PhD Dissertation

University of Trento

International Doctorate School in Information and Communication Technologies

Optimisation of Performance of 4G Mobile Networks in High Load Conditions

Alexey Baraev

April 2014

Abstract

The signalling subsystem of the LTE (Long Term Evolution) networks inherited some of the limitations of its preceding technologies. It is vulnerable to overload situations that occur as a consequence of unpredicted user behaviour. The most sensitive zones are the paging procedure and random access procedure and the signalling channels associated with them. Reliable paging procedure is particularly important. As of the current design, in case of overload of paging channels it blocks the possibility of modification of configuration of a cell, thus limiting the possibilities of the system to recover. This research proposes and analyses a solution to overload of the paging channels in LTE systems. It shows that there is a possibility to completely avoid overload of the paging channels in surging load conditions. The research develops and verifies a mathematical model of the paging procedure. This model is incorporated in the solution, thus allowing computation of the critical load thresholds that trigger the reconfiguration of the paging channels. The solution is explained by a detailed algorithm and validated in a simulator of the LTE paging channels. It is partially compliant with the 3GPP specifications. The research includes a compatibility analysis and underlines the operational procedures that must be defined in the standard. It is important that the implementation of the solution does not affect already deployed hardware but requires a modification of the eNb software. Thus it is possible to prevent development of the paging overload situations, and the solution can be implemented in the hardware that is already deployed in the LTE networks. The main result of this research is a reliable paging procedure that opens further opportunities for optimisation of other signalling procedures and channels.

Table of Contents

List of abbreviations

3GPP	3rd Generation Partnership Project
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BHCA	Busy Hour Call Attempts
CAPEX	Capital Expenditure
DCCH	Dedicated Control Channel
DL-SCH	Downlink Shared Channel
DRX	Discontinuous Reception
DTCH	Dedicated Traffic Channel
EMM	EPS Mobility Management
eNb	Evolved Node B (base station)
EPS	Enhanced Packet System
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FDD	Frequency Division Duplex
GSM	Global System for Mobile Communications
IE	Information Element
IE	Information element
IMSI	International Mobile Subscriber Identity
KPI	Key Performance Indicator
MIB	Master Information Block
MME	Mobility Management Entity
MO	Mobile Originating
MSC	Mobile Switching Center
MT	Mobile Terminating
NAS	Non Access Stratum
OMC	Operation and Maintenance Center
OPEX	Operating Expenditure
PBCH	Physical Broadcast Channel
PCCH	Paging Control Channel
PDCCH	Physical Downlink Control Channel
PD-SCH	Physical Downlink Shared Channel
PF	Paging Frame
PO	Paging Occasion
P-RNTI	Paging Radio Network Temporary Identity
QoS	Quality of Service
RACH	Random Access Channel
RRC	Radio Resource Control

SFN	System Frame Number
SGSN	Serving GPRS support node
SGW	Serving Gateway
SIB	System Information Block
SIM	Subscriber Identity Module
S-TMSI	SAE Temporary Mobile Station Identifier
TA	Tracking Area
TAC	Tracking Area Code
TAU	Tracking Area Update
TDD	Time Division Duplex
UE	User Equipment
UMTS	Universal Mobile Telecommunications System

1. Introduction

Mobile operators have made significant investments in their Long Term Evolution (LTE) infrastructure and continue to extend the LTE coverage and capacity. As the LTE networks acquire more customers, optimisation of the deployed networks requires higher attention and investment. Operators increase optimisation budgets, and suppliers of equipment and services seek to satisfy the operator's demand for value added services for their own benefit.

There two ways for the suppliers of equipment and services increase revenue from optimisation of networks. One way is the optimisation of already deployed infrastructure. It requires, first of all, experienced engineers and, possibly, additional equipment. It is lengthy work; optimisation projects run for a minimum of several months and often exceed a year. The second way is to offer the operators new features of the equipment. Following this line the vendors develop solutions and market them as optional software features that tune the operation of the networks, thus improving the performance of the networks.

Suppliers of equipment and services develop both described above lines of business simultaneously. The second approach – development of optimisation solutions – represents a research potential which, to a large extent, remains unrecognised by the research community.

Certain demand for optimisation solutions in the signalling subsystem of LTE is seen already. As the LTE customer base grows, proportional increase in the signalling load further stimulates the demand for optimisation. Previously, a similar trend developed in earlier mobile standards, such as GSM (Global System for Mobile Communications) and UMTS (Universal Mobile Telecommunications System).

The signalling subsystem of LTE is more flexible in comparison with the preceding technologies from the same family – GSM and UMTS. However it still does not have mechanisms to prevent overload of its signalling channels when, for any reason, the demand for capacity surges.

There are situations when the demand for signalling capacity significantly exceeds the installed capacity. This can occur as a consequence of an outage of several sites due to, for example, power cut. More frequently this is a result of unpredicted events or accidents that cause abnormally high concentration of users sometimes paired with an outage of a segment of the network.

Operators plan and dimension the signalling subsystem so that it allows full utilisation of the traffic channels in normal load conditions. The capacity is regularly upgraded to match the growing demand, but it is still based on normal user behaviour and concentration for geographical segments of the network.

Unpredicted events with abnormal concentration and behaviour of users can lead to development of situations when the demand for services significantly exceeds the installed traffic capacity; however this capacity remains partially unutilised because of overload of the signalling channels. The LTE channels that are most vulnerable to overload are the paging channels and random access channels. These signalling channels are the first to accept the surging load resulting from an unpredicted event. Among the two channel types, the paging channels can lose a share of their effectively capacity in overload due to the logic of their operation. This issue is addressed in this research.

The standard does not include a possibility of re-configuration of the paging channels in operation. This strictly limits the number of options available to vendors and operators for timely resolution of overload situations.

This research proposes and elaborates a solution to overload of paging channels in LTE.

Suppliers of infrastructure equipment and operators proposed a limited number of solutions related to the paging procedure which are normally formulated in patents. Section 0 offers an overview of these solutions.

Section 3 describes a mathematical model of the paging procedure that is used in the proposed solution. This section also includes a characterisation of a simulator of the paging procedure and provides an analysis of operation of the paging procedure.

The solution that is presented in Section 4 is accompanied by the results of its validation, technical requirements and a 3GPP (3rd Generation Partnership Project) compatibility analysis.

2. State of the art

Signalling is an expensive and the most technological element of any telecommunication system, wireline or wireless. Mobile networks produce especially strict requirements to their signalling subsystems because their users are not connected to the network constantly. Mobile user terminals spend most of the time in the idle mode and only periodically decode the system information from the air interface. All standards of the mobile networks apply this concept in order to optimise the use of the battery resource. LTE is not an exception. Moreover, in terms of the signalling, LTE is more complex than many other mobile standards. The reason for this is the fact that LTE is an evolutionary technology that incorporates network elements of several generations. Each of the LTE “sub-networks” has its own signalling subsystem and can operate independently. The “sub-networks” are GSM, UMTS and LTE.

With a large number of user terminals in the idle mode a mobile network must have mechanisms to locate user terminals and establish a connection when it is requested. This is necessary for both circuit-switched and packet switched service requests.

Service requests can be either mobile originated or mobile terminated. Mobile originated service requests, i.e. the requests initiated by user terminals, start with the random access procedure. The starting point for mobile terminated requests – the requests initiated by the network – is the paging procedure. It is followed by the random access procedure, if successful.

2.1. Choice of the paging procedure as the optimisation target

Paging procedure is the unique way of locating idle user terminals in a mobile network. While the network is in operation any change of the configuration of a cell involves paging. Therefore the paging procedure is the starting point for any real self optimisation in mobile networks. If the paging procedure is reliable, further optimisation of the signalling subsystem is possible, otherwise – it is not.

A reliable paging procedure is the one that offers zero probability of blocking of paging requests sent from the network to user terminals. Paging response partly relies on the random access procedure that must also be reliable, but it is an independent direction for optimisation which is outside the scope of this research.

2.2. MME paging for EPS services

The MME paging for EPS (Evolved Packet System) services is defined in the 3GPP specifications [1][2][3] as a part of the connection establishment procedure for the downlink data service requests. Evolved LTE systems apply MME paging in parallel with the MSC (Mobile Switching Center) and SGSN (Serving GPRS support node) paging mechanisms that are defined for service requests of other types.

Similar to the paging mechanisms of the older ETSI/3GPP standards, the MME paging request messages on the air interface can carry either the long and unique identifier IMSI (International Mobile Subscriber Identity) or a shorter identifier S-TMSI (SAE Temporary Mobile Subscriber Identity). 3GPP defines the composition of the subscriber identities in a dedicated specification [16]. S-TMSI is a 40-bit long identifier that is unique in a given tracking area (TA) list, i.e. in a network segment with a limited number of cells. IMSI is 64 bits long that uniquely identifies a SIM card (Subscriber Identity Module) on the global scale and thus in a particular mobile network.

The MME is the network entity responsible for paging for EPS services. Its task in the paging context is to initiate the paging procedure and manage it entirely. Thus MME generates S1-AP paging messages upon reception of the downlink data notification messages from SGW (Serving Gateway) [7]. MME sends an S1-AP paging message vie the backhaul links to all base stations in the TA list that the MME provided to the paged UE during the most recent TA update. At the moment when an S1-AP is sent the MME starts the T3413 timer for the paged UE. The T3413 limits the period validity of the paging message. MME considers expired any paging response received beyond the T3413 cycle.

A signalling flow diagram on Figure 1 illustrates the MME paging procedure.

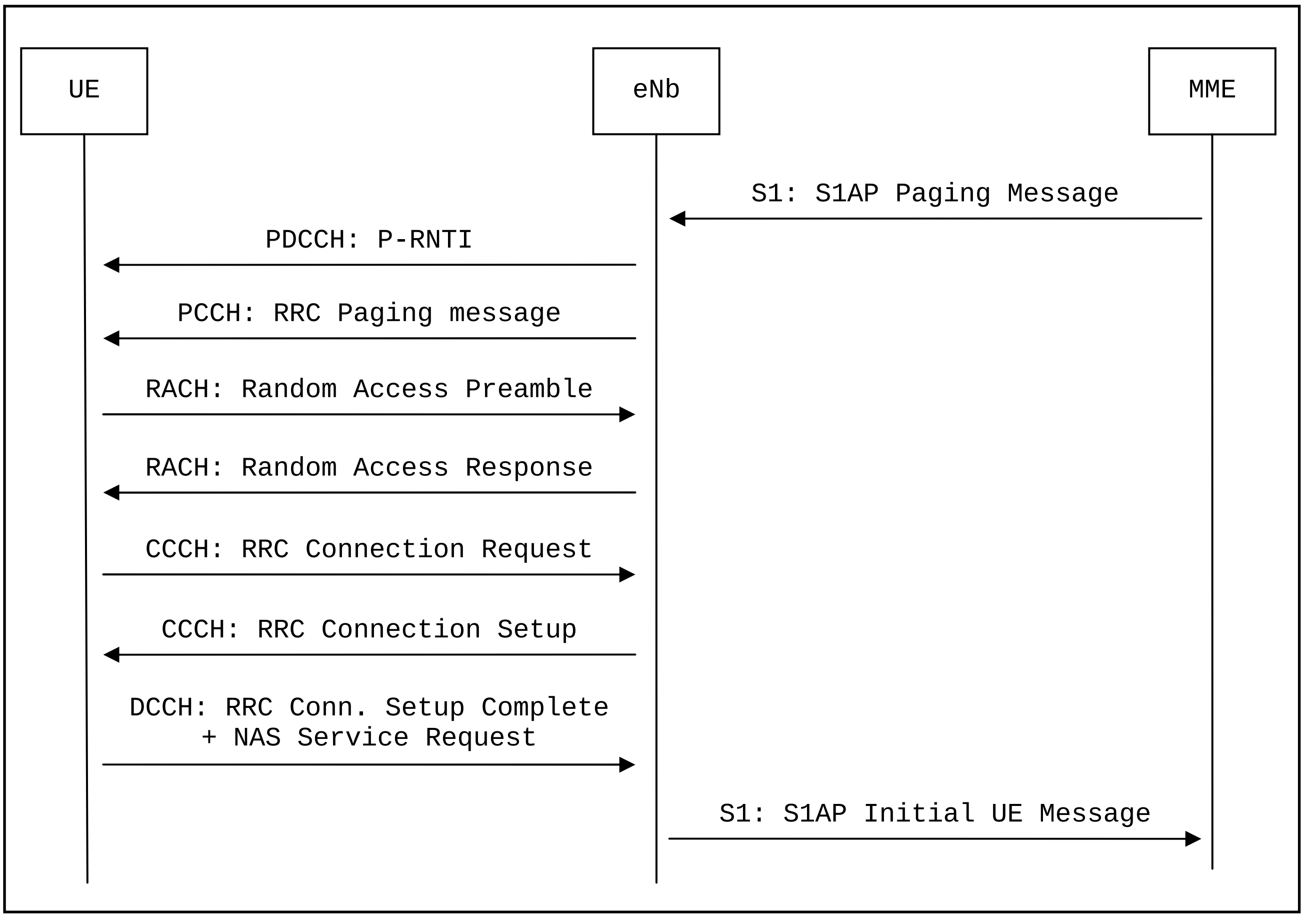

Figure 1. MME Paging Procedure.

All base stations that belong to the paged TA list receive the broadcasted S1-AP paging request, decode the UE identity from it and place a corresponding paging request into the paging buffer.

User terminals do not monitor the air interface continuously. LTE employs a concept of discontinuous reception in order to save the battery resource of the UEs. Thus, upon registration in the cell, each UE calculates its Subframe Number (SFN) that it has to monitor for paging. The base station performs the same calculation and transmits the paging with a given S-TMSI or IMSI only in the paging occasion (PO) with appropriate SFN. Thus the UE wakes up for a single short interval in the DRX (Discontinuous Reception) cycle to decode the paging message.

At every paging occasion the base station unloads a number of paging records from its paging buffer and transmits them in an RRC paging message on the air interface over PCCH (Paging Control Channel). The format of the RRC paging message is defined by 3GPP in the specification of the RRC protocol [4]. It is possible that the base station cannot send the paging request in the first scheduled paging occasion. In this case the base station keeps the paging record in the PCCH buffer and makes further attempts to transmit the paging request with the subsequent paging occasions. Base stations can have configurable settings that limit the maximum number of attempts to transmit a paging request or, alternatively, such settings can limit the period in which eNb attempts to transmit a paging request. This depends on the design of particular equipment and varies between vendors.

User terminals in the idle mode have to monitor the paging occasion that they calculate based on their IMSI and decode the UE identities transmitted in POs. This is defined by 3GPP as one of the UE procedures in idle mode according to the dedicated specification [2]. User terminal returns to idle mode, if it does not recognise its identity in the list of identities decoded from a PO. When an UE recognises its identity in a PO it has to respond to the paging by acquiring the resources via the random access procedure and transmitting a service request message. This is defined by 3GPP in [8] and [7]. User terminals have to ignore any paging request while undergoing any EMM (EPS Mobility Management) procedure or service request procedure at the moment a paging request is received [2].

Upon request from UE the base station grants access to the UE and receives a service request message from it. eNb must forward the service request to MME via its backhaul connection on the S1 interface. When the MME receives the service request it stops the T3413 timer associated with that particular IMSI and proceeds with the connection establishment procedure.

If the MME does not receive a response to a paging request within the period of time defined by the T3413 timer, it may repeat the paging and restart the T3413. This depends on the MME configuration. The maximum number of paging repetitions is a configurable parameter of the MME equipment. Operators normally limit it to one repeated paging request. The type of UE identity that is normally used in the repeated paging is IMSI, while S-TMSI is used for primary paging, as it saves the resources on the radio interface.

2.3. Overload of paging channels

Reliable paging is important for stable operation of mobile networks. The efficiency of paging depends primarily on the operation of the signalling channels that are used to transport the paging messages. A possible difficulty in the operation of the paging procedure is the overload of the paging channels on the radio interface. Most often it occurs as a result of an outage or unusual user behaviour. In terms of the traffic capacity, such events affect several sites in a tracking area - those sites which take the excess load. However, in terms of paging, such events affect entire tracking areas with dozens of sites since paging is performed on a basis of tracking areas. All cells in the affected tracking area experience an increase in paging load. Thus several sites that go out of service cause service disruptions on numerous other sites which are not affected by the outage directly and while they have fully operational traffic channels.

Figure 2 illustrates the consequences of an outage of several cells that resulted in increased paging load in their tracking area.

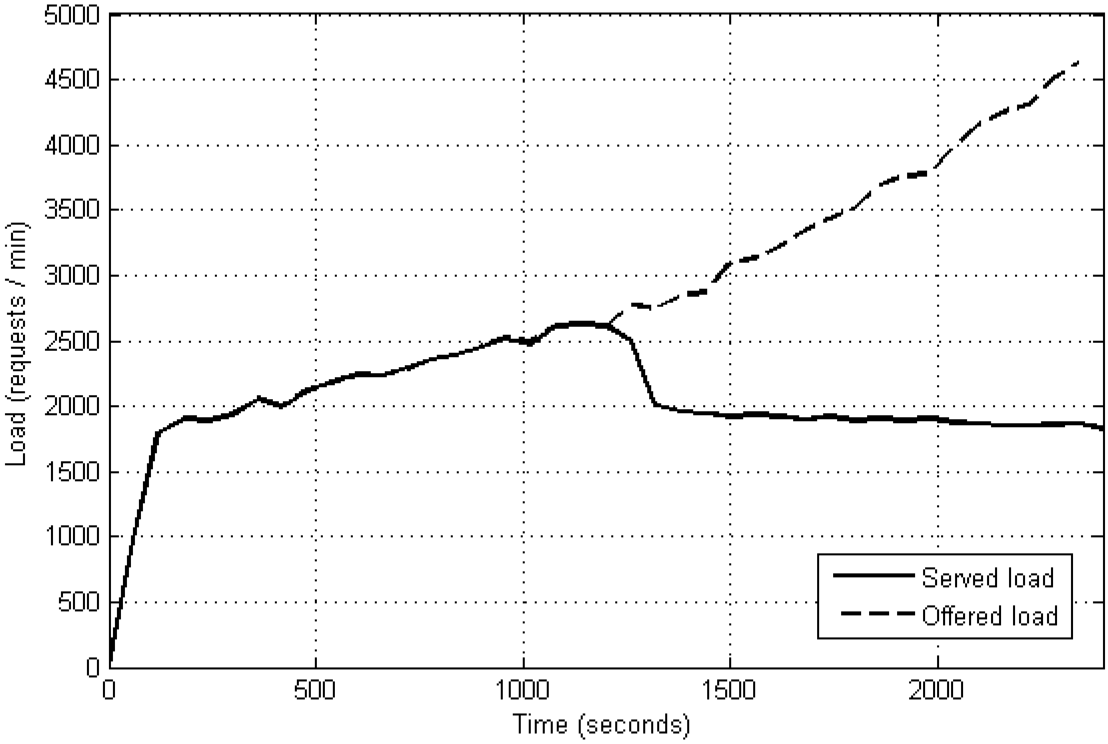

Figure 2. LTE cell is unable to load its traffic channels to 100% in a situation of paging overload.

Figure 2 is produced by the simulator of paging channels that is characterised in Section 3.2. It shows an LTE cell that accepts the user traffic that grows abnormally after an outage of several sites in the same tracking area. The cell serves 100% of the traffic until its installed paging capacity is exhausted approximately 1200 seconds after the start of the simulation. The installed traffic capacity allows higher load, however the overload in the paging channels makes it impossible to accept new user traffic.

The most undesirable effect can be observed on the graph after 1200 seconds of simulation. Due to overload in paging channels, the served user traffic drops by approximately 25% - 30%. This effect requires a detailed explanation.

A significant percentage of users drop from the network due to an outage of several sites in a tracking area. Some of the users previously served by the failed cells register in the neighbouring cells while a significant number of users remain out of coverage. The disconnected users are still registered in the tracking area. Therefore the MME continues to page them when it receives notifications from the SGW. The demand for the PCCH (Paging Control Channel) capacity surges as the MME receives no response from UEs and has to repeat the paging. The secondary paging is performed with the IMSI in the paging requests. It requires higher PCCH capacity than in the case of the primary paging which uses S-TMSI. Field statistics shows that it takes as little time as several minutes before the PCCH load reaches the installed PCCH capacity. After another 7-9 minutes depending on the design of the equipment, the user traffic in DTCH (Dedicated Traffic Channel) starts to decline as the overall time of transmission of paging requests now exceeds the timeout setting in the MME. Hence the cells in the affected tracking area enter such state in which they report a high blocking rate while having DTCH capacity available. However the available DTCH capacity cannot be utilized due to overload of the PCCHs.

If a network operates according to the current 3GPP specifications, the described situation of PCCH overload would require an intervention from engineers, or otherwise it cannot be resolved. Thus it is an opportunity for a research with an evident target which can lead to a release of a commercial solution. Equipment manufactures and network operators have already raised this issue, nevertheless it is neither addressed in the specifications nor in the equipment. Section 2.4 provides details on the relevant activities within 3GPP and describes some measures that vendors undertake to improve the performance of their equipment.

2.4. Solutions proposed by vendors, operators and research community

The LTE signalling subsystem undergoes continuous development. This process is driven by technical groups inside the 3GPP, which bring together representatives of major equipment vendors and service providers. Once the specifications are approved and accepted, the equipment vendors can still introduce modifications into the signalling subsystem of their infrastructure equipment, provided that the operational principles comply with the 3GPP specifications. Thus the freedom for making changes after the specifications have been accepted and investments in networks have been made is limited. Mobile networks normally use equipment from several suppliers, and the network entities have to be able to mutually interoperate.

It is reasonable that paging related solutions for LTE proposed in the recent years limit their scope to modifications that do not affect the existing infrastructure. This section presents several solutions that target directly the paging optimisation.

2.4.1. Cascaded paging proposal by Motorola

In 2007 Motorola proposed one of the most relevant solutions to reduce the paging load on the air interface. It is described in 3GPP technical reports [9] and [10].

The solution introduces an approach of cascaded paging. Thus the MME transmits the first paging message only to the eNb that serves the cell where the paged user terminal was last observed. The authors do not mention how this cell can be defined. However based on the RRC protocol and the corresponding procedures it can be the cell where the UE performed the last tracking area update or the cell from which it last time established a connection.

Paging to the last known cell is performed instead of paging to the entire TA list as defined in the 3GPP specifications. After the transmission of the first page the MME starts a timer to monitor the paging response. The authors suggest short waiting values for the timer, e.g. in the range of 100 ms. If the first page remains unanswered upon expiration of the timer, the MME transmits the second paging message. The second page is transmitted to entire tracking area – this is normal operation of the paging procedure as it is specified by 3GPP.

The authors claim that the paging load on the radio interface can be reduced by over 50% with the proposed solution. Hence the solution offers the following advantages:

- The declared decrease in paging load would undoubtedly lower the risks of PCCH overload and improve the technical vulnerability of the paging mechanism, thus making it highly reliable and scalable.

- The paging capacity a cells can be decreased, thus freeing the resources for traffic.

At the same time there are several negative aspects in the solution:

- The solution introduces a significant delay into the paging procedure (and as consequence into the connection establishment procedure). Speed of the connection establishment is a major KPI of a mobile network. Any solution that results in a degradation of this KPI will face a strong opposition from the technical teams of mobile operators.

- The proposed 100 ms interval for the new MME timer for the first page is insufficiently long for a large percentage of paged UEs. This is due to the concept of discontinuous reception that instructs the UEs to monitor only one PO subframe in the DRX cycle. The most commonly used DRX cycle is 1280 ms long. Thus the first page with 100 ms expiration timer would expire in approximately 92% of cases. Should we make a more precise calculation with the random access procedure taken into account, the percentage would increase even beyond this value.

- The solution requires additional capacity on the S1 interface. The backhaul resources are not as scarce and expensive as the radio resources, nevertheless this must be evaluated and taken into account in network planning and the CAPEX/OPEX budgets.

The solution naturally requires a modification of the 3GPP specifications. However it is important to underline that the RRC protocol on the radio interface remains unchanged with this solution. The required modifications on the S1 interface are less costly than those for the radio interface.

The most recent Release 12 specifications [3] do not include the modification that the authors proposed. The proposal was not accepted by the relevant 3GPP technical group. Nevertheless, the authors registered their idea in the prior art database where it is available for download [11].

2.4.2. Two-phase paging solution by China Mobile

In 2009 China Mobile proposed via the 3GPP a paging solution that targets a decrease in paging load. It is based on the company's expertise in operating the world's busiest 2G/3G network. In [15] the company provides a brief analysis of its 2G/3G solutions that it has in operation, and in [12] it targets the LTE and describes a resulting solution for the MME paging.

The new solution suggests that MME has to use the last known cell information. Cells in a mobile network have neighbouring cells – this information is stored by MME and eNb. China Mobile proposes to send the primary paging message to the last known cell and its neighbours that are configured in the database. There are two possibilities:

- Each cell in paged directly via its S1 interface (MME to eNb).

- Last known cell is paged directly from the MME via S1, while its neighbours are paged indirectly via their X2 interfaces from the eNb that serves the last know cell.

Thus, according to this scheme the primary paging request is sent to a limited number of cells, and the secondary paging is performed on the entire TA list.

This is beneficial from the perspective of the paging load reduction because such paging scheme limits the number of paged cells. Neighbour cells list rarely exceeds 14 items, while tracking areas are significantly larger.

The solution suggests that the repeated paging, nevertheless, is transmitted to all cells in the tracking area list, as it is defined in the specifications.

Precise definition of the last known cell is important. The authors refer to the China Mobile 2G/3G experience described in [15]. Hence the possible criteria are the last tracking area update and the last established connection.

The solution has the following benefits:

- Over 50% reduction of the paging load on the radio interface and in the backhaul on S1 interface.
- Reduced probability of paging overload.
- Lower paging capacity requirements on both radio and S1 interfaces.

The solution causes certain side effects that have to be assessed:

- Degradation of a major KPI - the average duration of connection establishment procedure. This is due to increased repeated paging; its volume is inevitably higher than with the standard 3GPP paging procedure.
- Unclear reliability and availability of the X2 links in already deployed LTE networks can limit the benefits especially on S1 interface.
- If X2 is used for the paging of the neighbours, this operation increases the load on X2 and thus may require an additional investment. The X2 capacity requirement must be evaluated.

While the solution offers evident advantages it requires further analysis. As the starting point, field statistics can help estimate the percentage of the UEs that are typically paged in their last known cell and in the neighbour cells to the last known cell.

The solution is particularly valuable because China Mobile operates equipment from all major vendors and hence its engineers have real possibilities to assess the technological vulnerability of the LTE standard that is not related to particular equipment.

The most recent Release 12 specifications do not include this paging scheme.

2.4.3. Smart paging lists by Verizon

In 2010 Verizon engineers proposed a user specific paging approach in a patent application [17]. Since Verizon did not have their LTE network in commercial operation on the date when the patent application was filed, it is likely that the proposed solution was based on the operator's experience in LTE trials. The patent was published in 2012.

Verizon proposes an algorithm to limit the area of the primary paging while keeping unchanged the scope of the repeated paging.

The patent describes three new functionalities:

- UE location logger
- UE location pattern analyzer
- Smart paging list generator

The listed functionalities may be integrated into MEE or assigned to a new network entity. By default the authors consider these functionalities as performed by MME.

Most paging optimisation solutions (including the solutions described in Section 2.4.1 and Section 2.4.2) aim at limiting the scope of the primary paging. They differ in the method of achieving this effect. The Verizon proposal differs from other considered solutions significantly since it brings new criteria into the consideration. Thus the solution considers handovers in the process of establishing the location of UEs. This is an addition to the 3GPP defined procedures of tracking area update and connection establishment that other solutions use for this purpose.

Another additional component of the solution is the use of the statistics on user behaviour. It defines a dedicated functionality "UE location pattern analyzer" that analyses the pattern of UE movements and locations depending on the day of the week and the time of the day and possibly some other factors. This statistics is used to define highly probable locations of any UE in the network.

The data from the EU location logger and UE location pattern analyzer is supplied to the "smart paging list generator". This routine can produce a list cells for primary paging. This result differs significantly from other proposals that consider one single cell for primary paging but not a list of cells.

There is a functional possibility that the smart paging list analyzer defaults to the primary paging based on the tracking area list for a particular UE.

The MME standard paging mechanism receives "paging instructions" for every UE to be paged from the smart paging list analyzer. The paging procedure that follows this step is standard, and it complies with the 3GPP specifications.

Although the Verizon solution is not backed by field statistics or analytical evaluation, it clearly offers at least the following benefits:

- It is not necessary to change the standard MME paging procedure defined by 3GPP. The solution operates in the background, and it provides an optimised paging list at the moment when a decision is required.
- A reduction in paging load. However it must be evaluated or tested. Reduced paging load lowers the risk of overload of the paging channels.

At present there is no information on whether this solution has materialised in the LTE equipment of any vendor. Alcatel-Lucent markets equipment with a paging optimisation solution of similar functionalities [18]. However it is not possible establish a connection between the two solutions due to the lack of information on the Alcatel-Lucent solution. A brief description of this solution follows in Section 2.4.5.

2.4.4. Allocation of paging areas by Huawei

A 2014 patent by Huawei [19] claims a method for allocation of paging areas that targets reduction of paging load. This method differs from the Verizon solution (that targets the same result) in two parts:

- Algorithm for prediction of UE movements
- Higher level of details that are disclosed in the patent

The Huawei solution particularly emphasizes the tracking of moving UEs. It accumulates statistics on movements of UEs in the network on the level of "basic paging areas". The authors define an apparatus for this task which includes two elements: a determining unit and an allocating unit. The determining unit determines the currently visited paging area. The allocating unit analyses the statistics, builds a prioritised list of possible paging areas that the UE is likely to visit subsequently and makes a decision of the composition of the next tracking area for the UE.

The patent claims several algorithms which define allocation of paging areas in a number of situations. The algorithms take into account the following factors:

- Initiating procedure – UE initiated TAU procedure, MME initiated TAU procedure or Attach Request

- Availability of statistical data for given UE and

Operational results and analytical evaluation for this solution are not available. If the solution reduces the paging load as it claims, this decreases the probability of overload in the PCCH channels. It is unclear whether the Huawei equipment currently incorporates the solution.

2.4.5. MME with smart paging feature by Alcatel-Lucent

Alcatel-Lucent, a provider of equipment and services to mobile operators, markets MME equipment that incorporates a paging optimisation solution. An overview of some signalling solutions of this equipment is available online [20] and is accompanied by a solution sheet [21] and an overview of optimised paging strategies for small cells in metropolitan areas [23]. This paging solution is compliant with 3GPP and relies on the specifications [3] and [22]. Technical documentation for the equipment is not available to non-customers, and due to this fact it is only possibly to provide a brief characterisation of the solution that leaves the technological questions unanswered.

Alcatel-Lucent engineers explain the company's motivation to optimise paging by providing a portion of field statistics from the LTE networks that run on Alcatel-Lucent equipment in the United States. According to the data presented by the company, 28.70% of MME signalling load is paging. In addition, the signalling load generated by the tracking area update procedure (TAU) is 4.90% of the total MME signalling load.

"Smart paging" is an optional feature of Alcatel-Lucent 9471 Wireless Mobility Manager (WMM). This equipment is a scalable MME/SGSN with all features attributed to these network elements.

Smart paging by Alcatel-Lucent is a technique that optimises the number of tracking area in UE TA lists. Similar to the Verizon solution described in Section 2.4.3, this solution uses the UE-related statistics and analyses the pattern of movements of each UE in the network. For a large number of UEs, i.e. human users, daily and weekly patterns are cyclic. This fact allows engineers to base algorithms on it that predict the cell where MME can find a particular UE with a certain probability. Thus WMM constantly updates tracking area lists associated with every UE.

The "Smart paging" solution reduces the volume of paging as well as the TAU load. The benefits in terms of the paging load can reach 80% reduction as the company claims. As a result, the risk of paging overload decreases, although the PCCH channels are not protected from overload.

2.4.6. Scientific contributions

There is limited volume of scientific work indirectly related to the paging overload or paging optimization aspects. A search performed by the author has not revealed any publications that would directly target paging overload. This area of development is dominated completely by suppliers of equipment and services and by mobile operators. This section offers a selection of relevant scientific publications on the topic.

A large share of attention in the research community is concentrated in the profile based optimisation of signalling procedures in mobile networks. Thus there are several articles, [24] dated 2002 and [25] dated 2007, by a group of authors who work on location management schemes. The authors propose paging strategies based on a mobility pattern-based scheme (MPBS). Such scheme maintains user profiles that include statistical information on movements of UEs. Simulation results suggest that MPBS can reduce both the paging traffic and the time required for the paging procedure.

Purely mathematical approach to paging optimisation is often proposed in publications [26]. This approach can be useful to define complementary mechanisms for paging solutions, such as computation of thresholds or prediction of paging load versus TAU load for compositions of cells in tracking area lists.

Non-standard methods of composition of tracking areas and tracking area lists are proposed in [27]. This work targets optimisation of the paging success rate by applying a "tabu search approach" for configuration of tracking areas. This approach considers a mobility factor and the volume of paging traffic in cell as the input parameters for optimisation. The algorithm predicts the paging success probability in that cell if it is transferred to another tracking area. If the computation gives a positive result, the tracking areas are re-shaped, and the previous tracking area of the cell is added to the tabu list. This algorithm repeats until it finds the best tracking area for the given cell. It is unclear however, how the paging success probability is computed.

2.4.7. Current standardisation status and commercial availability

A simple analysis of the relevant 3GPP specifications shows that the MME paging procedure has remained unchanged throughout all 12 releases of the 3GPP LTE. Hence it is unreasonable to expect significant changes in the upcoming Release 13 or the later releases, if any.

Although suppliers of equipment and services for mobile networks continue bringing proposals to discussion of the 3GPP technical groups, significant changes to the specifications are rarely adopted. This is natural because companies have already made serious investments into the LTE infrastructure. Proposals that consider modifications to the software or addition of independent network elements have higher acceptance rate. Hardware modifications are not even proposed at the current stage of the development of the standard.

Manufacturers of LTE infrastructure equipment have a high degree of freedom in implementing their preferred modifications to the signalling subsystem. They can implement solutions in the equipment, and if the impact of such solution is limited by the network entities supplied by a single vendor, there are no interoperability conflicts in the network. An example of such solution is the paging discard timer that is not defined in the 3GPP specifications. The timer allows the base stations to remove expired paging messages from processing. Realised by some vendors, it is their proprietary solution that resides inside the eNb and effectively improves the technical vulnerability of the paging mechanism without any negative impact on the network entities supplied by other vendors. However, the details on such proprietary solutions are only available to the vendors and their customers, and therefore they cannot be presented in this research

3. Modelling and analysis

3.1. Mathematical Model for MME Paging

This research proposes a solution that requires a mathematical model for the MME paging procedure at the base station. An appropriate model has been proposed and characterised in [6]. The present section delivers the basic explanations of the model and provides the details on its applicability in the paging overload solution.

The MME paging model represents a retrial queue similar to the one characterised in [5], [13] and [14]. It includes a service facility and an orbit. The service facility has a limited service capacity. When the service facility is not full, it serves the customers and they leave the system. If the service facility is full, customers join the orbit. They wait in the orbit a predefined period, that is equal for all customers, and attempt to enter the service facility. If the service facility is not full, they are served; otherwise they leave the system not served. Figure 3 illustrates this retrial queue system.

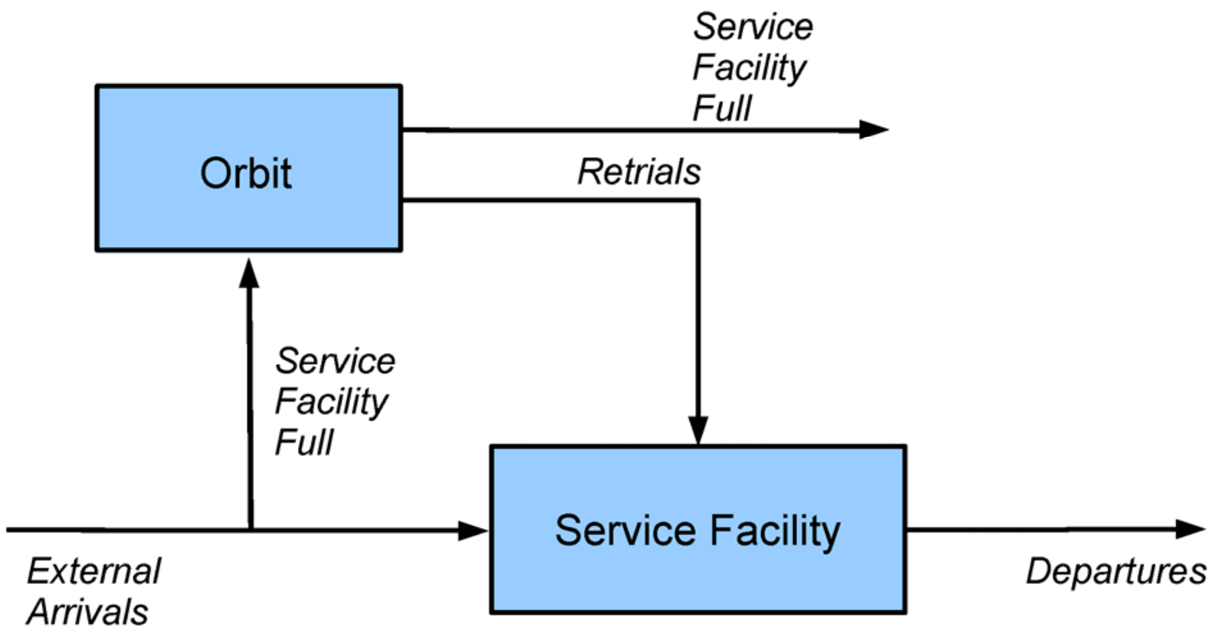

Figure 3. Retrial Queue Model.

The retrial queue model is applicable entirely to the paging procedure at the base station. The service facility describes the operation of the PCCH buffer in combination with the paging channel. Thus the paging messages arrive at the base station and are placed in the buffer, if it is not full. They are discarded, if the buffer is full, and secondary pages are issued by the MME upon expiration of the T3413 timer. It is a single-server queue since there is a single PCCH buffer in the base station that is unloaded by synchronized PCCH transmissions.

The mathematical model assumes that the primary paging messages arrive according to a Poisson process with intensity λ. PCCH transmissions are only possible in predefined subframes – paging occasions. The model assumes the serving time to be distributed exponentially with mean $1/\mu$.

The PCCH buffer has a limited capacity of K paging records. It discards the paging messages which arrive at a moment when the buffer does not have free positions. Thus there is a probability that the buffer is full – q. Discarded primary paging messages arrive in the orbit to wait the interval which is defined by the T3413 timer and modelled as an exponential distribution with mean 1/θ.

The model considers a single retrial attempt for every discarded primary paging. This is the normal configuration of the MME paging procedure. It is also normal when MME uses IMSI for the repeated paging for the reasons explained in Section 2.2. IMSI is a 64-bit identifier that is longer than the 40-bit long S-TMSI which is used for the primary paging. Thus the serving time of the repeated paging message can be related to the serving time of the primary paging message and is defined by an approximation $\frac{64}{40} \cdot \frac{1}{\mu} = \frac{8}{5\mu}$.

Having two serving rates, one for the primary paging and another for the repeated paging, complicates the model. Hence it is sensible to define a reasonable approximation for the average serving rate for both types of the paging messages. The arrivals rate for the primary paging can be described as $\lambda(1\text{-}q)$. This definition treats the primary paging flow that arrives from outside the system independently of the secondary paging. The arrival rate of the repeated paging, i.e. the flow from the orbit to the serving facility, is $\lambda\, q\, (1\text{-}q)$. Given the probability that the buffer is full – q, the probability of a new paging message originating from the orbit is $\frac{1}{1+q}\frac{q}{1+q}$. Hence the average serving rate that accounts for both primary and repeated paging is defined by the following equation:

$$M = \frac{1}{1+q}\,\mu + \frac{q}{1+q}\,5\,\mu\,/\,8 = \frac{\mu\,(1 + 5\,q\,/\,8)}{1+q}.$$

The process *(C(t),N(t))* describes the retrial queue model. *C(t)* is the number of customers in the service facility at time *t*. *N(t)* is the number of customers in the orbit at time *t*. With respect to the size of the PCCH buffer and its occupancy, the model has following transition rates of the aggregated process *(C(t),N(t))*:

$$(c,n) = \begin{cases} (c, n+1)\ with\ rate\ \lambda\ if\ c = K \\ (c, n-1)\ with\ rate\ n\theta\ if\ c = K \\ (c-1, n)\ with\ rate\ M\ if\ c > 0 \\ (c+1, n)\ with\ rate\ \lambda\ if\ c < K \\ (c+1, n-1)\ with\ rate\ n\theta\ if\ c < K \end{cases} \qquad (1)$$

The orbit system can be modelled as an *M/M/∞* queue with arrival rate λq and service rate θ. This is based on an assumption that events of paging discard are independent and distributed identically. The number of users in the orbit is described by a Poisson distribution with mean n* = λq / θ. This results in a plain process $\hat{C}(t)$ which is characterised by the following parameters:

- Arrival rate $\lambda + n^*\theta = \lambda\ (1+q)$, if $\hat{C}(t) < K$
- Arrival rate 0, if $\hat{C}(t) \geq K$
- Departure rate M

Thus the process $\hat{C}(t)$ is an *M/M/1/K* queue.

The equation (2) describes the steady-state distribution of the number of customers in the process $\hat{C}(t)$:

$$\Pi(k) = \frac{\left(1 - \frac{\lambda(1+q)}{M}\right)\left(\frac{\lambda(1+q)}{M}\right)^k}{1 - \left(\frac{\lambda(1+q)}{M}\right)^{K+1}}, k = 1, \ldots, K \qquad (2)$$

The probability of paging discard (serving facility is full) - q is the solution of the fixed-point equation $q = \pi(K)$ that results from an approximation of the process $C(\cdot)$ by $\hat{C}(\cdot)$. It is defined by the following equation:

$$q = \frac{\left(1 - \frac{\lambda(1+q)}{M}\right)\left(\frac{\lambda(1+q)}{M}\right)^K}{1 - \left(\frac{\lambda(1+q)}{M}\right)^{K+1}} \qquad (3)$$

The fixed point equation (3) has a unique solution that is found by fixed point iterations as it is defined in the equation (4). It converges for any initial $q^0 \in [0,1]$.

$$q^{(n+1)} = \frac{\left(1 - \frac{\lambda\left(1 + q^{(n)}\right)}{M}\right)\left(\frac{\lambda\left(1 + q^{(n)}\right)}{M}\right)^{K}}{1 - \left(\frac{\lambda(1 + q^{(n)})}{M}\right)^{K+1}} \qquad (4)$$

When the system is in a steady state there are equal probabilities of discard for primary and repeated paging requests, i.e. for the requests that are arriving from outside and from the orbit respectively. The basis for this is the fact that these events are independent and each position in the PCCH buffer (which is a part of the serving facility) can accommodate one paging request regardless of whether the request is primary or repeated.

Having the probability of discard of a paging request - q, the probability that both primary and repeated paging requests by one user are unsuccessful is defined by the following equation:

$$p_{fail} \approx q^2 \qquad (5)$$

In terms of the LTE system, the probability defined by (5) refers to the failure of a connection establishment attempt. The mathematical model established in this section is an important instrument in the solution to paging overload, and hence it is evaluated in Section 3.3 prior to incorporation to the actual solution.

3.2. Description of the Simulator

Development of the solution to the paging overload requires a precision simulator of the MME paging procedure and the software implementing the mathematical model that is characterised in Section 3.1.

This research employs Matlab to build a simulator that models the complete MME paging mechanism with the features listed below.

Functionalities of MME

- Generation and transmission of primary and repeated S1AP paging messages
- Monitoring of the T3413 timer

Functionalities of eNb

- Scheduling of DRX cycles
- Scheduling of paging frames and paging occasions

- Scheduling of BCCH modification periods
- PCCH buffer for S1AP paging messages and buffering mechanism
- Generation and transmission of the RRC paging messages
- Monitoring of the expiry

Statistics

- Aggregated statistics for all paging-related KPIs
- Real-time statistics for the PCCH buffer, S1 interface and air interface

Neglected aspects and procedures

- Scheduling transmission time on the S1 interface
- Propagation/transmission time on the air interface
- Random assignment delay and failures
- SFN assignment

Further details on the neglected procedures follow in Section 3.3.

Additional capabilities are introduced into the simulator to implement the paging overload solution. They are characterised in Section 4.

3.2.1. Configuration of the simulator

The model implements the frame structure type 1 with 5 MHz bandwidth allocated to the cells. Table 1 lists the system parameters that are used in the model for the evaluation of the operation of the paging channels the subsequent sections.

Table 1. Settings for the simulation.

Parameter	Value
System bandwidth	5 MHz
Frame structure	Type 1
Number of frames in the DRX cycle	128
BCCH modification period	2
nB coefficient (in proportion to DRX cycle)	1/16
Maximum number of S-TMSI paging records in the RRC paging request	7

Timer T3413	5000 ms
Maximum number of retransmission attempts	1
Primary paging type	S-TMSI
Secondary paging type	IMSI
Size of the paging buffer at eNb (in paging records S-TMSI or IMSI)	140

The system bandwidth of 5 MHz is the minimum allocated to operators by the regulator (it is 5MHz downlink and 5MHz uplink). This allocation represents the most unfavourable conditions for optimisation. In most countries operators receive multiples of 5MHz bandwidth.

Frame Type 1 corresponds to LTE FDD which are more common systems, while LTE TDD are rather complimentary systems at present.

The values of most parameters in Table 1 are a normal configuration commonly used by operators. This applies to DRX cycle, BCCH modification period, timer T3413, primary and secondary paging types.

Coefficient nB in combination with the maximum number of paging records in RRC paging request define the paging capacity. The values for the two parameters are selected to match very high load conditions that are considered in this research.

The size of the PCCH buffer varies from vendor to vendor. A PCCH buffer of 140 positions is chosen to match the settings of T3413 timer, as the paging messages must not expire in the buffer in an adequately configured system.

It is normal for operators to configure a single repeated paging transmission and use IMSI for it. S-TMSI, a more compact identifier, is always used for the primary paging.

3.2.2. Paging capacity on the radio interface

Computation of the installed paging capacity does not need to be implemented in the simulator as it operates on the basis of the Table 1 settings and the 3GPP specifications. This section provides an explanation of how the installed paging capacity of a base station can be calculated for the purpose of the performance evaluation.

Thus the installed paging capacity of a cell is defined according to the following algorithm with the numbers based on Table 1.

- Duration of the DRX cycle (discontinuous reception cycle) is defined by the number of frames in it. The frame is 10 ms, thus the DRX cycle of 128 frames has a duration of 1280 ms.

- Number of paging occasions in a DRX cycle is defined by the number of paging occasions configured in a single frame (1/16 in the experiment). Thus, a DRX cycle with 128 frames provides 8 paging occasions.

- Each of the 8 paging occasions provides capacity for a maximum of 7 S-TMSI paging records (one RRC paging request). Thus 56 user terminals can be paged with S-TMSI in a single DRX cycle of 1280 ms.

- 56 paged user terminals in 1280 ms DRX cycle corresponds to 43.75 S-TMSI pages per second.

More details on the paging capacity for a number of eNb configurations follow in Section 4.2.5.

3.2.3. Characterisation of the traffic

The simulator generates the primary paging messages on the S1 interface according to a Poisson arrival process with average inter-arrival time defined by the BHCA parameter (Busy Hour Call Attempts).

BHCA in this research defines the number of connection establishment attempts per hour and is used exclusively as descriptive but not precise term. BHCA must be differentiated from the number of paging requests per hour (paging load), since a connection establishment may require repeated paging.

The average inter-arrival periods considered in the presented experiments range from 12 to 72 ms. This corresponds to a range of BHCA values from 300,000 to 50,000 respectively.

Unreachable user terminals, i.e. the UEs that went out of coverage without communicating this event to the network, typically represent 3% to 5% of the total number of users registered in a tracking area. In a real network this percentage depends on the quality of the network planning, quality of maintenance and quality of optimisation, i.e. the professional level of the engineers. Although the simulator has a functionality of introducing unreachable users, the experiments use a value of zero percentage of unreachable users. This is due to the fact that unreachable user terminals do not affect the performance of the paging channels. They rather increase the paging load, however this can be achieved more effectively by adjusting the BHCA parameter of the primary paging load on the S1 interface.

The simulations include paging with target BHCA values in a tracking area. All cells in the TA have an identical configuration that provides paging capacity of 43.75 S-TMSI per second as described earlier.

Paging response messages (service requests in the uplink) and the corresponding control channels are not modelled with the 3GPP specification precision in the simulator. A paging message that was transferred on the air interface is considered to have reached the destination UE which has sent a paging response upon its reception.

3.3. Operation of paging channels

The results that are discussed in this section refer to the standard operation of the paging channels according to the 3GPP specifications without the optimisation which is proposed later.

Operation of the paging channels is characterized by a number of KPIs that are commonly applied by the LTE equipment suppliers.

- Average connection establishment time
- Average length of PCCH queue
- Average queuing time in PCCH buffer
- Average served paging load per hour
- Paging success rate
- Probability of connection establishment failure (paging contribution only)
- Probability of paging discard

It is reasonable to use the KPIs that are commonly recognised and applied by professionals, and thus the PCCH simulator monitors and reports all KPIs that are listed above. They are sufficiently accurate for a balanced assessment of the PCCH performance.

KPI: Probability of connection establishment failure

Probability of connection establishment failure is defined as a ratio of the total number unsuccessful connection establishment attempts to the total number of connection establishment attempts.

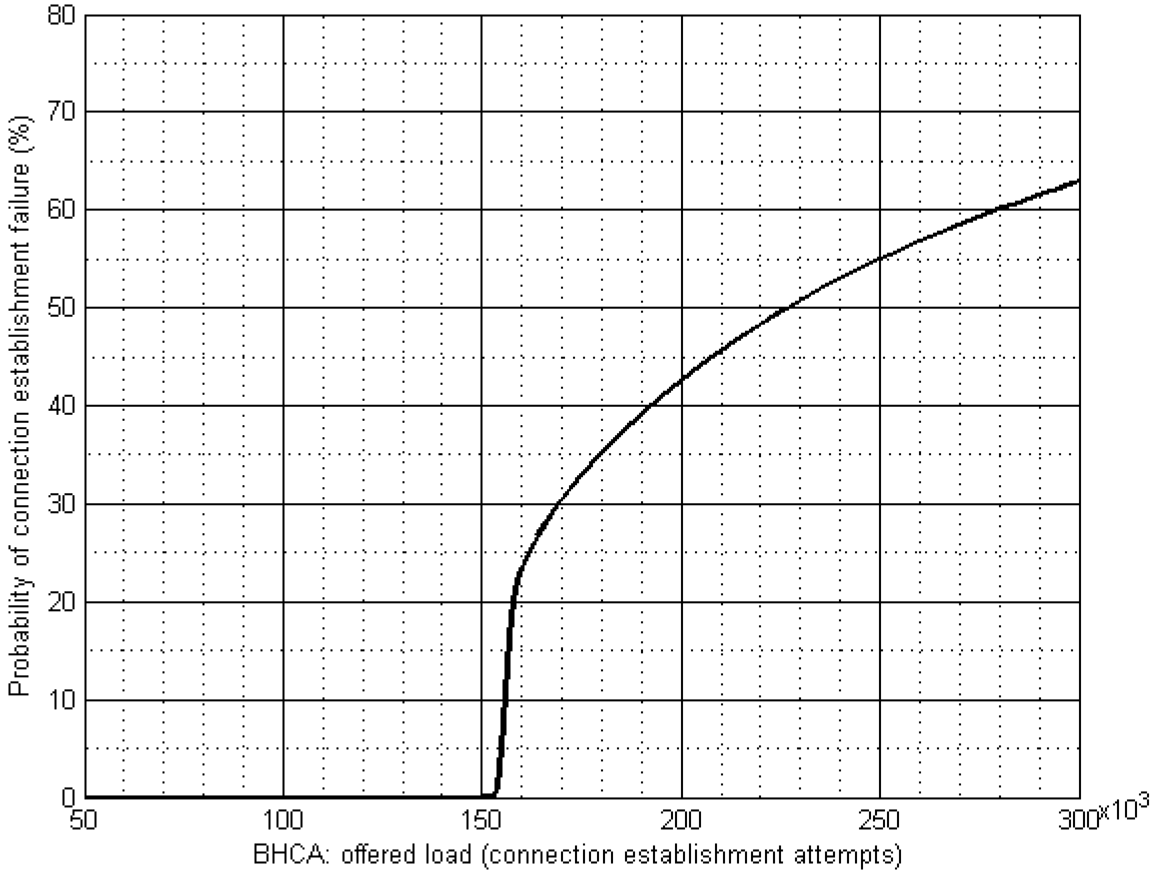

Figure 4. Probability of connection establishment failure is proportional to the paging load that cannot be served due to lack of paging capacity.

With the installed paging capacity of 157500 BHCA it is reasonable to expect non-zero probability of failure at the paging load slightly below this number. This is defined by the distribution of the arrival times that is never uniform in telecommunication networks.

The results displayed on Figure 4 underscore the fact that the probability of connection establishment failure rapidly grows to an unacceptable level at the paging load of over 150000 BHCA. This is approximately 5% lower than the installed paging capacity, and hence this experience will be considered the definition of the PCCH reconfiguration threshold.

The following factors have a limited influence on the paging success rate but are neglected in this research: (A) propagation / transmission time on the S1 interface and radio interface, and (B) RRC connection request intervals and failures. The reason for these factors being neglected is their small and limited contribution to the KPI. It is the order of microseconds for propagation time and milliseconds for RRC connection requests. This contrasts with a 5000 ms value of the T3413 timer for expiry of paging messages.

KPI: Probability of paging discard

Probability of paging discard is defined as a ratio of the number of discarded S1AP paging messages to the total number of S1AP paging messages generated by the MME.

An S1AP paging message is discarded when the PCCH buffer is full, i.e. it is not possible to process the message at the base station.

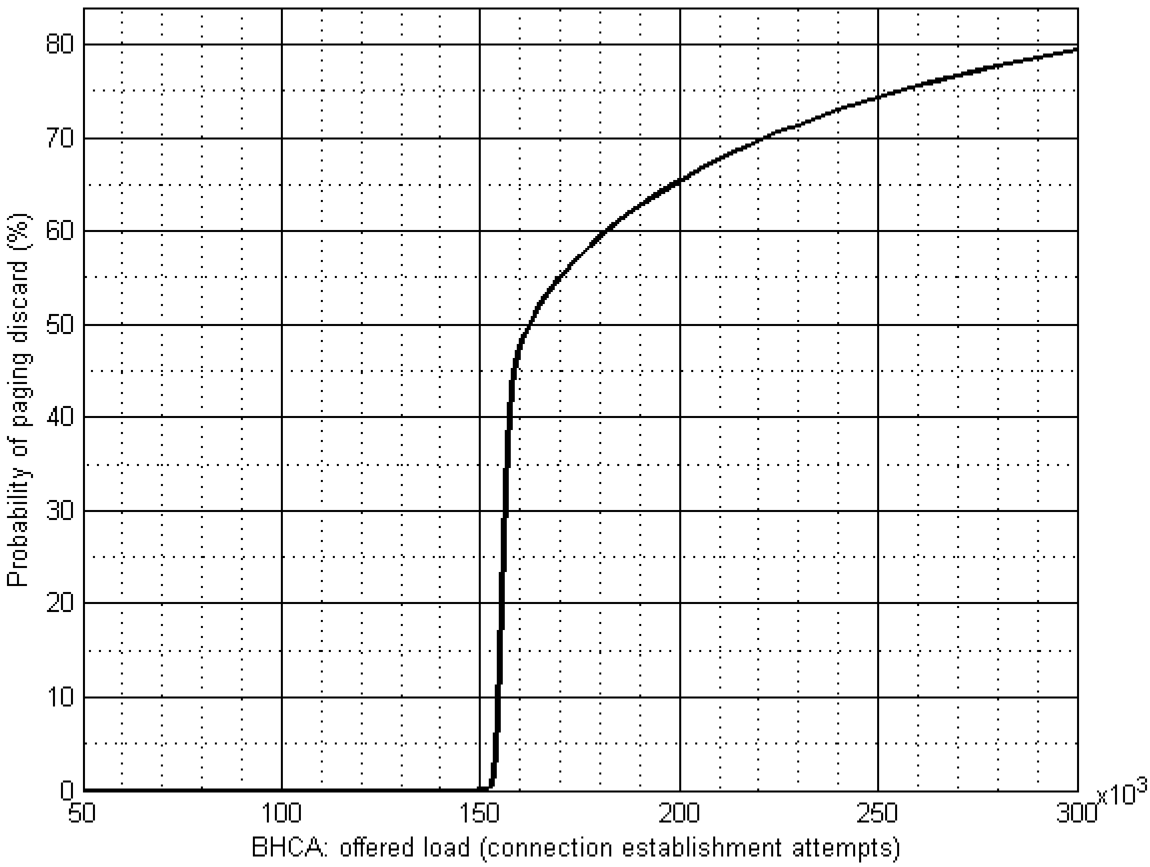

Figure 5. Probability of paging discard is higher than the probability of connection establishment failure due to repeated paging failures.

Probability of paging discard is not equal to probability of failure of connection establishment attempt. The reason for this is the possibility that a single connection establishment attempt may require repeated paging.

The assessment reveals a similar non-zero threshold for the KPI at approximately 150000 BHCA (Figure 5), thus underscoring the requirement for an earlier PCCH reconfiguration that must take place before the paging load reaches the installed capacity.

Auxiliary primary and secondary paging statistics is monitored in the simulator (Figure 6). It is not necessary for the analysis of PCCH operation since the aggregated statistics is more representative and sufficient. Nevertheless, the simulator maintains a functionality of separate monitoring for future use.

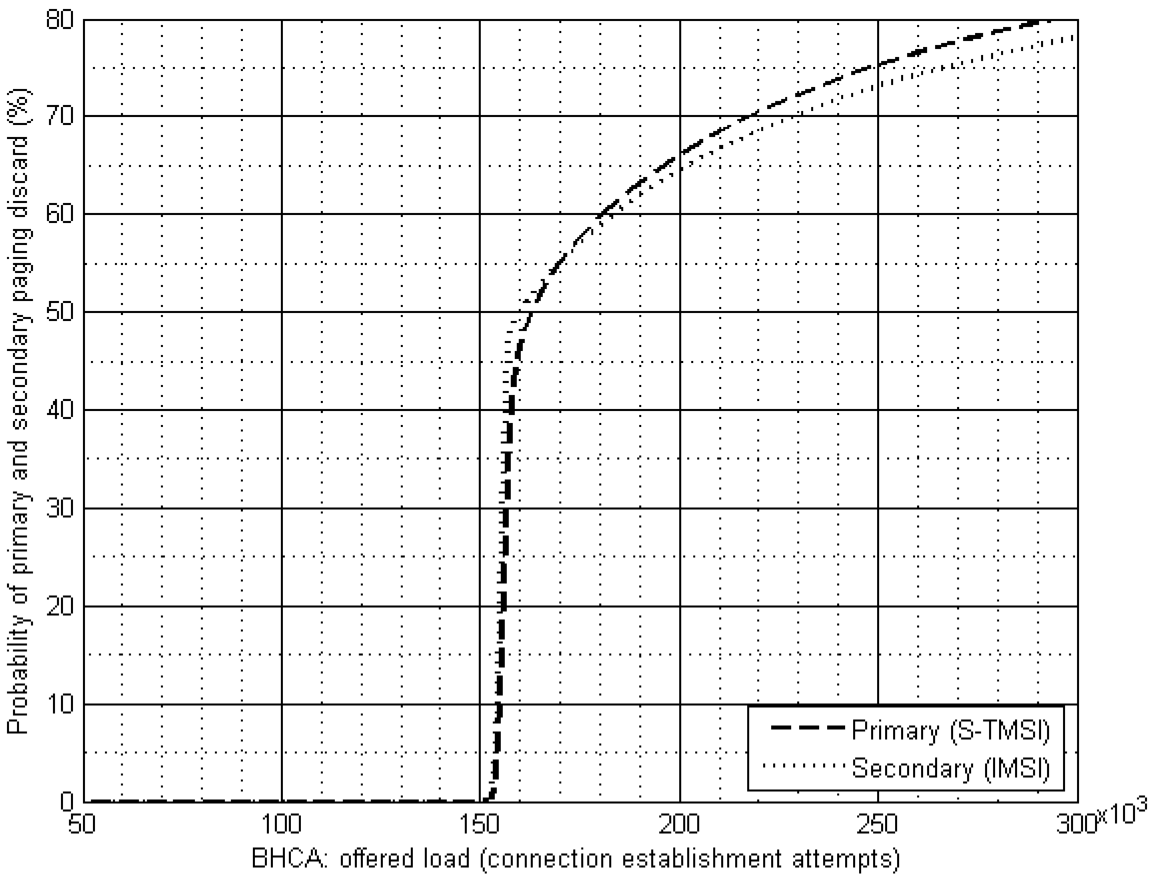

Figure 6. Separate statistics for primary and secondary paging is monitored in the simulator for future use.

<u>*KPI: Paging success rate*</u>

Paging success rate is defined as a ratio of the total number of the paging messages responded by user terminals of the total number of S1AP paging messages generated by the MME.

Propagation / transmission time on the S1 and radio interfaces and immediate assignment intervals and failures are not taken into account for the reasons explained earlier in this section.

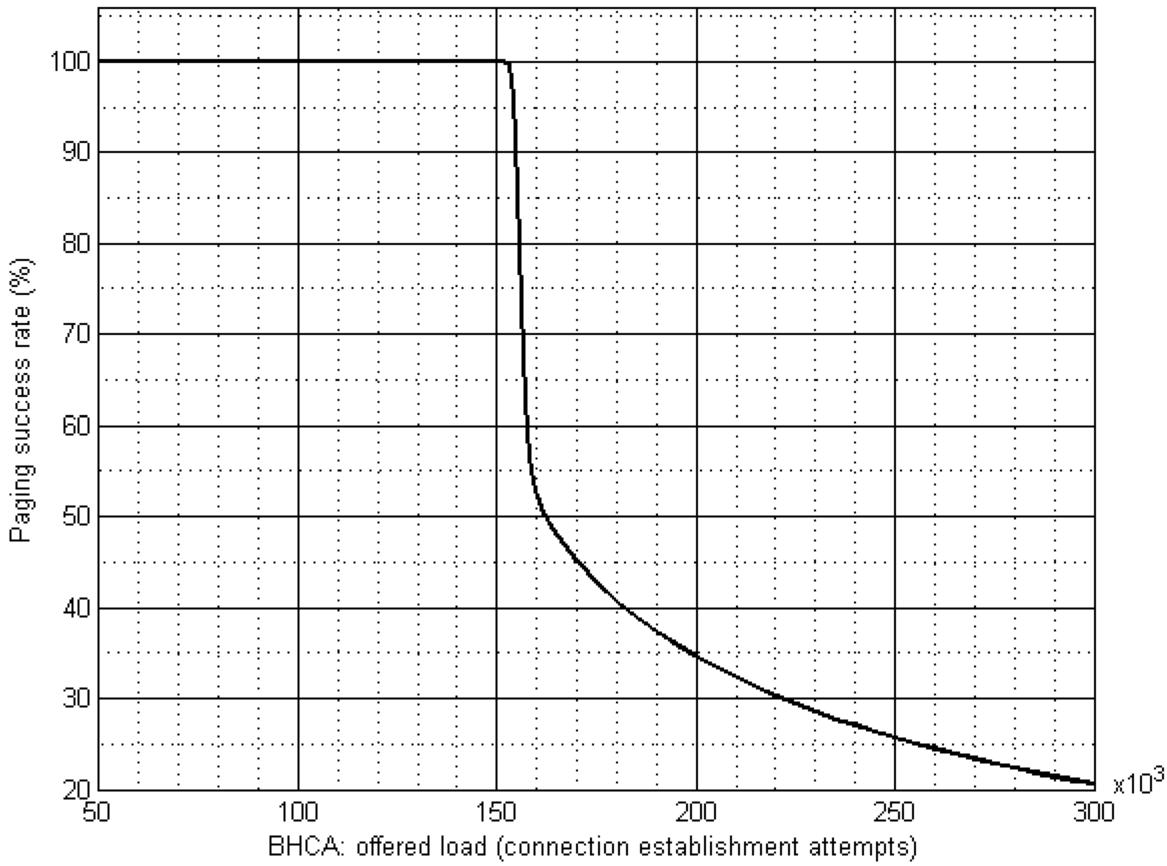

Figure 7. Paging success rate – the main paging-related KPI used by operators.

Paging success rate is among the major KPIs monitored by operators. It is analyzed on a basis of daily intervals. OMCs normally provide a possibility to zoom into the daily intervals and retrieve data for 30 minute intervals.

Typical values of paging success rate are above 96%. In well-managed networks the paging success rate is in a range from 98.00% to 98.80%. Such performance is achieved by a complex optimization of radio access network that includes numerous KPIs not directly related to the paging procedure. It is not possible in practice to reach 100% paging success rate as there are inevitably UEs that drop out of coverage with no notification to the network.

Paging discard observed on earlier graphs causes degradation of the paging success rate as it is seen on Figure 7. Same threshold of approximately 150000 BHCA marks the limit of reliable operation.

In a real network the paging success rate is normally worse than the paging discard because it is a broader indicator that includes a larger number of signalling procedures.

<u>*KPI: Average served paging load per hour*</u>

Average served paging load per hour is defined as the total number of successful paging attempts in an hour of operation of the network.

Served paging load may exceed the offered load (BHCA) due to repeated paging. It can also be lower than the offered load (BHCA) in a busy network due to paging discard.

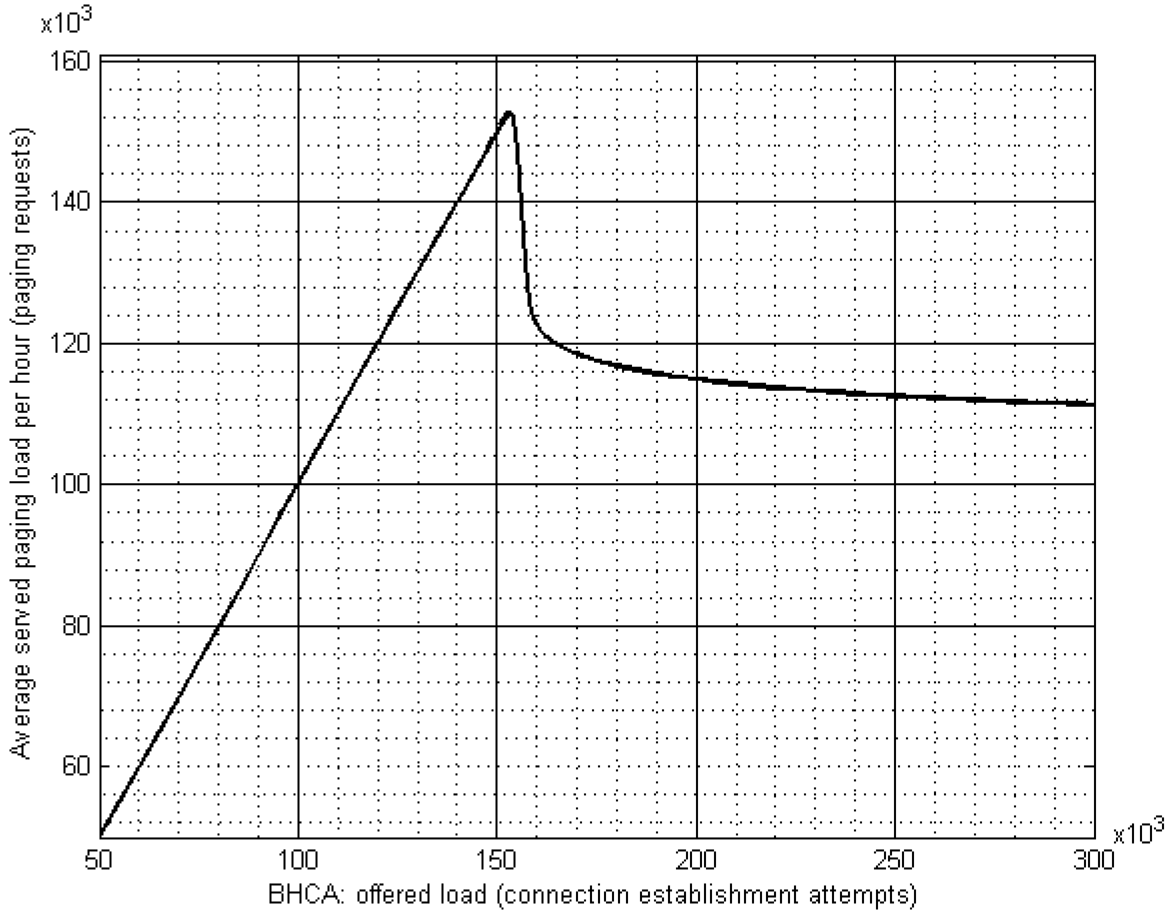

Figure 8. Served paging load and thus user traffic drop as a consequence of decreasing paging success rate.

Figure 8 demonstrates the major problem of paging overload: the served paging load and, as a result, the served user traffic decreases when the installed paging capacity is insufficient to establish connections. The figure shows an expected proportional increase of the served paging load until the installed paging capacity is exhausted, approximately 157500 BHCA. It is reasonable to expect the served paging and user load not to drop but remain stable at the maximum capacity. However this is not possible as the system triggers repeated paging with IMSI. A paging record containing IMSI requires more resources in the paging channels than the S-TMSI based paging. Thus the served load rapidly drops to below 120000 BHCA. As a result, over 20% of the installed traffic capacity cannot be utilized due to insufficient paging resources.

KPI: Average length of the PCCH buffer

The values of this KPI are based on the actual length of the PCCH queue. It is logged at every operation that involves the PCCH buffer: placement of S1AP paging messages in the PCCH buffer, discard of S1AP paging messages at the base station, transmission of RRC paging messages, expiry of a paging message in the PCCH buffer.

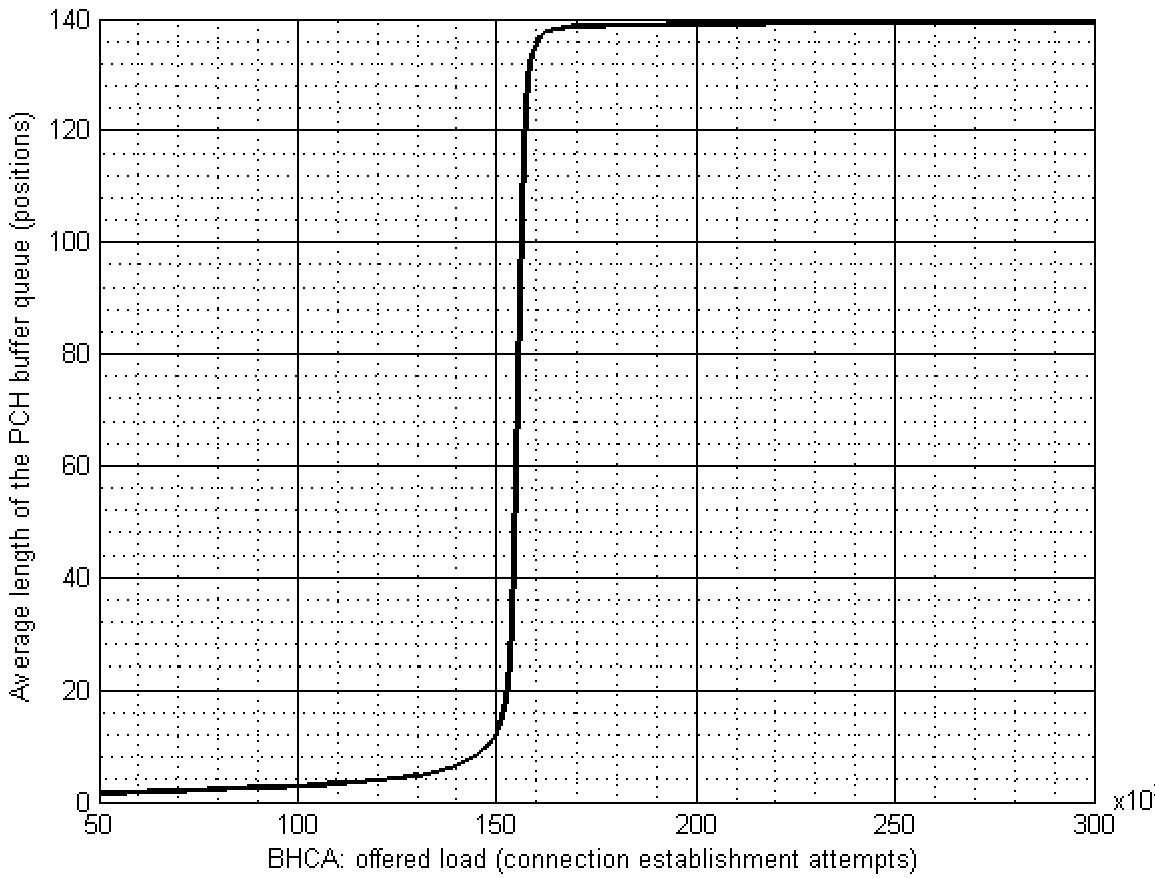

Figure 9. The queue in the PCCH buffer shall not exceed 15-20% of its total capacity in a healthy network and normal load conditions.

The queue in the PCCH buffer must not grow for prolonged periods when the capacity of the paging channels match the paging load. Momentary increase of the queue is normal, but a single DRX cycle must be sufficient to unload the buffer after such burst. 10% is a normal occupancy of the PCCH buffer. Prolonged operation at over 20% of the buffer capacity is potentially unstable.

KPI: Average queuing time in PCCH buffer

Average queuing time in PCCH buffer is defined as the time between the arrival of an S1AP paging message to the base station and the moment of transmission of the corresponding RRC paging message on the radio interface.

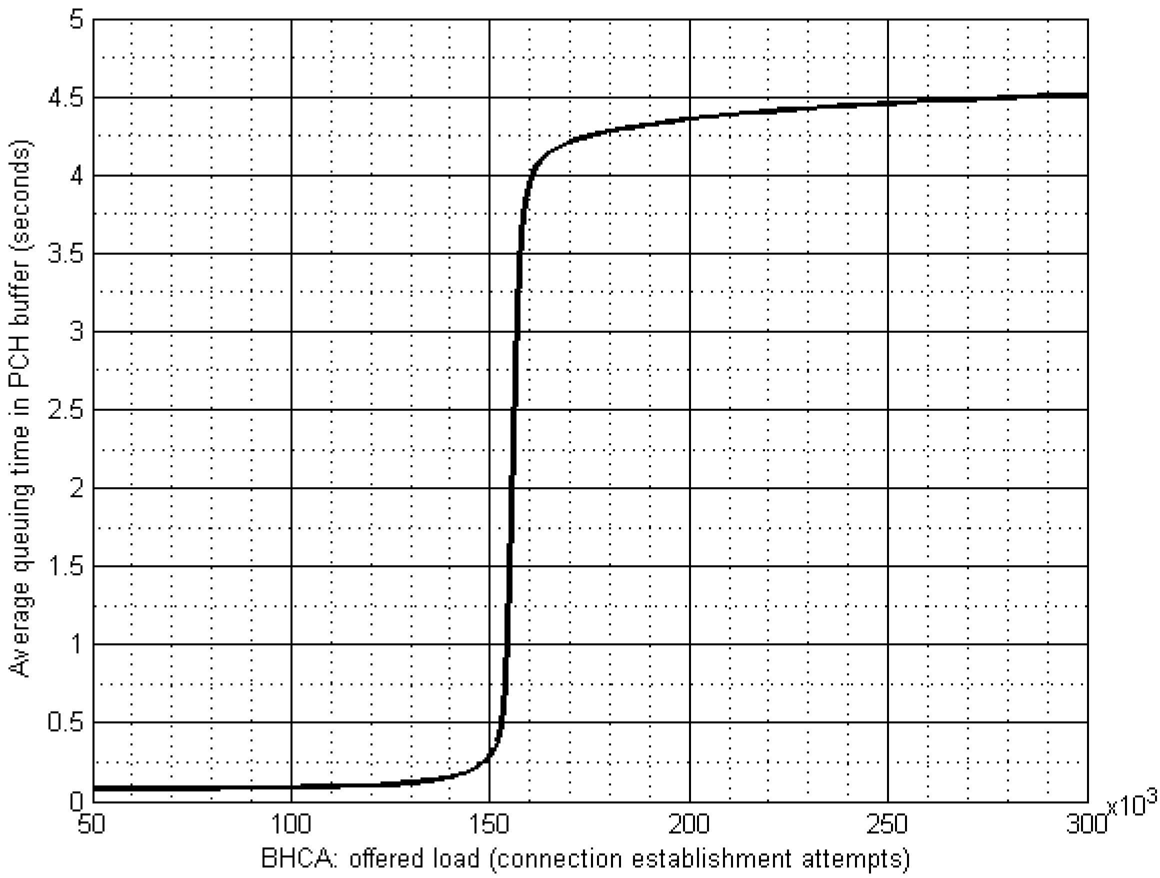

Figure 10. Queuing time in PCCH buffer must never exceed the value of the T3413 timer (5000 ms in the simulation).

Statistics on Figure 10 reveals unacceptable queuing times for paging loads over 150000 BHCA. The queuing time cannot grow above ~4600 ms for the configured paging capacity and the PCCH buffer (see Table 1).

Figure 10 paired with Figure 5 “Paging discard” show that when the paging load exceeds the installed capacity of 157500 BHCA, between 50% and 80% of paging messages are blocked and the remaining pages pass an entire waiting cycle in the buffer.

KPI: Average connection establishment time

Average connection establishment time is defined as the time between the arrival of the primary S1AP paging message to the base station and the moment of transmission of the corresponding RRC paging message on the radio interface.

In this research the scope of the parameter is limited to the paging procedure and excludes any influence of other signalling procedures that are necessary for establishment of a connection.

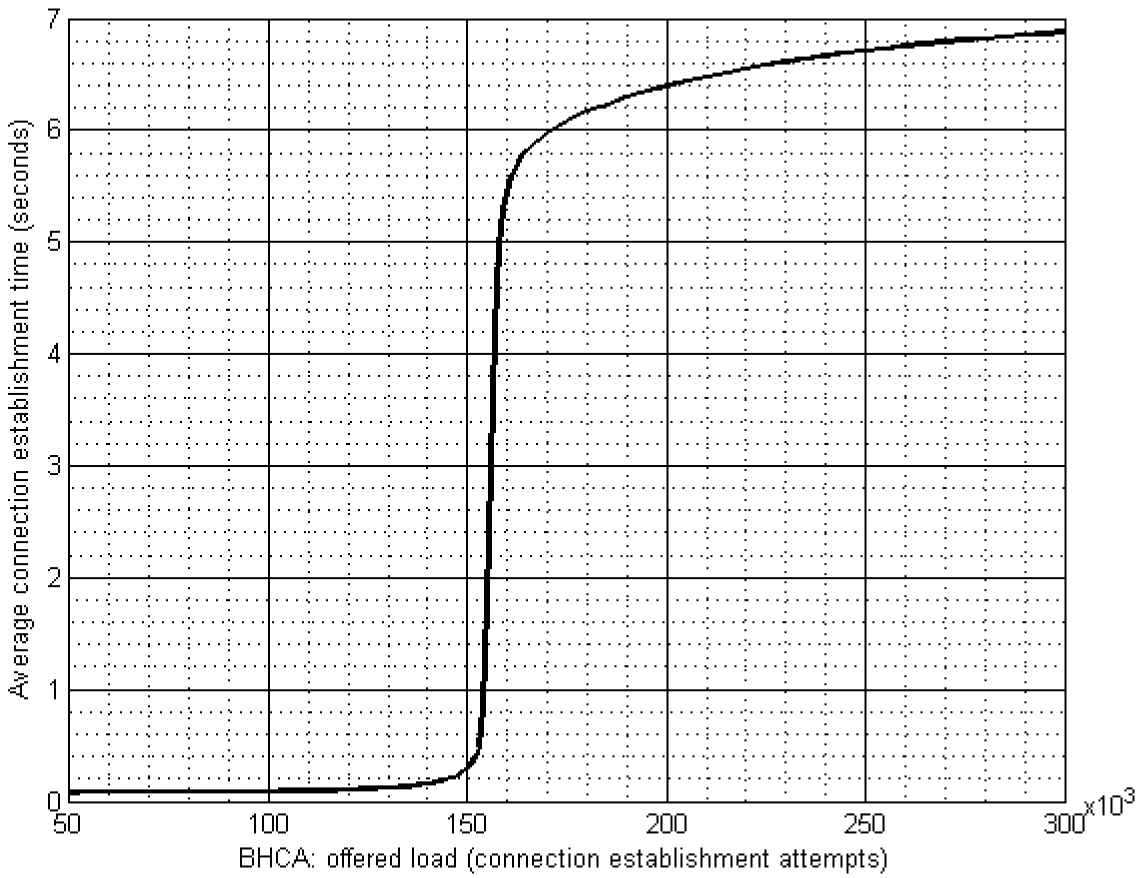

Figure 11. Connection establishment time is affected by the secondary paging.

Prolonged connection establishment intervals reported on Figure 11 are a consequence of PCCH buffer overflow (see Figure 9) and secondary paging. This is in line with the statistics for connection establishment failures reported on Figure 4.

An overall analysis of the operation of paging channels underscores that fact the air interface can process reliably a paging load that is 5% - 7% lower than the installed paging capacity of a base station. This result is of significant practical value because it helps to define the safety margins for the paging overload solution in Section 4.

3.4. Evaluation of the mathematical model for MME paging

Performance of the mathematical model defined in Section 3.1 is confronted with the results of the simulations that are presented in Section 3.3. The model targets the probability that both primary and repeated paging fail. It is the parameter q^2. This parameter corresponds to the KPI probability of failure of a connection establishment attempt which is defined in Section 3.3.

The simulator recognises a connection establishment attempt as successful when either primary or secondary paging request to a particular UE is successful. An unsuccessful connection establishment attempt is logged when both the primary and the secondary paging requests to a particular UE fail.

The mathematical model is processed in Matlab. The software programs the equation (4) and uses iterations in order to solve it. The final value of the probability is defined according to the equation (5).

The simulator applies the system configuration presented in Table 1. Each run simulates 2400 seconds of network operation. Figure 12 shows the resulting curve for the probability of connection establishment failure. It is based on the averaged values from 500 runs. Figure 12 reports the results that are acquired from the mathematical model in comparison with the results of the simulations.

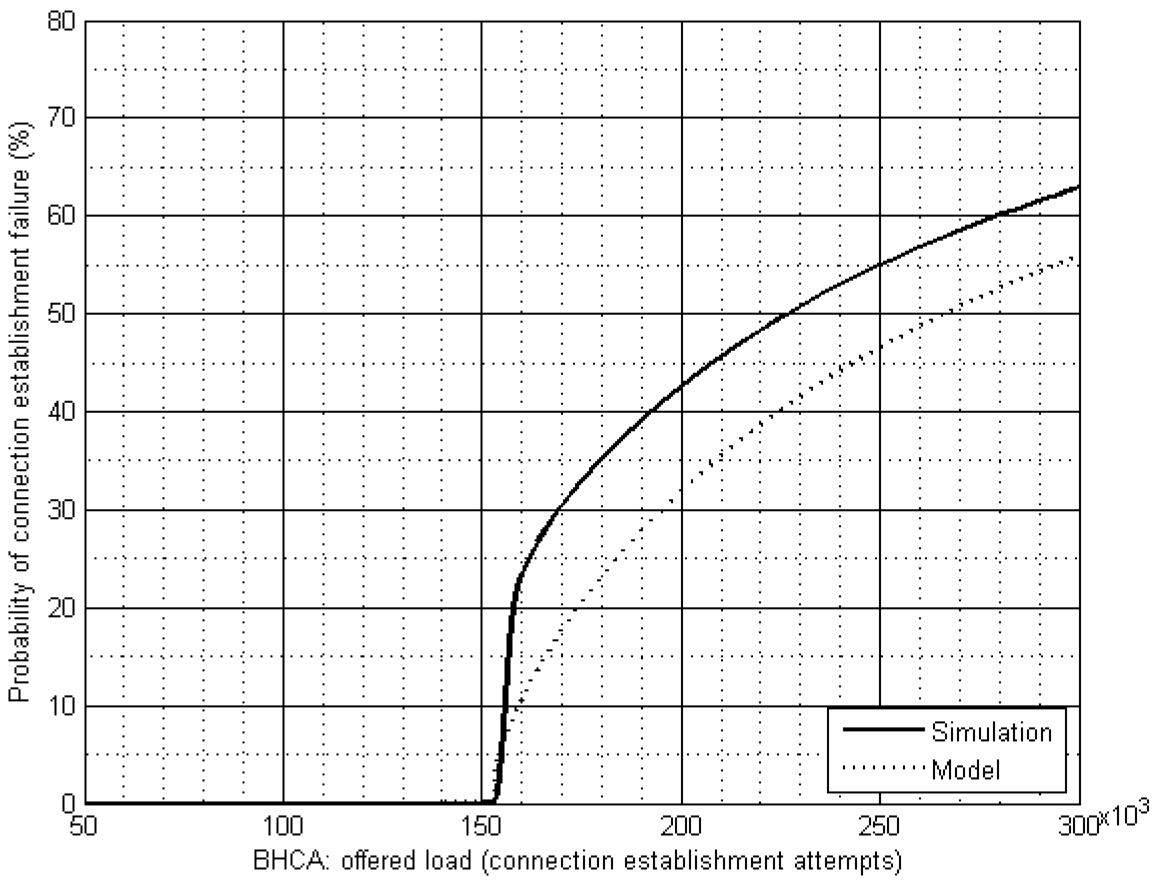

Figure 12. Probability of failure of connection establishment attempt acquired from the mathematical model and the PCCH simulator.

The results reported in Figure 12 reveal the fact that the mathematical model is precise for the paging loads below the non-zero threshold. It captures the non-zero threshold with a certain precision. Figure 13 zooms into the non-zero threshold zone of the graph in order to further assess the accuracy of the mathematical model.

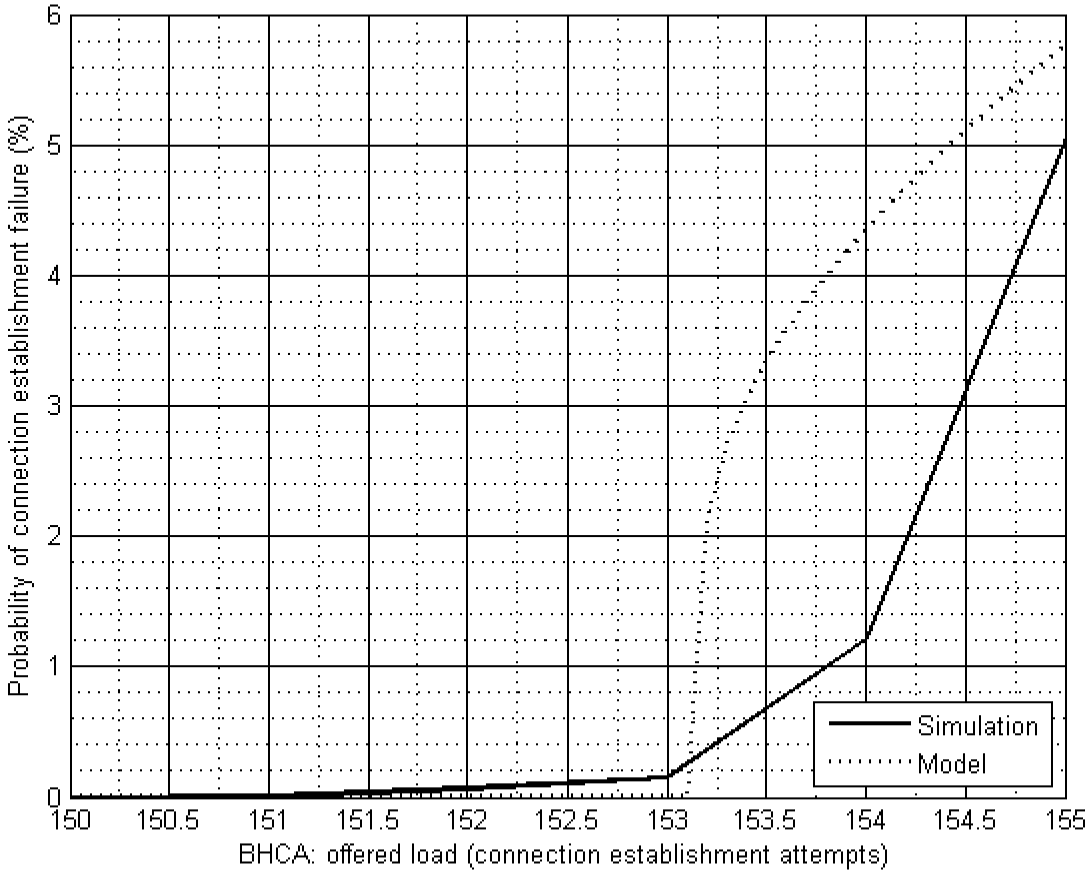

Figure 13. The non-zero threshold of the probability of connection establishment failure.

The results from the PCCH simulator show the probability of connection establishment failure rising from zero at 151000 BHCA where it reaches 0.0058%. It reaches 0.14% at the load of 153000 BHCA.

The mathematical approximation captures the non-zero probability threshold at 153200 BHCA where the probability is 2.11%. It is slightly more optimistic than the results obtained from the PCCH simulator. This fact will be taken into account in the development of the PCCH overload solution since the blocking in the signalling channels is unacceptable.

For the paging loads above the non-zero threshold the graphs show a growing divergence in the results acquired from the mathematical model and the simulator. The difference reaches a high of approximately 14%. The imprecision of the mathematical model for the loads above the non-zero blocking threshold is acceptable since these values of the paging loads are outside the operating range of the PCCH channels in their given configuration.

4. Solution and validation

Signalling traffic such as the paging is critical for the operation of the network. When the signalling channels are overloaded there are two possibilities to keep the network at the maximum load:

(A) Block the services requests that the network cannot allocate resources for, or

(B) Increase the signalling capacity.

The approach (A) “forced blocking” can prevent the decrease in user traffic – an undesirable network behaviour that is characterised in Section 3.3 and in particular on Figure 8. In addition to this, the approach (B) “increase in the signalling capacity” allows the network to accept the growing user traffic. This research focuses on the approach (B) and defines a modification to the operation of the paging channels in live LTE networks.

4.1. Increasing the MME paging capacity in operation

The MME paging capacity on the radio interface is controlled by two parameters:

- Number of paging records in a paging occasion (parameter *MaxPagingRecords*)
- Number of paging occasions per radio frame (parameter *nB*)

Full details on the MME paging capacity are provided in Section 2.2 “MME paging for EPS services”.

Depending on the system bandwidth a maximum of 7 paging records can be configured in a PO for 5MHz systems and 16 paging records in a PO for 10/15/20 MHz systems. If the system operates in a configuration with the maximally possible number of paging records in PO, additional POs must be configured in the DRX cycle in order to increase the paging capacity. This may require additional paging frames in the DRX cycle – if the parameter *nB* is currently set to a value smaller than the default paging cycle (*T*).

Reconfiguration of the paging channels is a routine task for OMC engineers. They perform it in the background mode, i.e. not in a live network, but in a frozen copy of the database. Once the new database is ready OMC engineers perform a swap in the hours with the lowest network load, typically between 02:00 and 05:00, because the operation is associated with a risk of outage of one or more base stations. This way to change the configuration of the network is normal, and it is implemented in the LTE systems of all equipment manufacturers. This research proposes a solution that requires a more complex action – configuration of additional paging capacity while the network is in operation without taking base stations out of service.

4.2. Modified operation of the paging channels

4.2.1. Monitoring procedures for the volume of paging and PCCH buffer

An LTE network in operation can provide two indicators that are useful for early detection of an approaching overload situation in the paging channels:

- Volume of paging
- Size of the queue in the PCCH buffer

The 3GPP specifications do not define monitoring mechanisms for the volume of paging and the PCCH buffer, thus leaving it optional for equipment vendors whether to have them. As of the current design, of the LTE equipment does not implement this monitoring. However it is required for the solution that is proposed in this research.

Thus, it is necessary to introduce new parameters that are not present in the 3GPP specifications and in the solutions by the equipment vendors:

LIMIT_PAGING_PC

LIMIT_PAGING_PC is the upper acceptable limit of load in the paging channels for their given configuration (measured in percentage points).

Value range: 50% - 100%.

This parameter is used to define the PCCH reconfiguration threshold in terms of the paging load. Thus it is confronted with the maximum paging load that the paging channels can serve without blocking according to the following equation:

$$PCCH\_RECONFIG\_LOAD\_THRE = \frac{LIM\_PAGING\_PC}{100\%} \cdot MAX\_PAGING\_LOAD$$

Where:

PCCH_RECONFIG_LOAD_THRE - the PCCH reconfiguration threshold “load” (measured in connection establishment attempts).

MAX_PAGING_LOAD – the maximum paging load that is associated with the zero probability of blocking (measured in connection establishment attempts). It is defined

by using the equations (4) and (5) from Section 3.1.
LIMIT_PCCH_QUEUE *LIMIT_PCCH_QUEUE* is the highest acceptable length of the queue in the PCCH buffer (measured in percentage points of its maximum capacity). Value range: 30% - 100%. This parameter is used to define the PCCH reconfiguration threshold in terms of the occupancy of the PCCH buffer according to the following equation: $PCCH\_RECONFIG\_QUEUE\_THRE = \frac{LIM\_PCCH\_QUEUE}{100\%} \cdot PCCH\_BUFFER\_SIZE$ Where: *PCCH_RECONFIG_QUEUE_THRE* - the PCCH reconfiguration threshold "buffer" (measured in paging positions). *PCCH_BUFFER_SIZE* – the size of the PCCH buffer that is defined by the equipment manufacturer or configured by the operator.

Based on *LIMIT_PAGING_PC,* the eNb software re-computes the reconfiguration threshold *PCCH_RECONFIG_LOAD_THRE* upon every variation in paging capacity that is introduced in this research.

Operators can configure the parameters *LIMIT_PAGING_PC* and *LIMIT_PCCH_QUEUE* individually for particular base stations and tracking areas. It is necessary to take into account at least the following factors:

- Capacity of tracking area
- User behaviour based on the statistics
- Speed of the PCCH reconfiguration (more on this in Section 4.2.6)

When either of the two thresholds, *PCCH_RECONFIG_LOAD_THRE* or *PCCH_RECONFIG_QUEUE_THRE*, is reached, the base station must configure additional paging capacity.

4.2.2. Configuration of additional paging occasions and paging frames

According to the solution that this research proposes, the growing paging load and the queue in the PCCH buffer trigger reconfiguration of the paging channels. In the LTE terms it means an addition of paging occasions to the DRX cycle. Additional paging frames to accommodate the new POs are required for the configurations of the DRX cycle with $nB < T$.

The solution is described by the algorithm that follows on Figure 14.

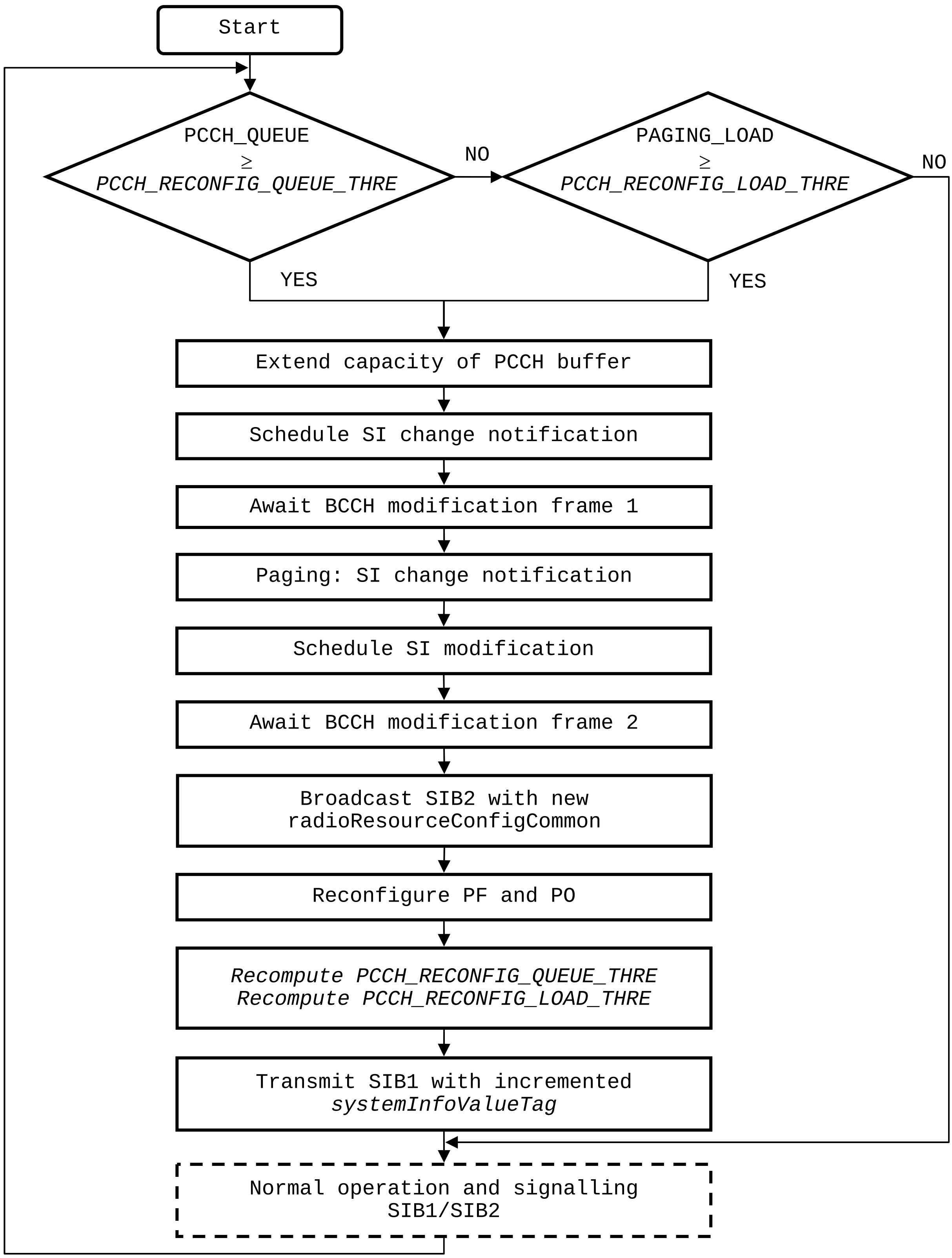

Figure 14. The algorithm for reconfiguration of the paging channels.

The subsequent Section 4.2.3 and Section 4.2.4 offer a detailed explanation of the algorithm and its functioning in normal and high load conditions.

There are 3GPP compliant and non-compliant procedures in the algorithm. Some procedures are features of the equipment. They are not part of the 3GPP specifications and do not affect the 3GPP-specified operation of the base station. A characterisation of the procedures from the 3GPP perspective follows in Section 4.4 "Compliance with the 3GPP specifications".

4.2.3. The solution in operation in normal load conditions

The solution does not introduce any modifications to the normal operation of the base station or the air interface as long as the reconfiguration thresholds are not reached. The only necessary and proposed innovation that is in use at this stage includes the monitoring mechanisms for the paging load and the occupancy of the PCCH buffer. The mechanisms are described earlier in Section 4.2.1.

When operating in normal load conditions, the solution does not require any modifications to the 3GPP specifications. The monitoring mechanisms are to be implemented in the eNb software and do not require any modifications to the eNb hardware.

According to the algorithm (ref. Figure 14), the eNb sends the Master Information Block (MIB) on BCCH → BCH → PBCH in accordance with the normal MIB transmission schedule and in full compliance with the 3GPP specifications [4].

The System Information Blocks Type 1 and Type 2 (SIB1 and SIB2) are transmitted according to the normal procedure on BCCH → DL-SCH → PD-SCH in full compliance with the 3GPP specifications [4].

4.2.4. The solution in operation in high load conditions

When either of the two PCCH reconfigurations thresholds is reached the eNb configures additional paging capacity. This action is associated with a change of the system information. Prior to the actual change the eNb must notify the UEs in its cells. 3GPP offers adequate instruments to change of the system information that it supplies in its RRC protocol specification [4].

The solution employs the following procedure, which is partly 3GPP compliant, in order to reconfigure the paging channels and change the relevant system information:

- **Step 1.** Re-dimensioning of the PCCH buffer. The buffer size is increased to match higher paging capacity and to accommodate the paging load that grows rapidly during the reconfiguration of PCCH channels.

- **Step 2.** The eNb schedules an SI change notification. This action is possible in certain radio frames as defined by the BCCH modification period settings in compliance with 3GPP [4]. UEs are synchronised with the eNb on this action as they calculate the BCCH modification period from the parameter *modificationPeriodCoeff* that they receive regularly in the IE *RadioResourceConfigCommon* broadcasted by eNb it in every SIB2.

  RadioResourceConfigCommon in SIB2

```
...
BCCH-Config ::= SEQUENCE {
modificationPeriodCoeff ENUMERATED {n2, n4, n8, n16}
}
...
```

  Change of system information requires two "SI modification-allowed" radio frames. They are separated by a BCCH modification period which is the major delay factor in the solution (further details on this in Section 4.2.6).

- **Step 3.** The eNb performs paging to the UEs in RRC_IDLE and RRC_CONNECTED states with the *systemInfoModification* field set to TRUE. The eNb employs the standard paging procedure and message formats as they are defined in the 3GPP specifications [4].

  Paging message

```
...
Paging ::= SEQUENCE {
     pagingRecordList          OMIT
     systemInfoModification    TRUE
     etws-Indication           OMIT
     nonCriticalExtension      OMIT
}
...
```

  This paging message does not contain any paging records. It is sent in all paging occasions of the first "SI modification-allowed" radio frame, thus making its reception possible by all UEs regardless of their SFN allocation.

  It is important that the paging message with *systemInfoModification* is sent immediately in the appropriate radio frame. It must not queue in the PCCH buffer. This approach is not compliant to the current 3GPP specifications.

  Thus the paging notifies the UEs of a system information change that they shall expect upon expiration of the next BCCH modification period. The PCCH configuration remains unchanged at this point. UEs do not have the information on the new PCCH configuration.

- **Step 4.** The base station schedules the actual SI modification and awaits the next "SI modification-allowed" radio frame.

- **Step 5.** The eNb broadcasts SIB2 to the UEs on BCCH → DL-SCH → PD-SCH. It contains the IE *radioResourceConfigCommon* with the new PCCH configuration according to the 3GPP defined format [4]:

RadioResourceConfigCommon in SIB2

```
...
PCCH-Config ::= SEQUENCE {
     defaultPagingCycle     ENUMERATED {rf32, rf64, rf128, rf256},
     nB                     ENUMERATED {fourT, twoT, oneT, halfT,
                            quarterT, oneEighthT,oneSixteenthT,
                            oneThirtySecondT}
}
...
```

The parameter *defaultPagingCycle* remains unchanged. The parameter *nB* is increased by one step.

Upon reception of SIB2 the UE recalculates its SFN allocation in the DRX cycle and is ready to monitor a new paging occasion according to its new SFN.

- **Step 6.** Actual reconfiguration of the PCCH channels. The base station configures additional paging occasions in the paging frames that are: (A) already available (DRX cycle with $nB \geq T$), or (B) configured at the same time (DRX cycle with $nB < T$). This is a function of the eNb software that has to be developed entirely. Although the 3GPP specifications provide all instruments that are necessary for this operation, it is not defined in the 3GPP specifications, and thus it cannot be considered 3GPP compliant in full.

- **Step 7.** The base station recalculates the following thresholds:

(A) PCCH_RECONFIG_LOAD_THRE

(B) PCCH_RECONFIG_QUEUE_THRE

The new threshold values are effective immediately. The parameters *LIMIT_PAGING_PC* and *LIMIT_PCCH_QUEUE* remain unchanged as they were initially set by OMC.

- **Step 8.** The eNb increments the *systemInfoValueTag* by 1 and broadcasts SIB1 with the new *systemInfoValueTag* value. This operation in defined by 3GPP [4] entirely. The channel mapping for this transmission is BCCH → DL-SCH → PD-SCH. *systemInfoValueTag* is primarily addressed the UEs that return from out of coverage. Changed *systemInfoValueTag* informs them immediately of the fact that their stored system information is no longer valid and they must update it from the coming system information messages.

Upon execution of the PCCH reconfiguration algorithm the base station returns to the normal operation as it continues to monitor the paging load and the queue in the PCCH buffer.

4.2.5. Quantification of the effect on the paging capacity

The solution proposes an increase of the parameter *nB* by one step when either paging load or the queue in the PCCH buffer reach the critical thresholds. It doubles the paging capacity on the radio interface.

Table 2. The paging capacity on the radio interface in relation to the parameter nB.

#	DRX cycle ms	Frame ms	**Coeff. nB**	**nB**	Max Paging Records	Num. of S-TMSI in DRX cycle	Paging capacity
1	1280	10	**0.03125**	**40**	7	28	78750
2	1280	10	**0.0625**	**80**	7	56	157500
3	1280	10	**0.125**	**160**	7	112	315000
4	1280	10	**0.25**	**320**	7	224	630000

The calculations of the paging capacity in Table 2 are based on the 3GPP specifications [4]. The following equation defines the paging capacity on the radio interface:

$$PAGING\_CAPACITY = maxPagingRecords \cdot \frac{nB}{Frame} \cdot \frac{1000}{DRXcycle} \cdot 3600$$

Doubling the paging capacity in a single step may appear to have an exaggerated effect. However, this is an adequate response to abnormal growth of the signalling traffic due to the following factors:

- When the paging load rises above the installed paging capacity of an eNb, this situation poses a threat of paging overload to other base stations in the tracking area. Hence the solution offers a PCCH reconfiguration that guarantees a significant safety margin and helps to cope with still increasing paging load.

- Reconfiguration of the PCCH channels is a lengthy operation (see estimates in Section 4.2.6), and thus frequent reconfigurations are undesirable. Moreover, every change of system information increases the probability of the following undesired events: (A) blocking of service requests, and (B) abnormal termination of already established connections.

- Complexity of the solution is lower with a single tuneable parameter. It simplifies the implementation of the solution and improves its reliability.

4.2.6. Speed of PCCH reconfiguration

3GPP specifies a concept of BCCH modification period for any change of system information [4]. This is a necessary and reliable mechanism, but it delays the PCCH reconfiguration inevitably. This section provides an estimate of the delay.

The solution employs a mechanism that modifies the system information. As per the 3GPP LTE design this action is only possible in certain radio frames. They are defined by the BCCH modification period – parameter *modificationPeriodCoeff*. It is a configurable parameter, and it can take one of the values defined by 3GPP in [4]. Table 3 shows the influence of *modificationPeriodCoeff* on the separation of "modification-allowed" radio frames.

Table 3. Separation of "modification-allowed" radio frames with the BCCH modification period.

DRX cycle ms	modificationPeriodCoeff	Interval between "modification-allowed" radio frames, seconds
320	2	0.64
320	4	1.28
320	8	2.56
320	16	5.12
640	2	1.28
640	4	2.56
640	8	5.12
640	16	10.24
1280	2	2.56
1280	4	5.12
1280	8	10.24
1280	16	20.48
2560	2	5.12
2560	4	10.24
2560	8	20.48
2560	16	40.96

A maximum of two BCCH modification periods have to pass before the solution actually reconfigures the paging capacity. Thus for the most commonly used DRX cycle of 1280 ms the total PCCH reconfiguration time can vary from a little over 5.12 to almost 10.24

seconds – in the most favorable scenario. The exact time depends on the moment when the solution triggers the PCCH reconfiguration process.

It is worth stressing that the system information that the UE previously acquired remains valid. Thus the paging channels are fully operational in their "old" configuration until the new system configuration arrives in SIB2.

4.3. Validation of the solution

Validation of the solution requires a model of the paging channels with a possibility to monitor their performance in real-time. Hence this research employs the simulator that is described in details in Section 3.2 and characterised in Section 3.3.

Table 4. Configuration of the system, paging channels and traffic in the simulations.

Parameter	Value
System bandwidth	5 MHz
Frame structure	Type 1
Number of frames in the DRX cycle	128
BCCH modification period	2
nB coefficient (in proportion to DRX cycle)	1/16
Maximum number of S-TMSI paging records in the RRC paging request	7
Timer T3413	5000 ms
Maximum number of retransmission attempts	1
Primary paging type	S-TMSI
Secondary paging type	IMSI
Size of the paging buffer at eNb (initial value in paging records S-TMSI or IMSI)	140
BHCA, growing gradually	110000 – 210000
LIMIT_PAGING_PC	100%
LIMIT_PCCH_QUEUE	80%

The operational parameters of the system that are not listed in Table 4 remain unchanged as they are defined in the Description of the Simulator, Section 3.3.

4.3.1. Reference simulation with the solution disabled

The user traffic that grows from 120000 BCHA causes an undesired situation of paging overload when the load approaches the installed PCCH capacity of 157500 primary paging requests. This point corresponds to approximately 1500 seconds of operation on the time scale on Figure 15. An LTE system as of its current design cannot recover from this event. This situation is characterised by blocking of S1AP paging messages as the PCCH buffer does not have the capacity to accommodate them (ref. Figure 15, upper graph).

System KPIs degrade due to paging overload as it was explained earlier. Thus the real-time monitoring shows the dynamics of a highly important parameter – PCCH queuing time (ref. Figure 15, lower graph). This parameter contributes to the overall connection establishment time; therefore its degradation to over 4000 ms takes a significant negative effect on the quality of service. Further degradation of PCCH queuing time is not possible in the modelled configuration as it is limited by the capacity of the PCCH buffer in order to match the timer T3413.

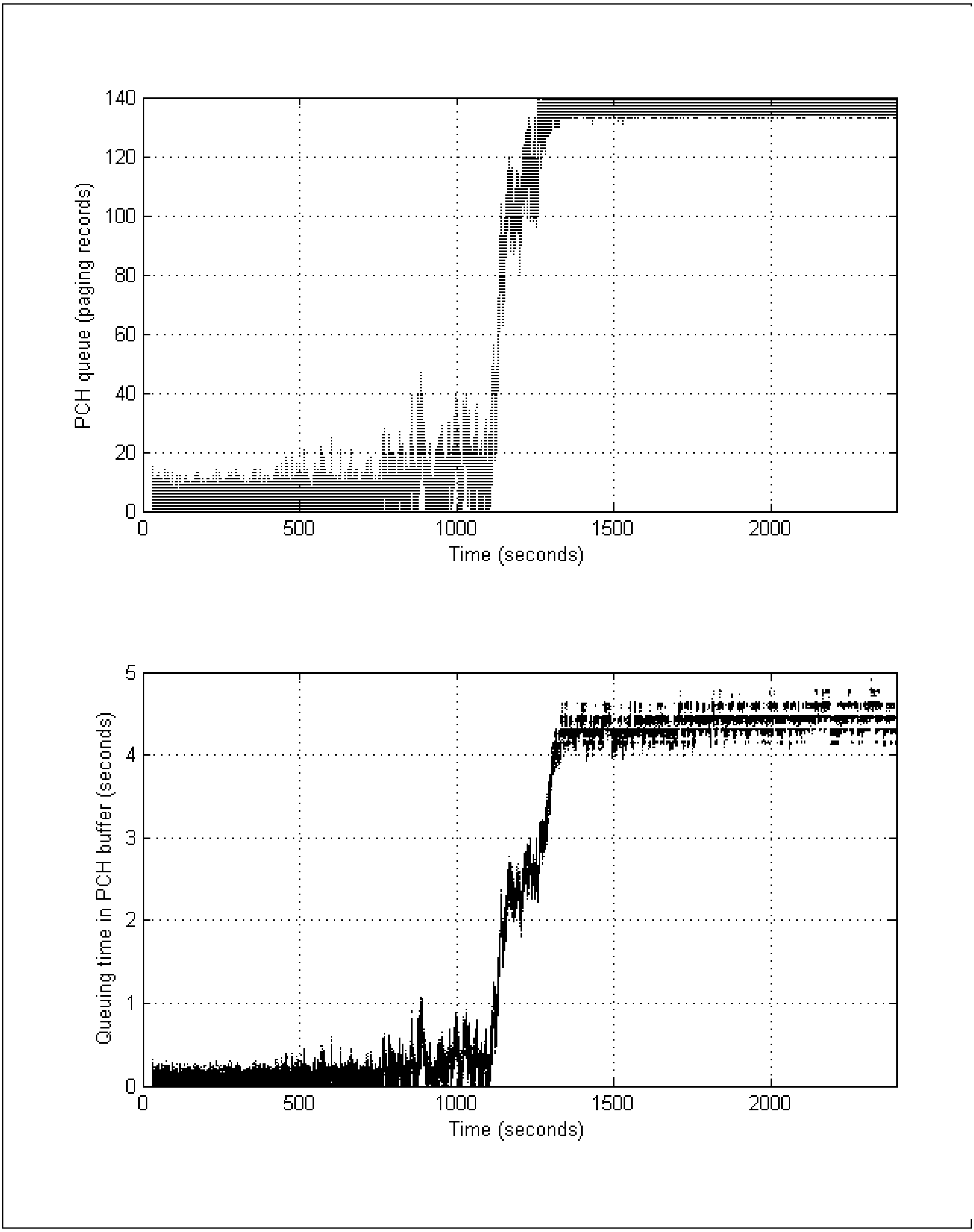

Figure 15. Solution DISABLED: real-time monitoring of the PCCH buffer. Paging overload and extended queuing time.

4.3.2. Simulation with PCCH reconfiguration

Simulations show that the system avoids paging overload when a PCCH reconfiguration is triggered at the PCCH buffer load threshold. However, during the reconfiguration the buffer accumulates a significant number of paging records. It takes then several DRX cycles until the PCCH buffer unloads.

Figure 16 and Figure 17 characterize the performance of the solution in real time.

It is important that the queuing time does not approach the limit of 5000 ms set by the T3413 timer in the model (Figure 16). Paging messages expire if they reach this limit, thus it is not acceptable.

Figure 17 zooms into the simulation interval when the actual PCCH reconfiguration took place. The reference algorithm for this operation is explained in Section 4.2. The figure displays the following steps:

- Marker "A". The algorithm triggers a PCCH reconfiguration when the growing queue in the PCCH buffer reaches the threshold *PCCH_RECONFIG_QUEUE_THRE*. The first step is immediate re-dimensioning of the PCCH buffer. The system doubles the capacity of the buffer to match the upcoming doubling of the capacity of the paging channels.
- Interval "A" to "B". The system awaits the first radio frame that allows SI change notification. New S1AP paging messages continue to arrive and are accommodated in the buffer. Paging records from the buffer are being transmitted over the air interface on the scheduled paging occasions.
- Marker "B". The base station starts transmission of SI change notification in every paging occasion of the radio frame. While all new S1AP paging messages are placed to the buffer, no paging records are transmitted over the air interface since the paging occasions are occupied by the SI change notification paging. This is the reason why the queue in the buffer starts growing rapidly at this moment.
- Marker "C". When the radio frame that carried the SI change notification ends, the paging channels return to their normal operation with the capacity still unchanged.
- Interval "C" to "D". The system waits a BCCH modification period. The paging channels remain fully operational but with the capacity still unchanged.

- Marker “D”. The base station broadcasts SIB2 that carries the new system information with new “PCCH-Config”. It uses a specific radio frame that is scheduled by the BCCH modification period. UEs decode the new SI, recalculate their new SFN and start to monitor their newly allocated paging occasion. New PCCH configuration is thus enforced.

Figure 17 shows that in less than 4000 ms following the reconfiguration the system unloads the PCCH buffer by transmitting all accumulated paging records.

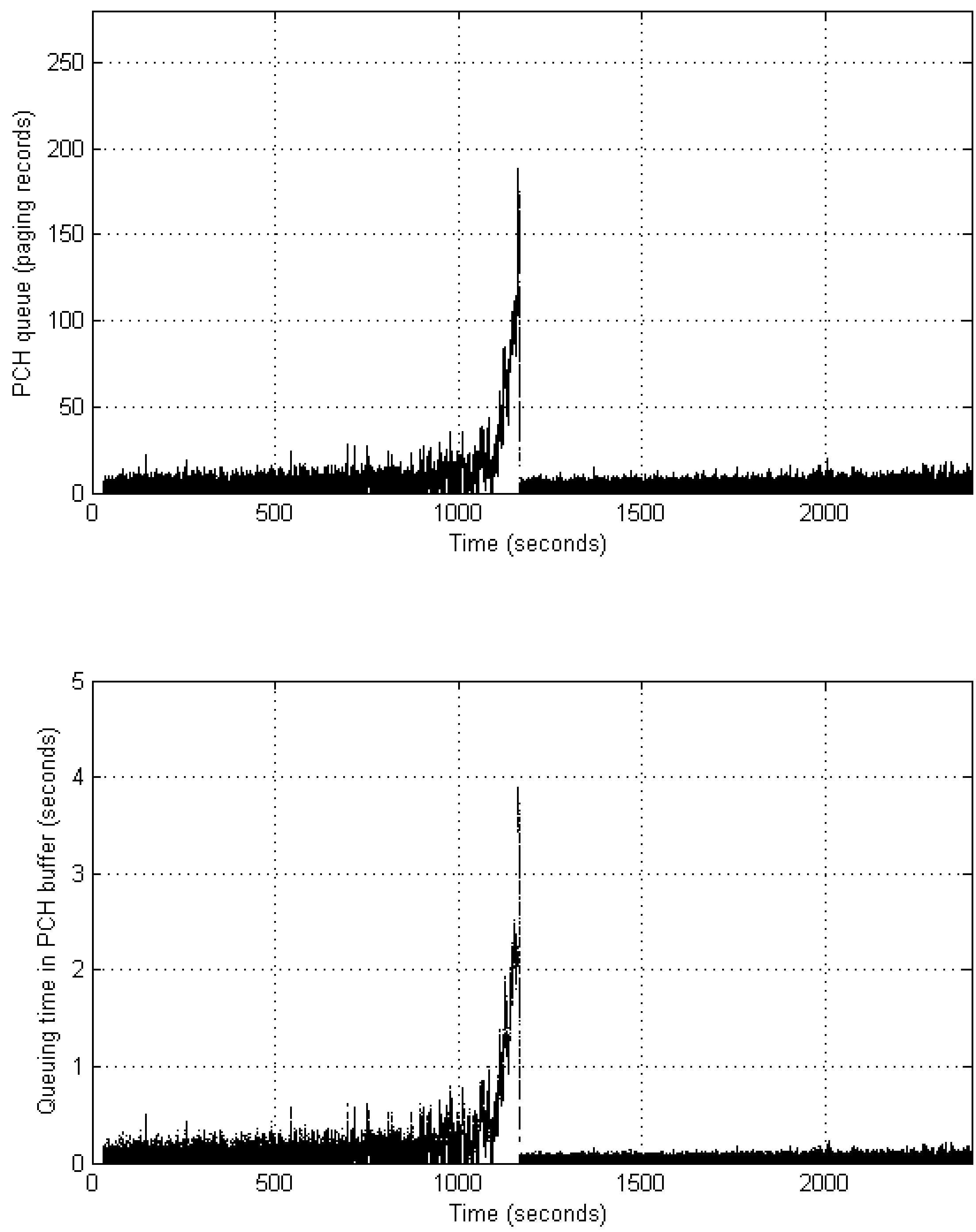

Figure 16. Solution ENABLED: real-time monitoring of the PCCH buffer. The system avoids paging overload while keeping the queuing time within the T3413 limit.

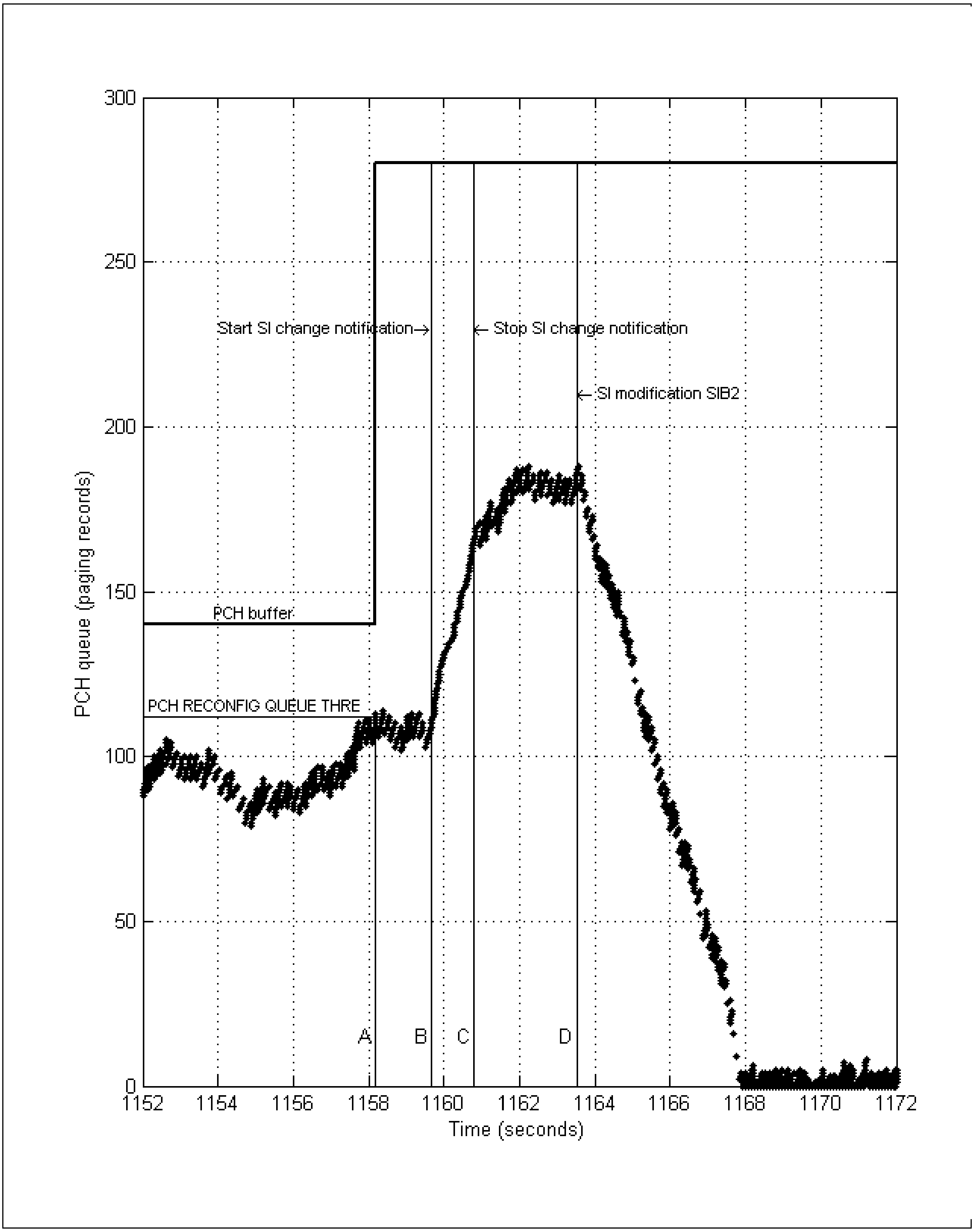

Figure 17. Solution ENABLED: real-time monitoring of the PCCH buffer. The effect on operation of the PCCH channels during the re-configuration.

The interval "B" to "C" on Figure 17 corresponds to a single DRX cycle and is characterized by a rapid aggregation of paging records in the PCCH buffer. This interval refers to the broadcast of the SI change notification since according to the 3GPP specifications SI change notification is carried by paging messages [4]. Thus other paging messages cannot be transmitted at the same time. Aggregation of a long queue in the PCCH buffer is an undesirable side effect. However it is not possible to avoid it. The only way to avoid it would be not using paging for transmission of SI change notification. Paging however is only available method to reach the UEs that are in RRC_IDLE state. Thus this side effect remains part of the solution.

Despite causing a limited queue aggregation effect, the solution prevents development of a paging overload situation without blocking. The queue aggregation is the major side effect; further analysis of the side effects follows in Section 4.5.

Since there is no PCCH blocking, as Figure 16 suggests, such KPIs as paging success rate and blocking probability report a perfect performance, and displaying them on graphs is of no practical value. There is one KPI, however, that demonstrates a difference in performance – average served paging load, Figure 18.

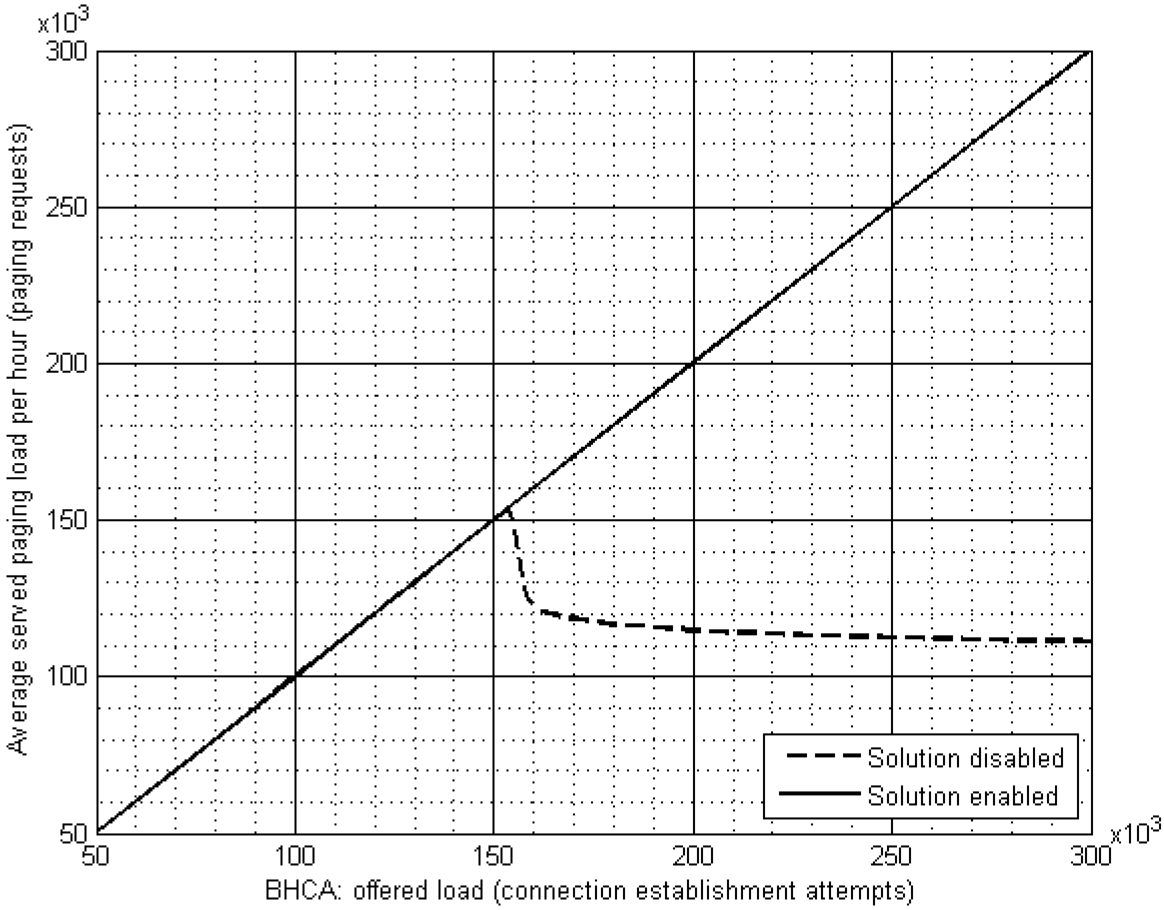

Figure 18. The solution allows the cells to accept surging load.

On Figure 18 the performance of the optimised paging procedure is confronted with the standard paging procedure. The solution allows the cells to accept the surging load while this load is blocked in non-optimised cells at the paging level. The non-optimised curve is from the experiments reported previously on Figure 8 in the evalution of the operation of the paging channels.

When a PCCH reconfiguration is triggered by the critical load threshold, the system shows an identical behaviour. Results of a simulation configured to prioritise the paging load threshold are reported on Figure 19.

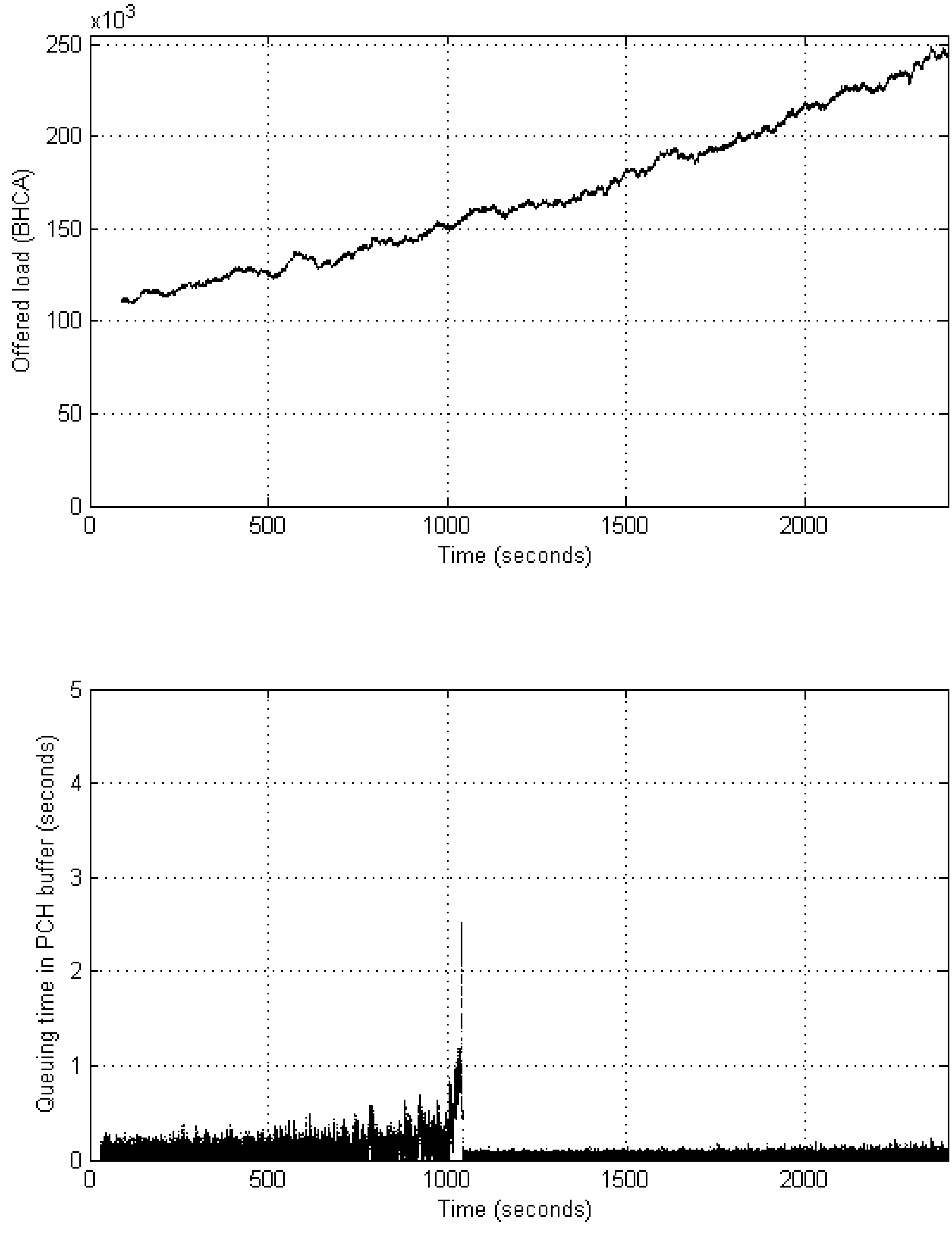

Figure 19. Solution ENABLED: real-time monitoring of the paging load and occupancy of the PCCH buffer.

The paging load threshold triggers a PCCH reconfiguration when the paging load hits 100% of the computed value 153200 BHCA. This moment corresponds to approximately 1040 seconds on the time axis and is conveniently monitored in the PCCH buffer.

Identically to the reconfiguration by the PCCH buffer threshold, the load-triggered reconfiguration causes a short-term increase in the buffer occupancy and the queuing time. At eh same time both parameters remain within their operational bands and do not cause blocking of paging requests.

4.4. Compliance with the 3GPP specifications

The solution incorporates signalling and operational procedures that rely on the 3GPP specifications. Sections 4.2.3 and 4.2.4 provide full technical details on the procedures. Not all procedures are fully compliant with 3GPP. Some procedures are outside of the scope of the 3GPP specifications which leave them to vendors to decide on their implementation. This section explains the relation between the procedures applied in the solution and the 3GPP specifications.

4.4.1. Operational procedures non dependant on 3GPP

The following operational procedures require modification of the software and are not part of the 3GPP specifications:

- Initial configuration and in-operation adjustment of the following parameters: *LIMIT_PAGING_PC, LIMIT_PCCH_QUEUE, PCCH_RECONFIG_LOAD_THRE, PCCH_RECONFIG_QUEUE_THRE*
- Monitoring of the paging load
- Monitoring of the queue in the PCCH buffer
- Re-dimensioning of the PCCH buffer

The procedures listed above depend on the implementation of particular equipment and the choices of its manufacturer. They do not need to be defined in any specifications and thus can be considered 3GPP compliant.

4.4.2. Signalling procedures fully compliant with 3GPP

The solution employs the following signalling procedures that are fully compliant to the 3GPP specifications:

- Scheduling of SI change in accordance with the concept of BCCH modification period
- Paging with notification of BCCH modification

- Broadcast of SIB1 and SIB2 without interruption of their standard transmission scheduling

The above mentioned procedures are described by 3GPP in [4] and are applied in this research unchanged in full compliance with their specification.

4.4.3. Signalling procedures non compliant with 3GPP

The solution defines the following procedures that are not compliant to the 3GPP specifications:

- Prioritization of the paging message with a notification of BCCH modification. This paging message must be transmitted on the air interface without being placed in the PCCH buffer. The solution cannot tolerate any delay in transmission of this paging as the BCCH modification must be completed in two particular radio frames, not later.

- Reconfiguration of the PCCH channels in operation. This operation employs the standard SI modification procedure defined in [4]. However 3GPP does not define this procedure to be applied in operation. Thus despite having the standard procedure provided by 3GPP, the reconfiguration of PCCH channels as it is performed in the solution cannot be considered fully compliant to 3GPP specifications.

4.5. Side effects and hardware requirements

In a situation when the capacity of paging channels is not sufficient, it is reasonable to reduce the capacity of the traffic channels in favour of the paging channels. If the subframes that are required to configure additional paging occasions are occupied with user traffic, the MME must free them up. Hence the user traffic must be dropped from these subframes. The exact way the user traffic is dropped is outside the scope of this research since the equipment and the OMC from all vendors have this functionality already. The decision on which users are dropped can be based on the pre-emption group policy or on any other algorithm that the vendors find appropriate.

As it was described earlier, the major side effect of the solution is the queue aggregation in the PCCH buffer during the reconfiguration. This effect however is limited to a single DRX cycle. Its possible negative consequences are eliminated by prior extension of the PCCH buffer.

The side effects of the solution are the following:

- Reduction of the traffic channels capacity in favour of the PCCH capacity
- Queue aggregation during one DRX cycle when eNb broadcasts SI change notification
- Increased complexity of the eNb software
- Additional load on the eNb CPU

Equipment manufacturers have to assess the additional CPU requirements in their base stations individually. Extra load on CPU results from deployment of two monitoring procedures: one for the paging load and another for the PCCH buffer. Although they will represent a small fraction in the overall CPU load, it must be taken into consideration. This is the major side effect of the solution since the operational efficiency of eNb and its scalability are highly sensitive to excess CPU load.

5. Conclusions

This research shows that excessive MME paging load can be processed on the cell level without degradation in KPIs and proposes an appropriate solution.

Moreover, the value of the proposed solution is not limited solely to the paging procedure. Prevention of paging overload is the starting point to any “in-operation” optimisation of other signalling channels defined in LTE. The reason for this is the logic of enforcement of system information modifications that necessarily must start with paging in order to notify the UEs which are in RRC_IDLE state. Thus reliable paging channels open numerous possibilities for further optimisation of the signalling subsystem in LTE.

The research delivers the following results:

- Mathematical model of MME paging
- Simulator of logical paging control channels (PCCH)
- Analysis of operation of logical paging control channels
- Solution to paging overload that employs “hot” PCCH re-configuration
- Validation of the solution, quantification of the effect, analysis of side effects
- 3GPP compliance analysis of the solution

All listed results are applicable exclusively to the MME paging procedure of LTE.

The solution completely prevents blocking in the paging channels by configuring additional PCCH capacity in “hot” mode without partial or complete interruption in service.

The research presents a stochastic model, based on retrial queuing systems theory, which describes the operation of the MME paging mechanism. The model predicts the paging failure probability with the required precision and catches the non-zero paging failure threshold. The solution to paging overload employs this model in order to set one of the two PCCH reconfiguration thresholds.

The solution to paging overload is partially 3GPP compliant. It includes four operational procedures that are not dependant on 3GPP specifications, three fully 3GPP compliant procedures and two procedures non compliant with 3GPP. The latter include prioritisation of BCCH modification paging messages and the change of PCCH configuration without taking eNb out of service.

Implementation of the two non-compliant operational procedures used in the solution requires a formulation in the 3GPP RRC protocol specification [4] and a modification to the eNb software.

No additional hardware is required to implement the solution. Modifications to already deployed eNb hardware are not necessary, provided that the CPU can accept extra load resulting from the solution. An evaluation of the increase in the CPU load is necessary and must be performed for every particular modification of eNb.

The solution offers a marketing potential to equipment suppliers. It can be realised as a purchasable software feature.

Returning to the aspects of applicability of the solution, it is of key importance that the solution stabilises the paging channels in high load conditions, thus opening opportunities for “in-operation” optimisation of other control channels and signalling procedures.

6. References

[1] 3GPP, “Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3” Specification 3GPP TS 24.301 version 12.4.0 Release 11. [Online]. Available: http://www.3gpp.org

[2] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode” Specification 3GPP TS 36.304 version 12.0.0 Release 12. [Online]. Available: http://www.3gpp.org

[3] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2” Specification 3GPP TS 36.300 version 12.1.0 Release 12. [Online]. Available: http://www.3gpp.org

[4] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification” Specification 3GPP TS 36.331 version 12.1.0 Release 12. [Online]. Available: http://www.3gpp.org

[5] V. G. Kulkarni and H. M. Liang, “Retrial queues revisited,” in Frontiers in queueing. Boca Raton, FL, USA: CRC Press, Inc., 1997, pp. 19–34.

[6] A. Baraev, U. Ayesta, I. M. Verloop, D. Miorandi and I. Chlamtac, “Technical vulnerability of the E-UTRAN paging mechanism”, IEEE WCNC, Paris 2012.

[7] 3GPP, “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP)” Specification 3GPP TS 36.413 version 12.1.0 Release 12. [Online]. Available: http://www.3gpp.org

[8] 3GPP, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access” Specification 3GPP TS 23.401 version 12.4.0 Release 12. [Online]. Available: http://www.3gpp.org

[9] 3GPP, “Paging load reduction in synchronized cells,” Technical document 3GPP R3-070965 of TSG-RAN-WG3 Meeting 56. [Online]. Available: http://www.3gpp.org

[10] 3GPP, “Paging load reduction in asynchronous cells,” Technical document 3GPP R3-081412 of TSG-RAN-WG3 Meeting 60. [Online]. Available: http://www.3gpp.org

[11] M. B. L. Lopes, J. Harris, “Scheme for paging optimization,” Technical disclosure IPCOM000 192 990D, 2010, by Motorola Inc. [Online]. Available: http://ip.com/IPCOM/000192990

[12] 3GPP, “Solutions for paging optimization,” Technical document 3GPP R3-092204 of TSG-RAN-WG3 Meeting 66-bis. [Online]. Available: http://www.3gpp.org

[13] J. R. Artalejo and A. Gomez-Corral, Retrial Queueing Systems: A Computational Approach. Springer, 2008.

[14] G. I. Falin and J. G. C. Templeton, Retrial Queues. CRC Press, 1997.

[15] 3GPP, “Paging optimization,” Technical document 3GPP R3-091698 of TSG-RAN-WG3 Meeting 65. [Online]. Available: http://www.3gpp.org

[16] 3GPP, “Numbering, addressing and identification”, Specification 3GPP TS 23.003 version 12.2.0 Release 12. [Online]. Available: http://www.3gpp.org

[17] J. M. Lopez, I. Shaikh, “LTE Smart Paging List”, US patent US2012/0115515A1, published May 2012.

[18] Alcatel-Lucent S.A., “9471 Wireless Mobility Manager”, description of commercially available equipment, available online http://www.alcatel-lucent.com/products/9471-wireless-mobility-manager, last visited on 03.04.2014.

[19] Yu Yin, Weihua Hu, Qing Zhou, “Method and apparatus for allocating paging areas”, US patent US2014/0024407A1, published January 2014.

[20] D. Nowoswiat, G. Milliken, “Managing LTE Core Network Signalling Traffic”, Alcatel-Lucent technical document, July 2013

[21] Alcatel-Lucent S.A., “Alcatel-Lucent 9471 Wireless Mobility Manager”, Solution sheet, May 2012.

[22] 3GPP, “Architecture aspects of Home NodeB and Home eNodeB” Specification 3GPP TS 23.830 version 9.0.0 Release 9. [Online]. Available: http://www.3gpp.org

[23] Alcatel-Lucent S.A., “The Impact of Small Cells on MME Signalling: methods to reduce and optimize MME core signalling caused by small cells”, Application note, October 2013.

[24] W. Ma, Y. Fang, “A new location management strategy based on user mobility pattern for wireless networks”, Proceedings of 27th Annual IEEE Conference on Local Computer Networks, pages 451-457, 2002.

[25] W. Ma, Y. Fang, P. Lin, “Mobility Management Strategy Based on User Mobility Patterns in Wireless Networks”, IEEE Transactions on Vehicular Technology, Volume 56, Issue 1, pages 322-330, 2007.

[26] Bar-Noy Amotz, et al., "Paging mobile users in cellular networks: Optimality versus complexity and simplicity", Theoretical Computer Science, Vol. 470, pages 23-35, 2013.

[27] Hyung-Woo Kang, Hyon-Goo Kang, Seok-Joo Koh, “Optimization of TAC Configuration in Mobile Communication Systems: A Tabu Search Approach”, Proceedings of 16th International Conference on Advanced Communication Technology, pages 5-9, 2014.