not need a dedicated slot in each frame, so a BS can have a large capacity of users within this profile compared to the RTa transmitting devices. To maintain a connection, this type of UE needs a dedicated slot every 5 s (according to the RTX use case). Slots that are dedicated to Listen-Only devices that only transfer metadata or DLC are shorter and also have a static size. A setup slot (further on referred to as Random Access Channel (RACH)), which is the last UL slot in transmission and re-transmission sub-frames, is used for UE's connection to a BS and facilitates a handover of a UE to other BSs.

The size of each GI is calculated dynamically and is equal. GIs occupy all the vacant time in the frame, which is not used for the dedicated DL and UL transmissions. The scheduler is responsible for calculating slots sizes and GI. This interval is further verified to be not smaller than the one defined in the system configuration.

Re-transmission is executed in every frame and takes half of its duration. It is equivalent to the transmission sub-frame, but to increase the chances of successful reception, it switches transport channel (changes the frequency which introduces additional delay, e.g. 244 μs). The structure of the frame of this system design has also been described in section 2.1 and is visualized in Figure 1.2.

2.1.2 Scheduler

The resource scheduler in the initial system design is executed before the start of a new frame. With no UEs connected to the BS, the scheduler only evaluates Broadcast Channel (BCH) and RACH slots sizes and sets GIs to take up the vacant time and divides it into an even duration among the intervals. BCH is dynamic and depends on the number of registered UEs to the particular BS. Meanwhile, RACH slot size is designed to be constant.

To connect to a BS, a UE scans different channel frequencies to locate a BCH signal. These channels are embedded in the device's configuration data. After a successful BS discovery and received Acknowledgement (ACK) through the subsequent BCH, a new device gets connected. The BS scheduler identifies the profile of the new member and adds it to the appropriate queue (either of Listen-Only or of streaming devices). Then the slot time for the dedicated device's UL slot (if it is needed) is evaluated and its start time is added to the schedule. The devices are served from the top of the queues sorted in First-In First-Out (FIFO) order and moved to the bottom of the queues after getting resources allocated. The slot position depends on the profile of a UE. The priority of slot assignment is given to the RTa UEs, then left unoccupied time is divided between Listen-Only devices. If not all the UEs got slots dedicated in the frame, they will be prioritized in the scheduling of the next frame. The same time properties from the transmission sub-frame are used for the re-transmission.

If the GI duration calculated is less than a minimum required, the scheduling activity is repeated and the last of the dedicated slots for a Listen-Only UE is removed to free up time in the frame. If there are no dedicated slots for the Listen-Only devices and the scheduling still fails to create a feasible schedule, streaming UE gets its dedicated slot removed from the frame. The process is repeated until successful scheduling execution and until there are dedicated slots to be removed, otherwise, the system setup is reported to be unschedulable.

2.2 System topology

The network of devices in the communication system contains a topology. The BSs are fixed and their locations are defined statically. On the other hand, locations of the UEs are dynamic and each device contains a mathematically defined trajectory that is changing over time. The distance between a BS and a UE is a major factor making an impact on the quality of a communication channel between the devices (expressed through the CQI). It directly affects the quality of transmission, error rate, thus the chosen MCS as well. Moreover, due to the change in the distance (when a user device moves away from one station while getting closer to another base station), the handover of a UE between two BS is triggered.

During the setup slot (RACH), a UE can scan a network to find other BS in the area. If it recognizes another station, it evaluates the quality of the channel with the new station. If the link is better than with the base station it is connected to, the UE disconnects and requests connection with the newly discovered BS. The BSs can also be decision-makers with regards to handover of a UE. Constantly exchanging information about the network topology, they can force the handover of a user device between each other without a need for a device to scan the network. The main advantages of such an approach are simplifying UE's logic (consequently, cheaper hardware) and achieving robust load balancing between the BSs. The design of this scenario is further discussed in section 2.5.

2.3 Sleeping mode

Through the use of ACKs between a BS and UE it is possible to design flexible means of re-transmission. During each frame a BS and the UEs no longer need to re-transmit all packets from the transmission sub-frame to provide high reliability of the connection. Re-transmission is not necessary for the UEs that had a dedicated slot in a transmission frame, successfully received DL data, successfully transmitted UL data to the BS and it was communicated between the two parts.

A user device can turn off the radio in the Wi-Fi chip to save power, thus enter the Sleeping mode. Entering this mode in the setup provided by RTX takes 500 µs and wake-up delay is 250 µs. If the whole chip would be turned off to save even more energy, the wake-up time including antenna start-up and time synchronization would take more than 2 ms which is considered to be an unsatisfactory-long period for frames as short as 20, 10 ms or even less. Thus, we assume that in the Sleeping mode only the antenna is turned off in a chip, but the chip itself stays on. In the use case of RTX which uses a QCA chip, turning off the antenna can still save up significant 70 percent of the energy consumed: going from 63 mA in the active mode down to 20 mA in the sleeping mode.

A UE that has no dedicated slots and successfully received a BCH from the base station in the transmission sub-frame can also enter the Sleeping mode during the re-transmission sub-frame. It turns the antenna back on just before the beginning of a new frame.

The Sleeping mode is mostly concerning the UEs that are portable devices. A BS, on the other hand, is assumed to have a stable power supply and no limitations on weight and size, so there is no need for it to have the Sleep mode. Moreover, it facilitates multiple Listen-Only devices without dedicated slots for optimal usage of the air resources, therefore the BS needs to re-transmit the BCH in every frame. Finally, the DL broadcast re-transmission helps the new entrants to discover BCH and connect faster.

2.3.1 Scheduler

To communicate successful initial transmission of data, a transmission sub-frame must include an additional DL transmission from a BS. As the transmission status report data can be stored in a single bit (for true or false evaluation) combined with the header data (e.g., around 20B) which makes up the bulk of the transaction, this DL slot's duration can be considerably short, for example, 2 µs of air-time with 60 Mbit/s data rate. Thus, the introduced delay is tolerable considering the gained benefit.

With the ACK signals available, the scheduler is executed before the frame starts and before the start of the re-transmission phase within that frame. It has to integrate transmission of DL ACK for each slot exclusively or a combined DL of all ACKs at the end of the transmission sub-frame. The two options are further explained in subsection 2.3.2. In the re-transmission scheduling the scheduler only keeps the dedicated slots of the UEs from which the BS did not receive expected data or the data was corrupted due to connection interference. The down-link ACK slots are included as well which helps to find a number of the UE packets lost after unsuccessful re-transmission. At the end of the frame, the ACK statuses of all connected devices are reset.

2.3.2 Frame structure

There are two approaches of communicating ACK between a BS and a UE. The first one only minimally impacts the structure of the frame. A BS uses a broadcast DL slot after all of the UL slots in the end of the sub-frame to collectively send the ACKs of all received ULs. According to the received data, a UE turns off the antenna or prepares for the re-transmission. If the BCH ACKs transmission failed and some or all user devices with dedicated slots did not receive it, the ACK statuses are also appended to the re-transmission BCH data. The devices can enter the Sleeping mode after they received re-transmission BCH as well.

Another approach that allows a UE to enter the Sleeping mode sooner is by

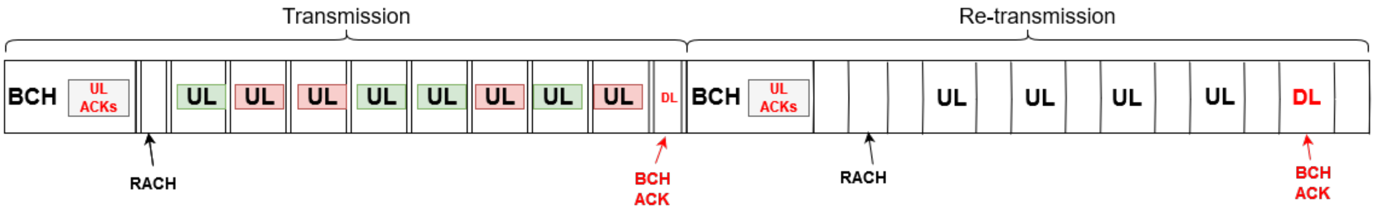

Figure 2.1: Frame structure of the system with enabled broadcast DL ACK

using a dedicated unicast DL slot for reporting an ACK or Negative Acknowledgement (NACK). In this case, a BS communicates the status of a transaction right after that transaction happened and before the next UL slot. The information needed to be sent in this DL is minimal as a UE will be expecting to receive a transmission within the ACK interval right after its UL air-time. If no DL burst was located (in cases of external interference), a NACK is assumed. As in the first approach, the BCH of the re-transmission data also includes the ACK statuses of all slots, thus the user devices can re-evaluate if the ACK status is received correctly and if the re-transmission is needed.

This approach increases the number of GIs in the frame schedule but allows the UEs to save more energy as a UE can start sleeping right after its dedicated transmission slot. Moreover, another benefit of having a dedicated DL communication is that only data of a single user is lost in case of a high connection interference during the transmission. Meanwhile, if the broadcast DL transmission was interrupted, the data of all connected UEs is lost.

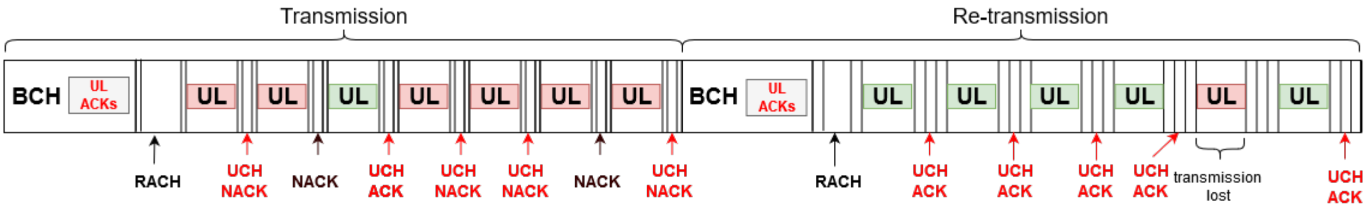

Figure 2.2: Frame structure of the system with enabled unicast DL ACKs

Figures 2.1 and 2.2 show the changes in the frame structure introducing each version of the ACK transmission. Both figures provide a scenario where some of the ULs were transmitted successfully (marked with a green rectangle) and some failed (in a red rectangle), thus the re-transmission sub-frame contains only the ULs slots that failed in the original transmission. In both frame structures, the RACH slot is moved to the beginning of the sub-frames - right after the BCH air-time. It is done so that the setup slot could be used by all UEs before some of them entering the Sleeping mode. All of the BCH DL data packets contain a replicated information of all previous sub-frame UL ACKs. A NACK colored in black means that there was no unicast DL burst received by the last slot owner. Therefore, it is initially assumed that transmission was unsuccessful and its status is checked again later in the subsequent BCH received.

2.4 QoS awareness

Each UE has a certain profile assigned to it. The profiles were introduced in section 1.1. Each profile imposes certain requirements for the radio resources allocated to a group of the UEs. The order of resource assignment is also prioritized according to the QoS requirements: the higher QoS required, the higher priority a UE gets.

Audio sampling is no longer frame-based as in the initial system design, but sample-based. It makes the system more dynamic and agile, although scheduling gets more complicated. The frame size depends on how many connected UEs there are and what their radio resource requirements are. As a BS is considered to be powered by an AC power supply and do not have strict requirements for weight, it can contain a powerful processor that would execute complex scheduling efficiently. Thus, we assume that scheduling delay is minimal and does not introduce additional latency compared to the initial system scheduler.

As each of the streaming UEs has different resource requirements and their sample rates are different, dedicated slots for the devices in a frame are no longer of even size. They are calculated dynamically considering the sample rate and minor frame duration (it is defined in subsection 2.4.2). Different MCS are chosen for the ULs to make packet transmission more robust in consideration of the CQI value. The higher an index of the MCS is, the higher speed is provided and the better connection must be between a BS and a UE as the error correction redundancy rate is lower.

To facilitate robust re-transmission with multiple retries available, we only use dedicated UL ACKs from this design step onward. A UE might transmit multiple times in a re-transmission sub-frame if there is a space available. Additional available air resources are originating from the successful transmissions of the ULs.

2.4.1 Scheduler

The scheduler starts resource assignment from the connected streaming devices that are in the queues sorted with regards to the priority, i.e. the QoS indication, and waiting time in the queue. According to the latency requirements and the system of connected UEs to the BS, the scheduler assigns to a UE a certain size of a slot and a start time. The slot start position in a schedule depends also on a UEs preparatory time needed for the upcoming transmission

Preparatory time

As the slot sizes assigned to the UEs are no longer constant, the transmission preparation time is widely varying. The scheduler must take the UE's preparatory time into account when placing the slots within a frame. The time between the BCH reception and the dedicated UL must be bigger than the time needed for QCA to load and process samples that are ready in the memory. It is because a BCH contains information on the current transmission schedule describing when a user device gets access to the air resource. If the condition is not met, the UL slot must be moved further in the transmission sub-frame. Control slots can then be placed closer to the beginning of the sub-frame to occupy the free air-resource as we assume that the preparatory time of control information is minimal and shorter than a GI.

Constant re-evaluation of preparatory time could be reduced creating a schedule strategy. After evaluating current and historical (if available) information the scheduler can create a plan for several frames and share it with the UEs. Then a UE can adjust the start of a UL arrangement and a BS can schedule air resources without additional restrictions. More details regarding this approach are provided in section 2.6.

2.4.2 Frame structure

Two new definitions are introduced in the frame structure as the frame size becomes dynamic: minor and major frame. The frame consisting only of a transmission and re-transmission sub-frame is a minor-frame. Its length is the Greatest Common Divisor (GCD) of all the latency requirements of the connected UEs. The stricter the latency requirements of the user devices are, the shorter is the minor frame. On the other hand, the major frame size is the Least Common Multiple (LCM) - the time during which all the connected UEs must get the required volume of air resource so that their QoS requirements would be met. Only audio latency requirements are taken into account evaluating minor and major frame values, meaning that disconnect value (which can be as high as 5 seconds) is not considered.

As in the previous frame designs, the BCH is sent at the beginning of a transmission. Afterward, the RACH takes place and the rest of the time is used for dedicated ULs each having a dedicated ACK DL slot. During the re-transmission sub-frame, the free air resource is utilized allowing multiple retries of unsuccessful UL. It is demonstrated in the first re-transmission sub-frame of the frame structure in part B of Figure 2.3.

In case there are no connected UEs, the frame of a BS only contains a BCH and RACH slot. Frequent broadcast transmission enables faster connection to a BS. A frame structure example of no user devices connected and one UE joined is

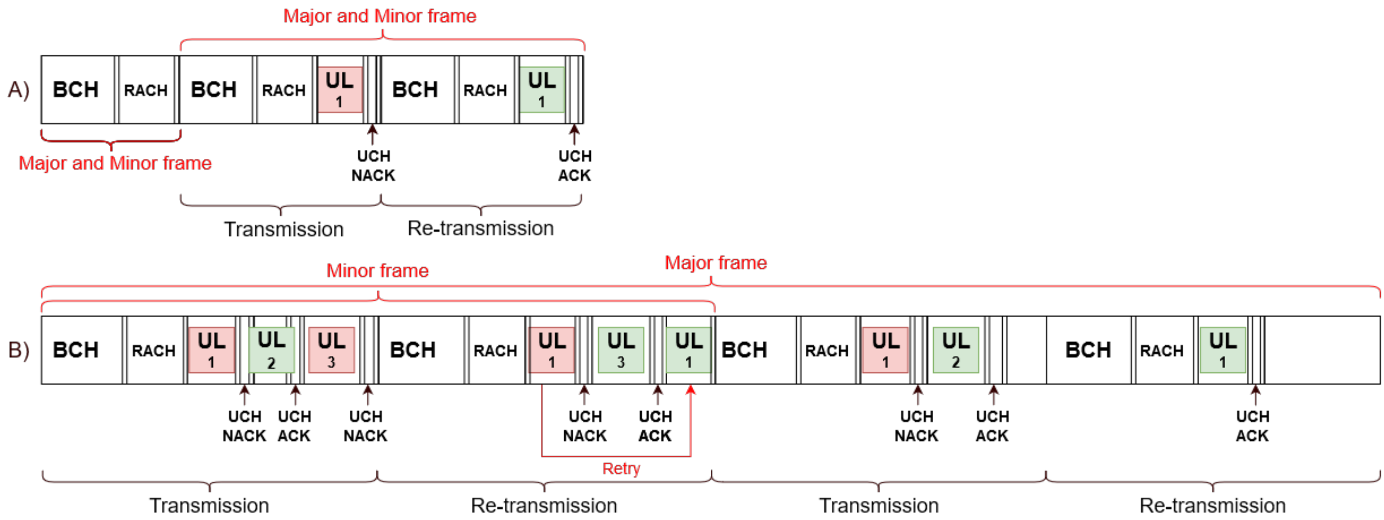

Figure 2.3: Frame structure of the system with dynamic frame size

illustrated in part A of Figure 2.3.

2.4.3 Load-adaptive re-transmission

The next step in improving re-transmission efficiency is detailed in this subsection. Re-transmission of exact data is wasteful as some of the audio sampling data becomes obsolete. It can be avoided by maintaining a moving window covering the relevant samples that are ready in the memory at any given moment.

Sample window

Each sample of audio data has a different duration of its relevance which can be called an expiration time. When this point in time is exceeded, the audio sample becomes obsolete thus its transmission should be avoided. Knowing the audio sample rate (r) and maximum audio latency (L) we can calculate the window size in a number of samples:

$$n = L(s) \times r(samples/s)$$

Only the latest samples that are ready and within this window are valid for the transmission. The size of the largest transmission of audio data corresponds to the size of the relevant sample window. During each UL slot (in transmission and re-transmission sub-frames) a data packet can be updated by moving the window: appending the most recent samples and dropping obsolete ones. This way, re-transmission not only re-sends previous data but also includes the newest data, if any. In the case when the data transference was successful, the streaming UE may require a smaller amount of air resource in the next frame which can be favorable to the other connected devices.

2.5 Load balancing between base stations

As the connection between the BSs can be conducted via fast wired communication, the stations can exchange their network status and condition rapidly. It allows to base decision making on the BSs where they control and enforce the handovers of the UEs. Thus, a UE can be programmed minimally - to be able to establish an initial connection. After a UE connects, a BS it is assigned to can change the connection autonomously. A simplified UE's workflow does not include regular network scanning for nearby stations and executes it only in case of a complete connection loss.

The BSs must still dedicated a slot for random access requests, but they might not be scheduled in each frame. In a scenario of high load on a BS, it can make use of a RACH slot time to deliver resources for the UEs with the high QoS requirements.

2.5.1 Frame structure

After a UE connects to a BS, the BS can at any time handover user device to another station. The handover preparation and communication between the BSs can happen independently from processes that are managing radio resources for the UEs. It can use vacant time of internal host processor or execute processes concurrently. A BS in its DL BCH includes necessary information (channel, bandwidth, time synchronization, encryption) of a BS from which a UE should prepare to receive a BCH.

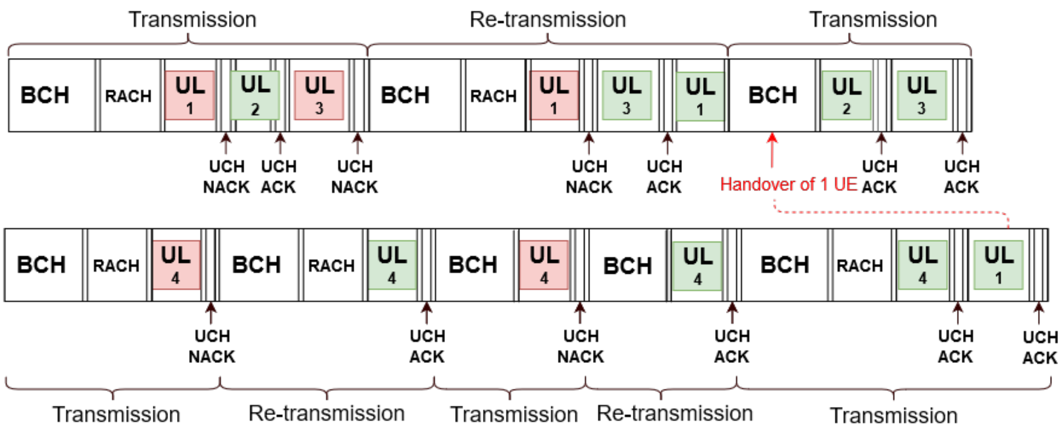

Figure 2.4: Frame structure representing optional RACH slot and BS-managed handover

Figure 2.4 shows a structure of a frame where the RACH slot is allocated sporadically and the handover is controlled by the base stations, that evaluate a location of the UE.

2.6 Strategy-based frame arrangement

We describe strategy-based scheduling as a way of planning multiple upcoming frames considering current network information and/or historical data. The main difference from the previous scheduling methods is that the scheduler is executed not during every frame, but every n-th frame. It creates a schedule and shares it with connected user devices. All the UEs that are newly connected to a BS enter the Listen-Only mode before the new strategy is created and then switch to the Streaming mode or stay in the Listen-Only depending on their roles.

The benefits of this multiple frame planning are optimized usage of radio resource time, elimination of packet preparation time as a constraint for the scheduler, and reduction of the average BCH DL packet size. As a BS has all information regarding the required service quality of the devices (audio quality and latency), their preliminary distance, and a behavioral model, it can organize radio resources for the n succeeding major-frames.

2.6.1 Frame structure

A minor frame has an equivalent size and structure of the transmission and retransmission sub-frames. Multiple minor frames can have varying duration, but after its arrangement the schedule of the current strategy is static. Although, the transmission data is dynamic and up-to-date. It means that if the UL transmission was successful, the device will be able to send a packet of new data during the retransmission sub-frame. Therefore the concept of transmission and re-transmission sub-frame becomes blurry. The UEs keep updating data during every dedicated slot, and the packet content depends on whether or not a BS successfully received previous packets. As mentioned in subsubsection 2.4.3, the packet size of a streaming device would not exceed the relevant sample window, thus the amount of radio resource assigned to the user device is sufficient to transmit the maximum amount of valid samples.

As the re-transmission sub-frame can also be used freely for transmission of new data, entering Sleeping mode during this sub-frame is less useful mainly for the streaming devices. The UEs that are in the Listen-Only mode can still have the full advantage of entering Sleeping mode after successful transaction during the transmission sub-frame.

Chapter 3

Model

To find out the advantages and disadvantages of proposed improvements to the RTX use case, we created a set of Uppaal models that represent BSs and multiple types of the UEs. Time-synchronized these devices emulate the behavior of the real-world communication system. The symmetry reduction technique is used extensively for the state space minimization and model optimization. This chapter covers the design of these models including color arrangement used, shared templates and structures used.

During the modeling of the system, frequent collaboration with the engineers from RTX was necessary. Through the discussion and model reviews, they provided feedback especially important to model the initial system as the further models are built upon it.

All versions of the communication network model are under the Git version control and are hosted in GitHub. They can be accessed using the following link: https://github.com/ruttam/TDMA_QoS_RRM.git

3.0.1 Color labeling

To make the models that represent the BS and UE behavior more readable they follow a common location's color system. This makes it easier to understand the states and transitions, find similarities of the behavior between multiples types of devices. In this subsection, these colors are briefly explained.

The states marked **purple** represent behavior of data processing (audio decoding, QCA processing, scheduling and similar). In most of the cases, these states introduce additional delay, unless processing can happen in parallel from QCA activities.

The **orange** color represents busy-waiting time. This time is mostly used to synchronize to the schedule: wait for the scheduling results, a frame start, a slot start, ACK transmission, scan for BCH DL.

The **dark violet** identifies transmission stage. It can be DL for the BSs and UL for the UEs. The transmission result is either **ACK** or **NACK**: light colors for UL, dark colors for BS DL results. If the UL transmission was successful, a user device enters the **sleep** state.

The **dark brown** color represents states of maximum latency violation, disconnection of a user device if the BCH was not received. Initial states of all models are white.

The BS model also indicates **transmission** and **re-transmission** phases. The **dark red** color marks channel switch that is executed in the models of all device types.

3.0.2 Sampling

There are two different types of sampling: frame-based and sample-based. They are both expressed through the Uppaal model. Figure 3.1 shows the cyclic behavior of audio sampling which is equivalent to both types. The first main difference is that sample time in frame-based sampling is equal to the frame duration. In the case of sample-based sampling, the single sample time is evaluated considering the audio quality (expressed in the bit rate or sampling frequency). Then a sample is transferred to the memory and is ready to be included in the packet for transmission. This is the second difference from the frame-based sampling where all samples are not stored in the memory but transferred to the Wi-Fi chip for transmission.

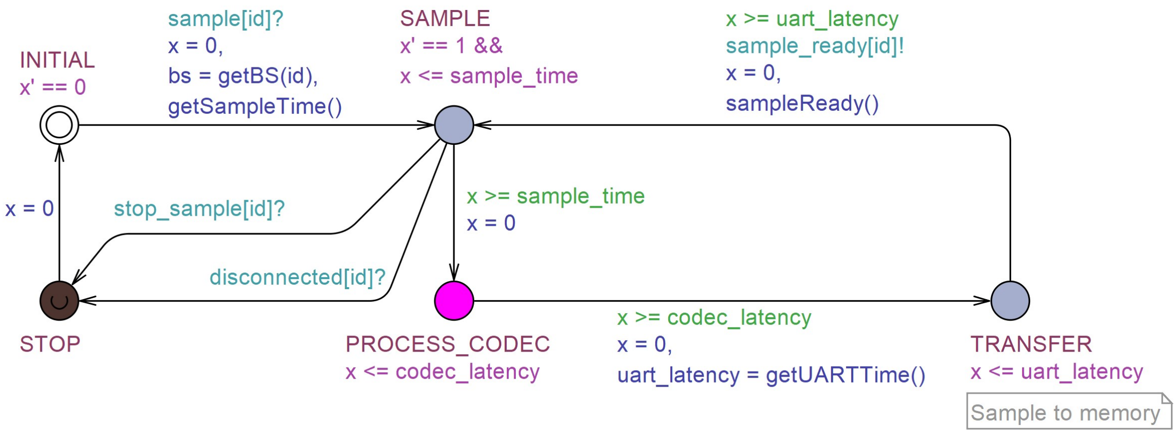

Figure 3.1: Sample-by-sample audio sampling model

Sample-based sampling is more beneficial as a UE can transmit immediately samples that are ready in the memory instead of waiting for a certain number of samples to be created and then transmitted to a Wi-Fi chip.

3.0.3 System topology

To make the model of the system more realistic, the Uppaal model contains the system topology. Each BS is stationary and has different coordinates. As the UEs are mobile devices, their location is expressed through algebraic curves that depend on the common motion timer and get closer to and further from the BS. The speed factor is equivalent between all of the UEs. The curves used are based on sine and cosine functions expressing ellipse, Talbot curve, astroid shaped trajectories. The approximate distance between the devices is then calculated using the Pythagorean theorem.

A BS evaluates the CQI of a UE when it receives a UL transmission. Thus, when scheduling the next frame and deciding on a MCS index used, it evaluates data which is already outdated information. Therefore, the shorter the frame duration is, the more precise evaluation of the CQI can be. An alternative way of capturing a more realistic distance evaluation between a BS and a UE is by the analysis of historical data. A BS can conclude a current trajectory of a UE from the previous positions of the device. This aspect is not further analyzed as it is not within the scope of this study.

3.0.4 Packet structure

Understanding of a packet structure: what data is necessary for different types of transmissions (broadcast and uni-cast, DL, and UL) and what are the main constituents, helps to model realistic transmission times. Moreover, this knowledge is necessary to find ways of optimization of packet sizes. This subsection 3.0.4 describes a general packet from the RTX use-case that containing header, MAC and user data.

Header data

First of all, the common header consists of 48 bits. It contains information about the burst so it could be successfully decrypted. The data in this part of the header is:

1. Flag for multicast/unicast (1 bit)
2. Local administration indication (1 bit)
3. Flag for A (transmission) or B (re-transmission) channel (1 bit)
4. Version of the communication protocol used (5 bits)
5. Communication protocol ID (8 bits)
6. System ID (14 bits)

7. Indication of BS or UE device (1 bit)
8. Device ID (8 bits)
9. Lower frame number (8 bits)

This information is equivalent for both UL and DL transmissions. Encryption information that takes 32 bits is also equivalent for both types of packets.

DL contains an additional 48 bits of burst data for encryption, describing which channel is used and similar. The DL also contains channel quality indication: 2 bits for every device that has a dedicated channel. Assuming that 16 devices can be scheduled in a frame, it adds 32 bits to a packet header. DL packet also contains 8 bits for number of UEs connected to this BS. Overall, there are 25 bytes of the necessary information.

MAC and user data

MAC and DLC space is shared - 30 bytes for information on all connected UEs with dedicated slots. MAC data contains physical layer information and it is event-driven - transmitted only after certain connection changes. DLC as user data is also event-driven, it contains information of updates from the user devices.

Differently from the DL, the UL header also contains 16 bits for transmission and re-transmission channel quality. Overall packet header size is smaller than the one of DL, around 14 bytes in this use-case. UL also contains DLC andMAC data (19 bytes that are shared between both types of information) and 41 bytes for audio data where the last 1 byte if the checksum. Shared bytes are prioritized for MAC data, where the left space is used for DLC user data. Sine MAC data is transmitted when it is updated (handover, changed connection scenarios), then most of the time no MAC data is sent, thus all shared space can be used by DLC.

3.0.5 Data processing

After a successful reception of data either in a BS or in a UE, it must be processed by the Wi-Fi chip and, if it contains the user data, by the internal processor. The Wi-Fi chip processes MAC data and, according to the RTX use-case, introduces a static delay of 25 ms. The user data are transferred through the UART connection to the internal host processor where it is processed. But this processing time is non-blocking in either the BS or the UE. The only constraint is the availability of UART RX stream. E.g., when the BS receives a UL with audio data, a request to transfer it from the QCA to the internal processor is called and the QCA can continue execution without waiting for the transmission completion. However, the next UART transfer request must not be called before the on-going data transmission is completed. Thus, the distance between the reception of two ULs should meet the requirements of minimum GI and maximum UART latency. If these conditions are not met, an additional delay of waiting for UART availability might be introduced corrupting the slot schedule.

Depending on the chosen hardware, the QCA might include additional memory for customer applications that could handle the user data processing in the same chip - so-called System-on-Cip (SOC). It would allow faster data sharing mechanisms and eliminate UART data transfer delay.

The Uppaal model of the user data processing is shown in the Figure 3.2. The model is equivalent to both a BS and the UEs.

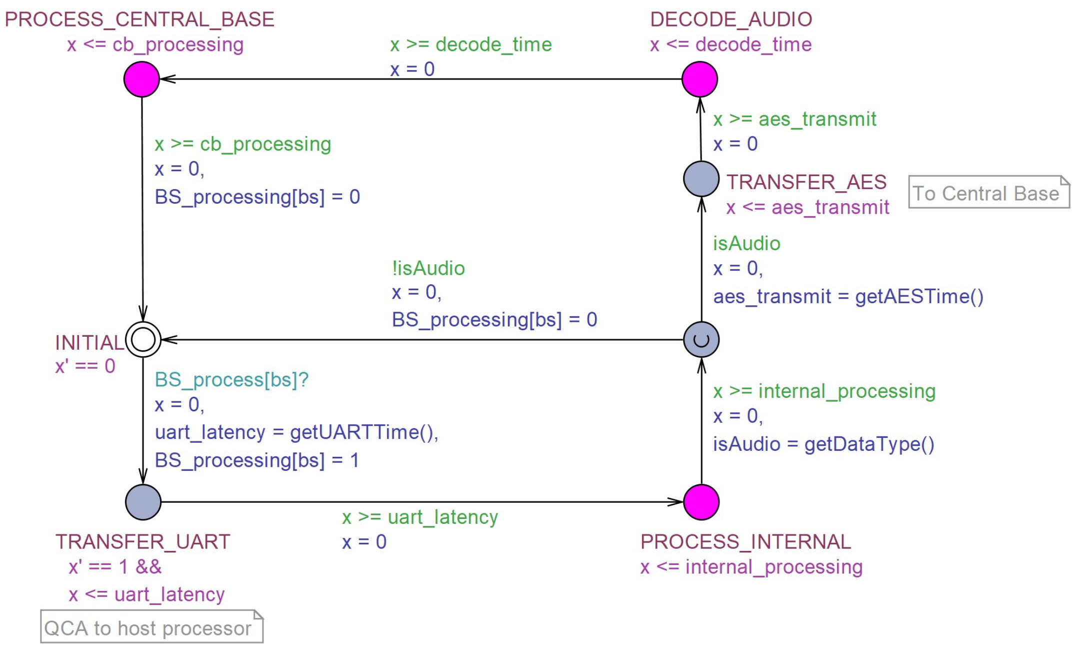

Figure 3.2: Base station model

3.1 Base Station (BS) model

The BS is the main part of the system. It has the responsibility of radio resource management and allocation, time synchronization between devices, and maintenance of a network of the UEs. It is a decision-maker within the system.

Each BS executes a cyclic behavior of scheduling the radio resource among the connected devices, transmitting broadcast DL, waiting for the UL transmissions and connection requests and acknowledging the UL reception. At the end of the frame, the channel is switched to a secondary re-transmission channel. This behavior is expressed through the Uppaal model and presented in Figure 3.3. This model is common for all of the system versions: from the initial to the dynamic frame model. But the scheduler functionality, BCH structure, and other data-related parts of a BS are different between different versions that represent the different designs presented in chapter 2.

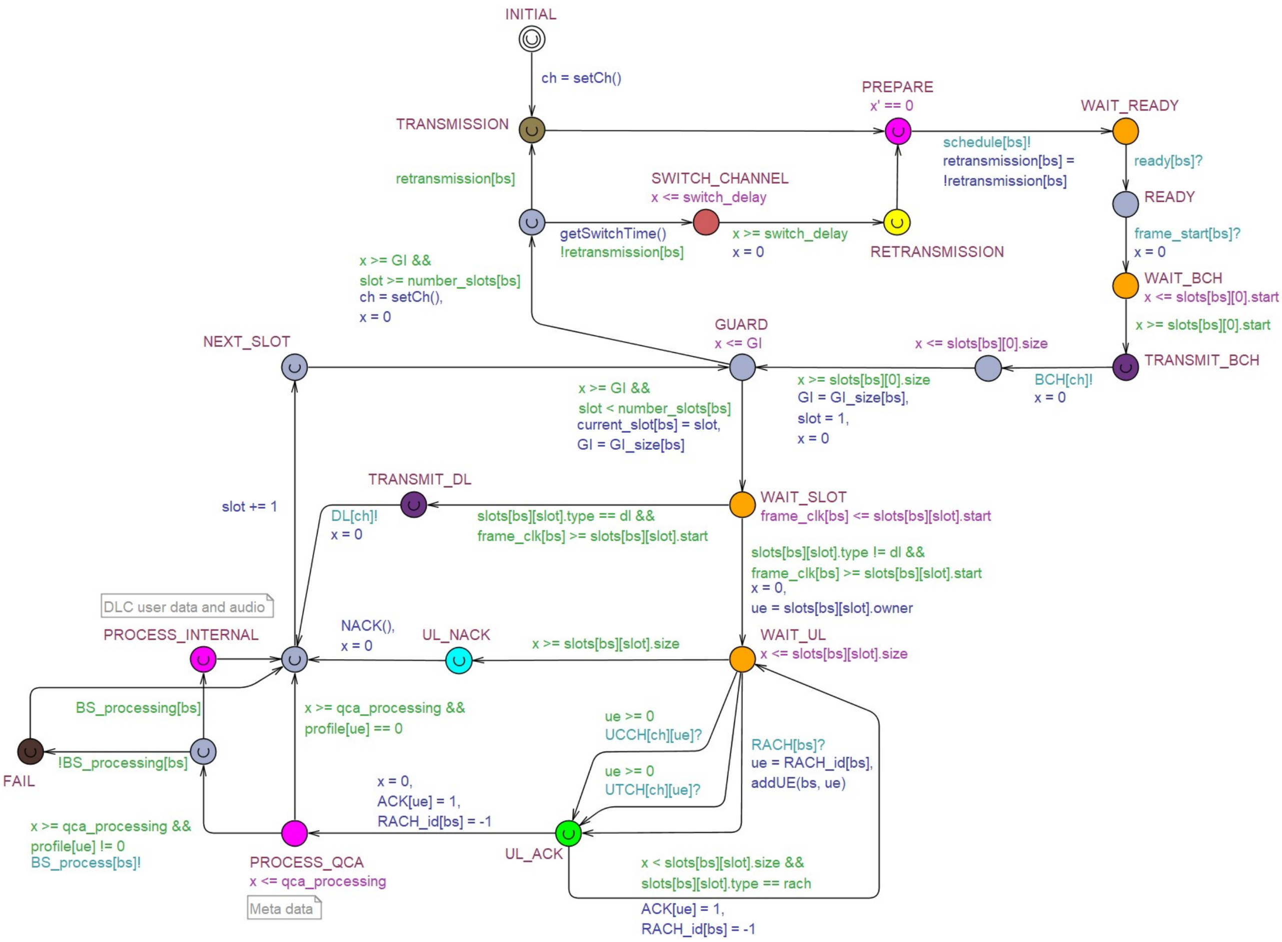

Figure 3.3: Base station model

A BS is designed to support a custom number of devices. When implementing this design, an exact number depends on the specifics of the internal host processor used. The device information is stored in a structure of the C language within the circular priority queues based on the array container. Multiple queues store UEs of different kinds: Listen-Only, Streaming, and Initial mode devices.

3.2 User Equipment (UE) models

Different types of UE are expressed through 3 different models: Initial, Listen-Only, and Streaming.

The Initial model represents a phase of a device when it only keeps the connection and exchanges MAC and DLC user data. It only has one timing constraint: a disconnect timeout when a BCH DL is not received. This disconnect timeout is a considerably long interval compared to the frame size and audio latency requirements. The radio resource requirements are granted to these UEs using the best effort method as they have the lowest priority. These devices can also stay in the sleep mode during most of the disconnect timeout.

The Listen-Only model represents a device that maintains a connection and listens in the system - receives audio data through the BS's DL. The model is designed to both expect to receive the BCH DL within the disconnect timeout like the Initial model, but also within the audio latency to meet the audio requirements. The Listen-Only model is also used to represent a UE that streams audio sporadically. When it does not stream, it stays in a Listen-Only mode. Listen-Only devices can only sleep through the re-transmission sub-frame if the initial transmission of BCH DL was received successfully.

Finally, the Streaming model represents a user device that samples and streams audio. It can have different requirements for audio quality and latency while it maintains the disconnect requirement like the other UE models.

The probability of successful BCH reception and successful UL transmission depends on the CQI of the channel which is individual for each UE. Thus, we decided that the UE's models evaluate it and represent it as a synchronization signal, which is sent with a certain probability that is dependent on the CQI. Otherwise, there is no signal generated. The reception of a DL is modeled similarly, where the UE evaluates its current position from the BS and then after receiving a synchronization signal the decision is made if the packet is received successfully or a transmission failed.

Figure 3.4 presents the model of a UE that is streaming audio data. As a BS every UE also has a cyclic behavior of waiting for the BCH, preparing for the data transmission, waiting for a dedicated UE slot, transmitting and waiting for the ACK signal. If the transmission was successful, a user device can enter the sleep mode for the re-transmission phase. Otherwise, it waits for a subsequent BCH BS starting a new cycle of actions.

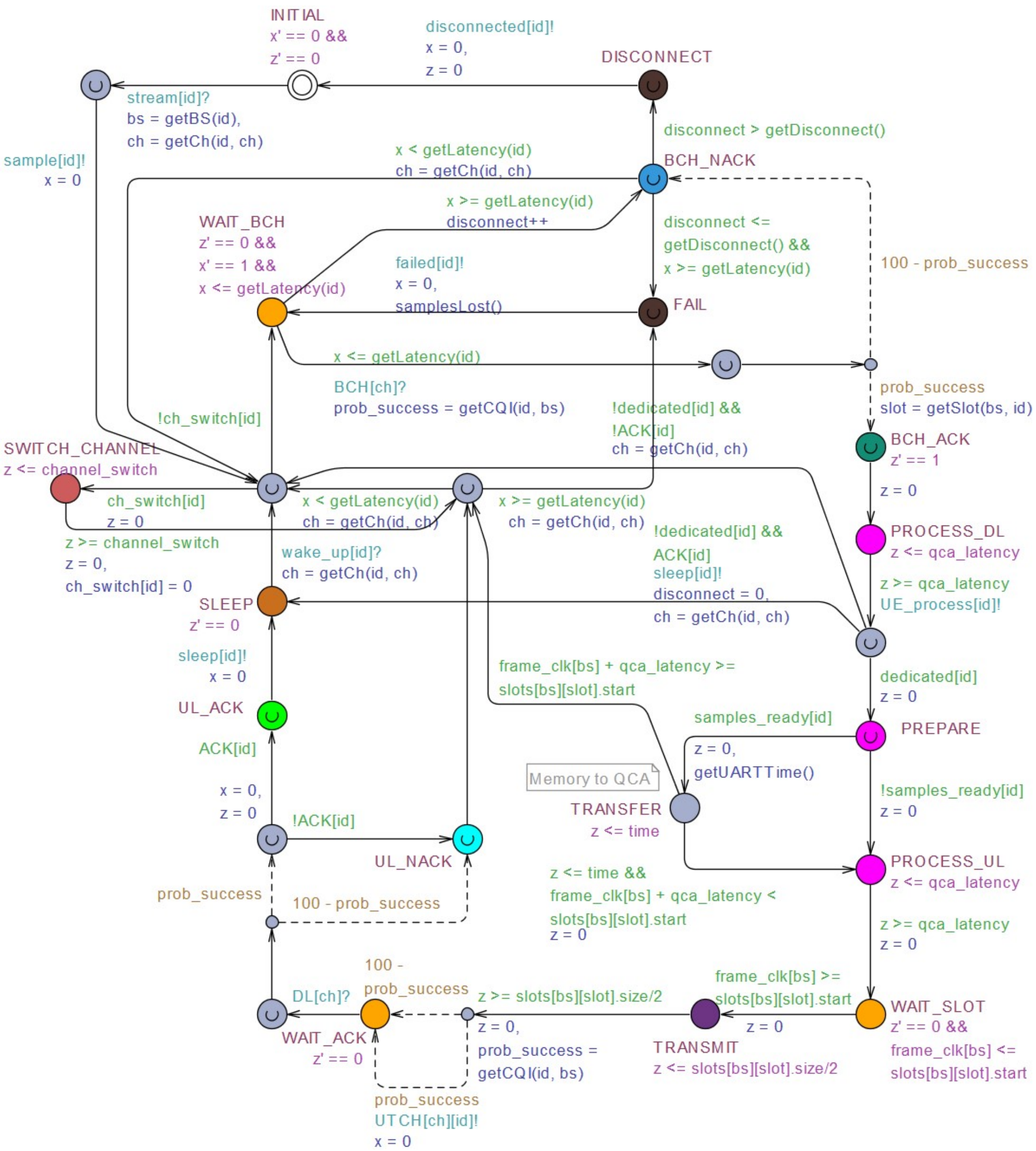

Figure 3.4: Streaming user equipment model

Chapter 4

Evaluation

We performed an evaluation of the Radio Resource Manager (RRM) optimizations expressed through the Uppaal models using the Uppaal Verifier tool. We have compared two Sleeping mode versions and assessed the performance of the dynamic frame approach compared to the one with the static frame structure.

The emulations were executed using a setup consisting of 2 BSs and 20 UEs. The UE were assigned multiple configurations of profiles. E.g., for comparison of Sleep mode gain - a percentage of sleep time within a frame - all devices were assigned a Listen-Only profile. It is because the streaming logic would not make a distinct difference in the sleep time comparison test.

4.1 Sleep mode

The Sleeping mode clearly brings many possibilities and agility to the system. First of all, it allows a UE to save energy by turning off an antenna which consumes most of the device's power. Moreover, a device which enters the Sleeping mode is not using air resource during the re-transmission sub-frame. Thus, vacant time can be utilized for multiple retries of the failed ULs or providing additional radio resources for the UE.

Yet, depending on the way the Sleeping mode is implemented, the overall capacity of transmission time in a frame decreases. The explanation of this is presented in subsubsection 4.1. We will compare two Sleeping mode designs: version 1 with a common BCH DL acknowledgment and version 2 with dedicated DL acknowledgment.

Version 1

We run the Uppaal verifier's estimation query:

```
E [<=2000000000; 5] (max:  sum(i:  UEid_t) sleep_percent[i])
```

to calculate the average sleep percentage in 50 000 frames.

Table 4.1: Estimation results of version 1

Runs	Sleeping time percent
1	45.1
5	43.1
10	45.7

After multiple runs presented in Table 4.1, it appears that the mean sleeping time of all 20 UEs is around 45 percent of the frame time. It is an expected result as the user devices can enter the Sleep mode after receiving the DL ACK transmission. In this version the ACK DL is broadcasted and contains ACK information of every device, thus it is transmitted just before the end of the transmission sub-frame.

Version 2

The percentage of sleep time during the frame in version 2 is dependent on the frame schedule and UEs slot position in it. The closer to the end of the transmission sub-frame a UE dedicated slot is, the shorter its Sleep mode period is. This is because the dedicated DL ACKs are transmitted right after each UL. Naturally, there are more GIs and twice as many transmissions - even though a dedicated ACK DL takes a fraction of the time.

Table 4.2: Estimation results of version 2

Runs	Sleeping time percent
1	63.5
5	72.2
10	63.5

Thus, the best-case scenario of sleeping duration in the Sleep mode version 2 is distinctly higher than Sleep mode version 1, while the worst-case sleeping duration when evaluating the device with a dedicated slot at the very end of a frame is likely to be worse. The average sleeping percentage within a frame is displayed in Table 4.2. The average results of version 2 are approximately 40 percent better than those of version 1.

To conclude, both versions of Sleeping mode implementations have their advantages and drawbacks. The smaller the frame size is, the more beneficial version 1 is and the gain of version 1 gets closer to one using version 2. But the bigger the frame size is, the greater the power usage reduction is in version 2. System developers should decide whether power saved through version 2 is worth the frame capacity loss. But as the dedicated ACK transmissions can be useful for further improvements towards the efficient system design, uni-cast and multi-cast DL should not be discarded.

4.2 QoS awareness

To evaluate some of the improvements of the dynamic frame and the QoS awareness in the system, we will look at the estimation and simulation results of the device connection time to the BS, the error rate - packages lost within the same duration in the two types of models. Finally, we will look at an average throughput of the system measuring data exchanged during the same period between a UE and a BS. We will compare the QoS-aware model with dynamic frame length to the static frame model.

The benefit of the dynamic frame size not only makes the system overall adaptive but allows new users to enter the network quicker. We evaluated the average connection speed of the equivalent systems in the models of static and dynamic frame designs. The results are presented in Table 4.3.

Table 4.3: Connection times results

Query	Static frame	Dynamic frame
$E[<= 2 \times 10^8; 1](max : sum(i : UEid\_t)ConnectTimer(i).x)$	31.15×10^5	15.72×10^4
$E[<= 2 \times 10^7; 5](max : sum(i : UEid\_t)ConnectTimer(i).x)$	44.16×10^5	18.75×10^4

The connection time analysis reveals a vast difference in the speed of the UE connection time to a BS. It is due to the more frequent BCH DL sent, thus numerous more chances to locate the burst while scanning through the channels compared to the static frame size model.

Even more significant assessment is that of an error rate between the two models. We count the number of samples dropped after they could not be successfully transmitted and their latency requirement was violated. The reason for failing transmission is low CQI which is dependent on a UE location. The results of the average number of samples lost are displayed in Table 4.4.

Table 4.4: Average number of lost samples per device

Models	$[2 \times 10^8; 1]$	$[2 \times 10^7; 5]$
Static transmission model	14922	3312 ± 588
Dynamic transmission model	9284.9	775.45 ± 55

The average number of samples lost in the dynamic and QoS aware model per UE was around 40 percent lower in two runs compared to the model of the Sleeping mode, which keeps the static frame size. It is due to having a frame-based sampling and transmitting larger bulks of samples. As the transmission is less frequent, the larger amounts of audio data are lost due to the interference, the fewer chances of re-transmission of data are available.

The periodic motion of the UEs that has a major impact on the transmission success is expressed through the distance to a BS in Figure 4.1.

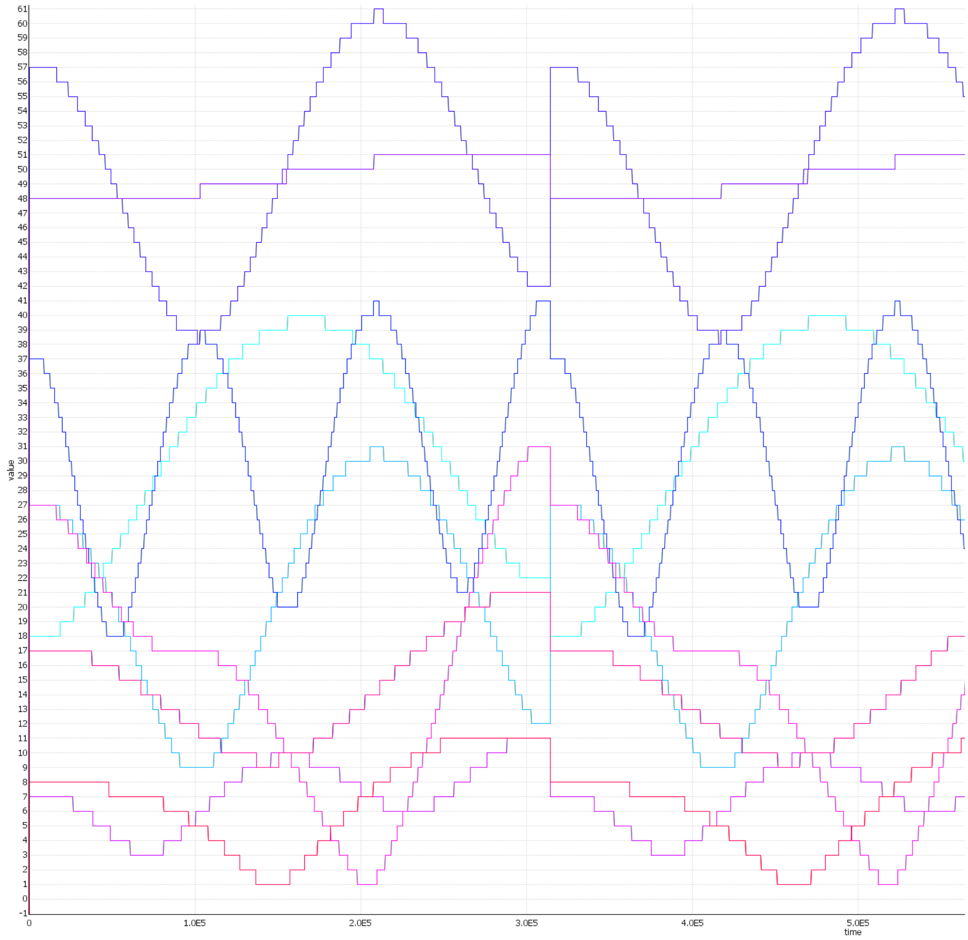

Figure 4.1: The UE location variation from the viewpoint of a BS

The last aspect of the comparison that we have examined is the average number of bytes transmitted throughout the same time period. The results of different period lengths are displayed in Table 4.5.

The reason to have such a high difference between data transmitted in the first shorter period is also influenced by the faster connection time of the model with the adaptive frame size. Running the longer verification query the difference becomes smaller, although it is still as high as around 70 percent.

Table 4.5: Average number of lost samples per device

Models	$[2 \times 10^5; 5]$	$[2 \times 10^7; 5]$
Static transmission model	174.2	251719
Dynamic transmission model	16240	1220410

The various benefits of a dynamic frame scheduling prove its practicality, especially in the networks where the devices are heterogeneous. It efficiently uses the radio resource adapting to the changes in the network setup and enables the development of the systems supporting various QoS requirements.

Chapter 5

Related work

As the purpose of this study is to analyze the use-case provided by RTX A/S and present optimization options for the scheduler and efficient design solutions, the scope of the problem is considerably wide. Thus, finding related work that would be closely correlated was complicated. Thus, we looked at the related work through 3 aspects:

- what has been discovered in the field of QoS-aware radio resource management;
- papers which covered scheduling optimizations using the TDMA channel access method;
- was there any projects on radio resource scheduling applying model checkers like Uppaal or Spin;

A lot of attention during recent years was dedicated to QoS-aware scheduling with regards to Cognitive Radio technology and 5G network ([3], [9]). The majority of the papers consider the usage of the FDMA, distributed decision making, and spectrum-sensing by the UEs. A review of RRM and packet scheduling algorithms with guarantees for QoS was presented in [1]. They expressed algorithm improvements employing multi-objective optimization techniques. Unlike the goal of our study, the authors were focusing on the FDMA channel access method. Similar to our design, Jinchang Lu and Maode Ma in their article [9] from 2011 classify users into 4 classes that define different QoS requirements and determine the required slot qualities.

Authors of the article [7] from 2010 analyze slot assignment using TDMA as a way to share the medium considering an ad-hoc network with no centralized control. Our study covers an infrastructure network instead of an ad-hoc network. The paper of M. W. Khan et al. [8] published in the journal "Ad Hoc Networks" in 2020 is considered to be the most relevant study for our project. The authors classify users concerning the nature of data transmitted. These users enter related queues where they are sorted by their latency and required throughput. They formulate scheduling as a multi-objective optimization problem and optimize it using a Back Propagation Neural Network to maximize throughput and minimize latency. The scheduling was implemented using Matlab and optimizations were executed using the Keras framework in Python.

Considering studies where the experiments were executed using Uppaal modeling and verification tools, there were a few papers in the communication field. Although they are mostly analyzing protocols: Mathijs Schuts et al. [14] in 2009 proposed a distributed, single-channel MAC protocol that helps improving network throughput in the 802.11 ad hoc networks. More recent research from 2015 "Model-Based Verification of the DMAMAC Protocol for Real-time Process Control" [16] presents a Dual-Mode Adaptive MAC protocol for process control applications in a wireless sensor network. In 2009 V. Rosset [12] proposed protocols for the development of safety-critical applications that communicate using dual scheduled TDMA communication protocols.

To summarize, we were not able to find related work that would be closely related to or would overlap the problem definition and especially the approaches used in this report. Our project covers classifying the users into types as most of the studies of QoS-aware communication do. But we extend the most relevant studies adding consideration of the CQI dependent on distance and movement of portable devices. We also consider TDMA shared access control instead of the FDMA which is prevalent in the research of QoS-aware and intelligent radio. As mentioned in chapter 1, TDMA uses the spectrum more efficiently and the consumed power levels of the UEs are lower. Finally, we chose to design and experiment with the QoS-aware network using the Uppaal modeling and verification tool.

Chapter 6

Conclusion

In this project, we considered a use case of communication network setup provided by RTX A/S and introduced the possible design decisions to make it QoS-aware and more efficient. We provided design descriptions starting with the power saving possibilities through acknowledgments at a program level. Based on it, we proposed the scheduling approach of the classified UEs using the priority queues in a dynamic-length frame. Lastly, we suggested the concept of a frame structure without the re-transmission sub-frame to improve the utilization of the air resource. We designed four versions of the system using the Uppaal modeling and verification tool: the initial design, two types of Sleeping mode implementation, and a model representing the dynamic frame concept. Collaboration and feedback from the RTX engineers with regards to the modeling of different design decisions ensured that our models would as closely conform to the real-world requirements as possible. We examined the advantages of each modification over the initial system design. The models contained the movement trajectories of the UEs and distances from the BSs, UL MCS allocation based on the CQI and realistic packet sizes based on the customized 802.11 protocol provided by RTX.

Through the experimental results simulating the behavior of the network of devices, we showed how each of the examined modifications outperforms the initial system design. Even though there still are some aspects of the different models left unestimated, e.g., the fairness of the scheduling algorithm, worst-case behavior, and overall capacity of each BS, the benefits are evident. The proposed changes of the design can create a system that is capable of adapting to the constantly changing environment of the network while making the best use of the radio resources available at the lowest cost.

6.1 Discussion and future work

We observed that the Sleeping mode is primarily advantageous for the UEs that only maintain a connection with a BS or are in the Listen-Only mode. We can also conclude that a shorter frame has some significant advantages over a longer one: fast adaptation to the changes in the network and robust scheduling that can facilitate strict QoS requirements. Moreover, a common BCH ACK can be used as it is more beneficial than the dedicated ACK DLs for the frames with a small number of UL slots. As one of the subjects for future work can be the investigation of ACK transmission at the protocol level. Thus, the dedicated ACKs would not decrease or minimally impact the frame capacity.

Basing a CQI on the distance to a BS is a simplification which in the real-world scenario would not necessarily result in an increased level of interference. For example, even if a UE is moving away from a BS, it can still have a clear line-of-sight to it. Thus, the additional external interferers like devices that use the same frequencies without first listening if they are busy, walls of buildings, must be taken into account. One of the future work goals could be the creation of a more realistic CQI evaluation model. Another aspect is the distribution of the CQI with regards to the MCS assignment. We were considering a linear model while the geometric (or similar) distribution of CQI and a chosen MCS would be more reasonable.

As we use the RACH UL transmissions for the connection of a new user device to the system, a more intelligent method of collision avoidance could be investigated. We assumed a randomly generated back-off interval to reduce a chance of collision.

An in-depth analysis could be made investigating UE movement and impact on the transmission. E.g., in case a UE moves away from the BS, its UL transmission time grows with the increasing distance a signal must travel. This might violate the strictly scheduled slot borders.

Finally, additional simulation experiments of Uppaal models could be done. The examination of how much of the average transmitted data is UL data and DL to find efficiency. Also, the evaluation of the BS's capacity in all versions of the system model and analysis on the fairness of the slot scheduler.

6.1.1 Frames without re-transmissions

In section 2.6 we have already mentioned that as the streaming devices are constantly updating audio data, constructing the frame from the transmission sub-frame and an equivalent re-transmission sub-frame is purposeless. Thus, our future work proposition is to assume that a frame consists only of a transmission phase where the frame size is dependent on the latency requirements of the connected streaming UEs. Other devices that are set to Initial state or Listen-Only mode are served using the best-effort method with a consideration to their connection requirements.

The scheduler would prioritize the audio-streaming devices over the UEs having other profiles. The lower-ranking UEs would receive radio resources left unused by the prioritized devices. Assuming that a RTa device successfully transmitted in the previous transmission, during the next transmission it might only have a few new samples generated to update, thus it would not use the whole slot dedicated to it. After receiving the smaller packet of data, a BS can evaluate the slot time left and estimate if a lower-priority device can use it for the user-data transmission. Then, sending a multicast DL ACK signal (for a successfully transmitted UE and a lower-priority UE at the top of the request queue) the BS would instruct an actively waiting device to transmit. The frame structure of this use case is visualized in Figure 6.1. The streaming devices have dedicated slots from 1 to 3 and a lower-priority non-streaming device is allocated the slot 4.

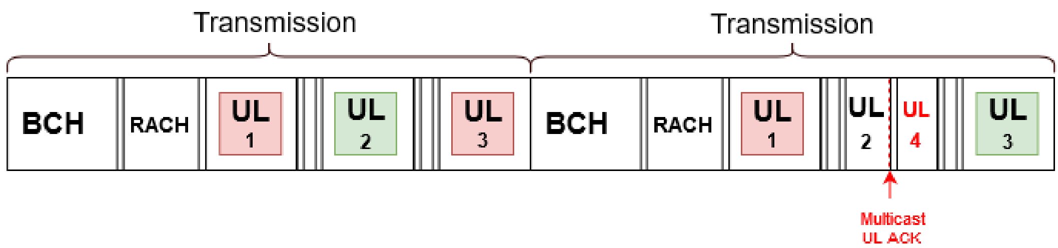

Figure 6.1: Frame structure without division to transmission and re-transmission parts

In case a transmission failed, a non-streaming UE device would be actively waiting for a BCH and/or a multicast DL signal. Meanwhile, it would keep the information updated: if it is a power level, it would make sure that whenever a time slot is allocated, it will transmit (or re-transmit) an up-to-date power level data.

As mentioned in section 2.5, the RACH slot can be excluded from some of the frames, especially if they are considerably short. The space of a RACH slot could act as a buffer air resource for the lower priority devices to allocate space for the pending transmissions to meet the latency requirements.

Bibliography

[1] Ayesha Haider Ali and Mohsin Nazir. "Radio resource management with QoS guarantees for LTE-A systems: a review focused on employing the multi-objective optimization techniques". In: *Telecommunication Systems* 67.2 (June 2017), pp. 349–365. DOI: 10.1007/s11235-017-0342-z. URL: https://doi.org/10.1007/s11235-017-0342-z.

[2] Alex Brand and Hamid Aghvami. *Multiple access protocols for mobile communications: GPRS, UMTS and beyond*. John Wiley & Sons, 2002, p. 213.

[3] Ioan-Sorin Comsa, Antonio De-Domenico, and Dimitri Ktenas. "QoS-Driven Scheduling in 5G Radio Access Networks - A Reinforcement Learning Approach". In: *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. IEEE, Dec. 2017. DOI: 10.1109/glocom.2017.8254926. URL: https://doi.org/10.1109/glocom.2017.8254926.

[4] Linda Doyle. *Essentials of cognitive radio*. Cambridge Univ. Press, 2009, p. 2.

[5] Bruce Alan Fette, ed. *RF & wireless technologies*. Newnes know it all series. OCLC: ocn171111165. Amsterdam ; Boston: Newnes/Elsevier, 2008, pp. 254, 255. ISBN: 9780750685818.

[6] Timo Halonen, Javier Romero, and Juan Melero. *GSM, GPRS and EDGE Performance: Evolution towards 3G/UMTS*. John Wiley & Sons, 2003.

[7] S. M. Kamruzzaman and Mohammad Shamsul Alam. "Dynamic TDMA slot reservation protocol for cognitive radio ad hoc networks". In: *2010 13th International Conference on Computer and Information Technology (ICCIT)*. IEEE, Dec. 2010. DOI: 10.1109/iccitechn.2010.5723844. URL: https://doi.org/10.1109/iccitechn.2010.5723844.

[8] Muhammad Waqas Khan et al. "QoS-aware traffic scheduling framework in cognitive radio based smart grids using multi-objective optimization of latency and throughput". In: *Ad Hoc Networks* 97 (Feb. 2020), p. 102020. DOI: 10.1016/j.adhoc.2019.102020. URL: https://doi.org/10.1016/j.adhoc.2019.102020.

[9] Jinchang Lu and Maode Ma. "Cognitive radio-based framework and self-optimizing temporal-spectrum block scheduling for QoS provisioning in WiMAX". In: *Computer Networks* 55.6 (Apr. 2011), pp. 1291–1309. DOI: 10.1016/j.comnet.2010.12.003. URL: https://doi.org/10.1016/j.comnet.2010.12.003.

[10] Jan Maes, Marc Vercammen, and Luc Baert. *Digital audio technology: a guide to CD, MiniDisc, SACD, DVD(A), MP3 and DAT*. Focal, 2001, p. 97.

[11] John Rogers, Calvin Plett, and Ian Marsland. *Radio frequency system architecture and design*. Artech House, 2013, p. 12.

[12] Valério Rosset. "Services for safety-critical applications on dual-scheduled TDMA networks". In: 2009.

[13] Michael Ryan and Michael Frater. *Communications and information systems*. English. OCLC: 50579648. Canberra, A.C.T.: Argos Press, 2002, pp. 100–103. ISBN: 9780958023801.

[14] Mathijs Schuts et al. "Modelling Clock Synchronization in the Chess gMAC WSN Protocol". In: *Electronic Proceedings in Theoretical Computer Science* 13 (Dec. 2009), pp. 41–54. DOI: 10.4204/eptcs.13.4. URL: https://doi.org/10.4204/eptcs.13.4.

[15] Jeff Smith, Jake Woodhams, and Robert Marg. *Controller-based wireless LAN fundamentals: an end-to-end reference guide to design, deploy, manage, and secure 802.11 wireless networks*. Cisco Press, 2011.

[16] Admar Ajith Kumar Somappa, Andreas Prinz, and Lars Michael Kristensen. "Model-Based Verification of the DMAMAC Protocol for Real-time Process Control". In: *VECoS*. 2015.

[17] Peter Stuckmann. *The GSM evolution: mobile packet data services*. John Wiley & Sons, 2003, p. 99.