(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 August 2014 (28.08.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/128239 A1

(51) International Patent Classification:
H04W 28/02 (2009.01)

(21) International Application Number:
PCT/EP2014/053388

(22) International Filing Date:
21 February 2014 (21.02.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13000940.0 25 February 2013 (25.02.2013) EP

(71) Applicant: **NOKIA SOLUTIONS AND NETWORKS GMBH & CO. KG** [DE/DE]; St. Martinstrasse 76, 81541 Munich (DE).

(72) Inventors: **HAUTH, Stephan**; Maienweg 71, 89081 Ulm (DE). **WALLMEIER, Eugen**; Bussardstraße 14, 82223 Eichenau (DE). **THEIMER, Thomas**; Forstenrieder Weg 1g, 82065 Baierbrunn (DE). **WEERAWARDANE, Thushara**; Kotelawala Defence University, Ratmalana (LK). **ZAKI, Yasir**; New York University Abu Dhabi (NYU AD), Abu Dhabi, P.O.Box 129188 (AE).

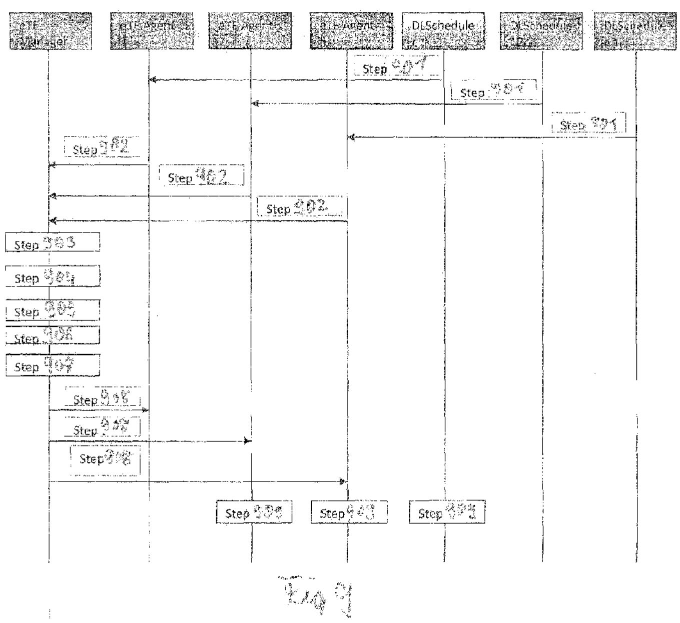
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) Title: METHOD, MANAGING ENTITY, AGENT ENTITY FOR CONSISTENT BANDWIDTH ALLOCATION



(57) Abstract: A method for a consistent allocation of bandwidth shares for non-guaranteed bit rate bearers on air interfaces and in a mobile backhaul network section of a packet based mobile service network is provided, wherein the method comprises: estimating an overall transport bandwidth available for non- guaranteed bit rate bearers in the mobile backhaul network; estimating a fair share of the overall transport bandwidth for each non-guaranteed bit rate bearer and its respective traffic rate; and in case of a potential congestion throttling the respective traffic rate of a non-GBR bearer to its respective fair share transport bandwidth.

WO 2014/128239 A1

METHOD, MANAGING ENTITY, AGENT ENTITY FOR CONSISTENT BANDWIDTH ALLOCATION

BACKGROUND OF THE INVENTION

Field of the Invention

The invention relates to the field of mobile service networks. More specifically, the invention relates to a method, a managing entity, an agent entity, a mobile service network, and computer program product for consistent bandwidth allocation, in particular downlink bandwidth allocation, and in particular in case of air interface and mobile backhaul congestion in a packet oriented mobile service network.

As an example some different kinds of mobile service networks are depicted schematically in Figs. 1 and 2. As an example, in a mobile service network 100 at least one radio transceiver (Radio Transceiver) 101, connected via an access and aggregation network (Mobile Backhaul Network; MBH) 102 to at least one gateway (Gateway) 103, provides through a radio interface connectivity between a plurality of (in most cases mobile) subscriber devices (user equipment, UE) 104 among each other and with other (mobile) subscriber devices, servers, or other components or devices (not shown), reachable from the at least one gateway via the Public Switched Telephone Network (PSTN), through the Internet 105, or using other, potentially dedicated, wired or wireless (fixed or mobile) service networks. Mobile Core Network functions (not shown) may be incorporated in or associated with the mobile service network to support mobility management, to implement operator policies, and to perform service control.

User equipment (UE) comprises any fixed or mobile devices, systems, or arrangements in the hands, or at a site, or under control of a subscriber (or user) of the mobile service network and capable of connecting to the network via the radio interface provided by a radio transceiver.

The radio transceiver (RT) may be a base station (BTS), a NodeB, an enhanced NodeB (eNodeB) or any equivalent device providing regional (and preferably cellular) radio access using technologies as specified e.g. in the 2G, 3G, 4G/LTE, or other relevant radio standards.

The gateway may be any device, system, or arrangement capable of providing access to other service networks such as, or through, the PSTN, the Internet, or any other kind of application and/or transport service network. For example the gateway may be a PDN gateway, SAE gateway or any other suitable gateway providing an interface, e.g. a packet data network interface, to a network, e.g. the Internet. Typical applications, among others, could be location based or streaming services. The gateway may to a large extent be implemented in, or comprise, computer program software, which, when loaded into the memory and executed on a computer, causes the computer to implement respective gateway functions. Consequently, a gateway device, system, or arrangement may comprise computer hardware and software and it may be capable to, or actually do, provide and/or share hardware and software resources with other system functions not necessarily specific for the gateway function.

Radio transceivers and gateways may be arranged in redundancy schemes for a better availability and reliability of their respective services.

The mobile backhaul network (MBH) may comprise any components and technologies suitable to interconnect radio transceivers and gateways as described above. More recent systems preferably use packet-oriented data transfer and related MBHs preferably use packet based transport with protocols and formats as e.g. specified by Ethernet or Internet Protocol (IP) related standards. Reuse of existing infrastructures and use of off-the-shelf routers and switches enables cost efficient solutions. Physical transmission may comprise any kinds of technologies including microwave radio, optical and electrical systems. Packet Data Network (PDN) Gateways interface to the Internet or dedicated packet oriented service networks using a packet data network interface (PDNI) (Fig. 2).

Related Art

3GPP TS 23.002 V12.1.0 (2012-12) (as well as other versions of the TS 23.002 document) presents possible architectures of a packet oriented mobile service network based on various radio access technologies and an Evolved Packet System (EPS) as specified by 3GPP. A short summary of a related architecture is provided by F. Firmin in “The Evolved Packet Core” (retrieved on Jan.31, 2013 at http://www.3gpp.org/The-Evolved-Packet-Core). As these documents emphasize on the mobile service architecture, they do not show the MBH transport.

A variety of different services is provided to subscribers and users of such kind of mobile service networks via a user equipment as described above. Services may comprise voice, video and data in various combinations of unidirectional and bidirectional, realtime or non-realtime, interactive, messaging, streaming type, or any other modes of communication. Accordingly, a variety of different service and Quality of Service (QoS) requirements have to be respected for conveying respective traffic flows across the packet oriented mobile service network.

Such requirements are reflected in the concept of bearers as specified e.g. in the standardization document 3GPP TS 23.401. The most recent version of this document with respect to the instant application is 3GPP TS 23.401 V11.4.0 (2012-12), issued on Dec. 18, 2012. A bearer uniquely identifies traffic flows that receive a common treatment between a user equipment and a gateway. Packet filters are associated with the bearers to identify the traffic flows belonging to each bearer. All traffic mapped to the same bearer receives the same bearer level packet forwarding treatment, i.e. routing, queuing, scheduling, rate shaping, etc., in the network and thus exhibits the same QoS behaviour. Actions performed on bearer traffic by individual components of the network may differ according to the different roles of the components in the network (user equipment, radio transceiver, MBH, gateway), but the rules applied to individual packets within a component will always be the same for traffic belonging to the same bearer.

3GPP distinguishes between guaranteed bit rate (GBR) bearers and non-guaranteed bit rate (non-GBR) bearers. A QoS Class Identifier (QCI) is associated with each bearer as a reference to access node-specific parameters that control bearer level packet forwarding treatment (e.g. scheduling weights, admission thresholds, queue management thresholds, etc.) in the mobile service network nodes. Dedicated transmission resources are allocated and blocked for the transfer of GBR traffic. An Aggregate Maximum Bit Rate (AMBR) is assigned to each access point to the network and shared between all related non-GBR bearers. In uplink (or upstream) direction this access point name AMBR (APN-AMBR) is enforced by the UE and the PDN gateway. In downlink (or downstream) direction it is enforced by the PDN gateway. In a similar way a further AMBR value is applied across all non-GBR bearers of a user equipment (UE-AMBR) and enforced by the radio transceiver.

QCI, GBR and AMBR values enable a mandatory and fair allocation of resources assigned to the different bearers on the (by its nature) shared radio interface. Whereas dedicated resources are blocked for exclusive use by admitted GBR traffic, remaining resources are shared according to respective subscription levels (e.g. premium or economy) between the non-GBR bearers currently active on the interface. A fair allocation of resources, reflecting respective subscription levels, is the target, even in times of congestion. When a premium type bearer has three times more weight than an economy type bearer, then this ratio is targeted when assigning resources to respective competing bearers independently of the number of bearers currently being served per type. Bearer based traffic shaping and related fair traffic shares can also be applied by the radio transceiver to upstream traffic propagated towards the PDN gateway via the MBH. Note, that the number of subscription levels is not necessarily limited to two.

Whereas traffic can be treated individually on a per bearer basis in the radio transceiver and the gateway, this is completely different within the packet based MBH. Class based traffic management is applied instead of bearer based traffic control. Traffic classes are distinguished e.g. by DiffServ Code Points (DSCP) in IP based networks (IETF RFC 2474, RFC 2475, RFC 3260 and others), P-bit values with Carrier Ethernet (IEEE 802.1Q), or EXP bit values with MPLS (RFC 3270), but no bearer individual information can be used.

3GPP TS 23.401 suggests a potential mapping between QCI values of EPS bearers and DSCP values. Ekström ("QoS Control in the 3GPP Evolved Packet System", IEEE Communications Magazine, February 2009) explains that the gateway and the LTE RAN implements a QCI to DSCP mapping function to make a translation from bearer level QoS (QCI) to transport-level QoS (DSCP), and he concludes that in the transport network the bearer is not visible and hence the traffic forwarding treatment of each individual packet is based on the DSCP value.

For Mobile Operators it becomes increasingly important to offer differentiated QoS to their customers. For instance, an Operator may want to distinguish between "business users" and "economy users" or offer "gold, silver and bronze services". The QoS which the end user of an LTE network perceives in case of network congestion is determined by various QoS mechanisms applied in different parts of the network.

Congestion can occur at the air interface and also in the Mobile Backhaul Network (MBH). Due to the high peak rates possible in LTE networks, it is not economical for an operator to take maximum cell capacities into account when dimensioning the MBH.

The Downlink Packet Scheduler of an eNodeB handles congestion at the air interface by supporting multiple QoS classes identified by the Quality Class Identifier (QCI). The Scheduler takes into account the QCI each time, when it allocates resources to an individual Radio Bearer. Business users and economy users, and gold, silver and bronze services respectively, each have to be distinguished and individually mapped to different QCIs and respective bearers.

Mobile Backhaul Networks use simpler QoS mechanisms than the air interface schedulers of an eNodeB. In IP based Mobile Backhaul Networks typically class based traffic management (based on the DiffServ concept) is used to handle congestion. Individual Radio Bearers are not visible in the MBH.

In LTE networks dedicated resources are reserved for Guaranteed Bit Rate (GBR) bearers, e.g. using strict priority queues and Admission Control (AC) procedures. This raises the issue of how to distribute the remaining bandwidth among non-GBR bearers.

BRIEF SUMMARY OF THE INVENTION

It may thus be an object to provide a method, managing entity, agent entity and computer program product which may enable a consistent allocation of bandwidth shares, e.g. downlink bandwidth shares, for non-GBR bearers on air interfaces and in the MBH section of a packet based mobile service network.

The object is achieved by a method, managing entity, agent entity and computer program product as specified in the independent claims. Further embodiments are described in the dependent claims.

According to an exemplary aspect a method for a consistent allocation of bandwidth shares for non-guaranteed bit rate bearers on air interfaces and in a mobile backhaul network section of a packet based mobile service network is provided, wherein the method comprises: estimating an overall transport bandwidth available for non-guaranteed bit rate bearers in the mobile backhaul network; estimating a fair share of the overall transport bandwidth for each non-guaranteed bit rate bearer and its respective traffic rate; and in case of a potential congestion throttling the respective traffic rate of a non-GBR bearer to its respective fair share transport bandwidth.

According to an exemplary aspect a method for a consistent allocation of downlink bandwidth shares for non-guaranteed bit rate bearers on an air interfaces and in a mobile backhaul network section of a packet based mobile service network is provided, wherein the method comprises estimating an overall transport bandwidth available for non-guaranteed bit rate bearers in the mobile backhaul network; estimating a fair share of each non-guaranteed bit rate bearer and its respective bandwidth; and in case of a potential congestion throttling a respective traffic rate of a non-GBR bearer to its respective fair share bandwidth; wherein the throttling is done by discarding incoming packets to a radio transceiver, thereby invoking respective TCP traffic control mechanisms.

In particular, the method may relate to a consistent allocation of downlink bandwidth shares and/or uplink bandwidth shares. In particular, the estimated fair share of each non-guaranteed bit rate bearer relates to a fair share of bandwidth and/or fair share of traffic rate.

It should be noted that the term “estimating” should be read in a broad sense. In particular, it may encompass a true estimation, i.e. calculation or even estimating, of an unknown value, for example the overall transport bandwidth available. In this case the estimated or calculated value may include an error. However, the term may as well encompass a determination of an in principle known value, e.g. in cases where the overall transport bandwidth is already known or can be easily calculated, in which case the resulting “estimated value” may include only a very small or even no error at all.

According to an exemplary aspect a managing entity for a mobile service network is provided, wherein the managing entity comprises: a calculation unit configured to estimate an overall transport bandwidth available for non-guaranteed bit rate bearers and configured to estimate a fair share of the overall transport bandwidth for each non-guaranteed bit rate bearer and its respective traffic rate; and a communication interface configured to communicate information indicative of the estimated fair share.

In particular, the managing entity may be a traffic enforcement manager or an enhanced traffic enforcement manager and/or may be part or may form a gateway. Alternatively, the managing entity may be part of a radio transceiver or may form a radio transceiver, for example. The term “managing entity” may particularly encompass a hardware component like a suitably programmed processing unit or computer as well as the pure software component which may be run on a processing unit. In particular, the overall transport bandwidth available for non-guaranteed bit rate bearers may be calculated for all bearers of the mobile service network or for a subset of all bearers of the mobile service network, e.g. by a subset corresponding to a cell and/or which are associated with a specific eNodeB, NodeB or base station.

According to an exemplary aspect an agent entity for a mobile service network is provided, wherein the agent entity comprises a counting unit configured to count incoming data packets per non-guaranteed bearer; and a communication interface configured to communicate with a managing entity according to an exemplary aspect.

In particular, the agent entity may be a traffic enforcement agent or an enhanced traffic enforcement agent. For example, the agent entity may be part of or may form an eNodeB, a NodeB, a base station or a radio transceiver. Thus, the agent entity may be a radio transceiver which may be adapted to perform a method according to an exemplary aspect. The agent entity may be further configured to increment a virtual queue length for at least one or each non-guaranteed bit rate bearer, in particular the virtual queue length may be incremented according to bytes received by the agent entity in a time interval (t-x,t). In addition or alternatively the agent entity may be configured to send the counted number of incoming data packets or bytes arrived in the time interval (t-x,t) and/or a fair share of an air interface bandwidth for the point in time (t) to the managing entity. In particular, the agent entity may be configured to perform a throttling by discarding packets or data packets and/or by using Explicit Congestion Notification.

According to an exemplary aspect a packet based mobile service network comprising a managing entity according to an exemplary aspect and an agent entity according to an exemplary aspect is provided.

It should be noted that the managing entity and the agent entity may be two different entities, physical and/or logically/virtually. However, the two entities may be implemented by one and the same physical entity implementing both functions.

According to an exemplary aspect a computer program product comprising software is provided, which when loaded into the memory of a computer enables the computer to execute any of the steps of the method of an exemplary aspect.

By providing a method and entities which are adapted or configured to perform an estimation or calculation of a fair share for non-guaranteed bit rate bearer and are further adapted or configured in case of a potential congestion to throttle the respective traffic rate of the non-GBR bearer to its respective fair share transport bandwidth, it may be possible to enable a consistent allocation of bandwidth shares, e.g. downlink bandwidth shares, for non-GBR bearers on air interfaces and in the MBH section of a packet based mobile service network.

In the following exemplary embodiments of the methods are provided. However, the described components and features may also be used in connection with the managing entity, the agent entity, the packet based mobile service network, and the computer program product.

According to an exemplary embodiment of the method the throttling is done by discarding incoming packets and/or by using Explicit Congestion Notification.

In particular, the throttling may be done by discarding incoming packets to a radio transceiver, agent entity, NodeB, base station or eNodeB. The discarding of a number of incoming packets may be a suitable way to achieve that a calculated or estimated fair share of transport bandwidth or corresponding traffic load is observed. In addition or alternatively the chances of a potential congestion may be reduced or even avoided by using Explicit Congestion Notification which may lead to the fact that one or more bearer may reduce their data rate.

According to an exemplary embodiment of the method respective TCP traffic control mechanisms are invoked for the throttling.

According to an exemplary embodiment of the method the estimating of the fair share of the overall transport bandwidth a weighting factor is considered for each non-guaranteed bit rate bearer.

In particular, the weighting factor or weight may be associated with a Quality of Service Class Identifier (QCI). The use of dedicated or selected weighting factors may be a suitable way to ensure that the different non-GBR bearers get a fair share of the overall transport bandwidth.

According to an exemplary embodiment the method further comprises determining whether an actual traffic rate of respective non-guaranteed bit rate bearer is below its respective fair share of transport bandwidth; and in case it is determined that the actual traffic rate of the respective non-guaranteed bit rate bearer is below its respective fair share of transport bandwidth redistributing the difference between the estimated fair share traffic rate and the actual traffic rate to other bearers.

In particular, the other bearer(s) may be non-guaranteed bit rate bearers. The provision of a redistribution possibility may enable an efficient use of the overall transport bandwidth. The redistribution may be performed once or several times, e.g. in form of an iterative process, till all potential free or unused transport bandwidth is allocated to the bearers, in particular to the other active bearers.

According to an exemplary embodiment of the method the estimation of the fair share of the overall transport is performed for a given time interval.

In particular, the estimation of the fair share of the overall transport bandwidth may be based on a time interval preceding the given time interval. For example, the estimation may be based on a single preceding time interval or a plurality of preceding time intervals. In particular, the estimation may be based on an average of the plurality of preceding time intervals or may be based on the directly or immediately preceding time interval.

According to an exemplary embodiment the method further comprises deciding whether a non-guaranteed bit rate bearer is an active or inactive bearer.

In particular, an inactive non-GBR bearer may be a bearer which is not active, i.e. has a traffic load or carried load of less than a predetermined threshold. For example, the predetermined threshold may be 10 kbit/s or 2 kbit/s or the like.

According to an exemplary embodiment of the method the estimation of the fair share of transport bandwidth is only performed for bearer determined to be an active bearer.

By estimation the fair share only for active bearer it may be possible to reduce a necessary processing or calculation power.

In the following exemplary embodiments of the managing entity are provided. However, the described components and features may also be used in connection with the methods, the agent entity, the packet based mobile service network, and the computer program product.

According to an exemplary embodiment of the managing entity the calculation unit is further configured to decide for each non-guaranteed bit rate bearer whether it is an active bearer or not.

In particular, the decision may be based on information received by the managing entity from an agent unit, for example. For example, the information may be based on or may be indicative for a carried load of the respective bearer during a given time interval.

According to an exemplary embodiment of the managing entity the estimation of a fair share is only performed for active bearers.

According to an exemplary embodiment of the managing entity the managing unit is configured to perform a bandwidth redistribution.

In particular, the bandwidth redistribution may be performed for the active bearer, i.e. only the active bearers may be considered by the redistribution, while the bandwidth share of the inactive non-GBR bearers are redistributed or allocated to the active ones. Also in case active bearers are below their fair share, the difference between their actual traffic and their fair share may be redistributed as well.

According to an exemplary embodiment of the managing entity the calculation unit is further configured to calculate an enforcement rate for at least one active bearer.

The invention may be guided by the idea that the bandwidth available for (active) non-GBR bearers at any bottleneck in the system should always be shared based on the same rules as used by the air interface schedulers, independent of the location of the respective bandwidth bottleneck. A key principle may be a preventive one. Proactive throttling of bearer traffic based on prediction of potential transport bottlenecks avoids the potentially devastating effects of the purely class based traffic control mechanisms in the transport.

In the DETAILED DESCRIPTION below an example of an “enhanced Traffic Enforcement (eTE)” algorithm running in the eNodeB is described, which triggers TCP congestion control in order to control the rates of non-GBR bearers inside the MBH. In case of transport congestion eTE ensures that the bottleneck bandwidth is distributed among the Radio Bearers based on the weights $w_{k,l,j}$ (or their equivalents in case of different air interface conditions for different bearers) defined for the air interface schedulers. This means, the same Quality of Service (QoS) policy is applied independently of the root cause for congestion (transport congestion or air interface congestion). It should be noted that the eTE algorithm may perform best, in case all non-GBR bearer in one MBH has associated the same DSCP value. Thus, it may be advantageous that for all non-GBR bearers the same DSCP value may be assigned.

Exemplary embodiments builds and relies on the specific properties and the behavior of traffic that on one hand is greedy on bandwidth (and capable of getting hold of available bandwidth), but on the other hand obeys to rules of congestion control that aim at a fairly shared usage of the available bandwidth between multiple traffic sources. This typically applies to TCP based traffic, which forms the vast majority of non-GBR traffic in the system.

The system comprises or consists of an eNodeB (eNB) connected to a transport network for mobile backhaul (MBH). The transport network may use any kinds of switching and transmission technologies. It may form a bottleneck for traffic coming from various sources and destined for users connected to the eNodeB through the air interface. It is a characteristic of the exemplary embodiments that the DOWNLINK schedulerin the eNodeB (indirectly) controls the downlink BW of the transport section bottleneck located UPSTREAM to it, i.e. the bandwidth of the data streams coming down to it. In fact, it does not really manage the bandwidth of the respective downlink transport system, but it influences the transmission rates of the downlink data streams by playing with their intrinsic congestion control mechanisms in order to prevent any bearer from exceeding its instantaneously calculated fair share. The invention specifies the mechanisms to control the system and a specific algorithm for determining and allocating the respective resources for the different bearers.

It should be noted, that major parts of the algorithm may be performed outside of the radio transceiver (or eNodeB), e.g. in any other control instance of the system, which is equipped with a computer. Such instance could e.g. be a network management system or a policy controller.

It should be noted further, that most (or even all) steps of the method or parts of the related algorithm may be implemented in software, i.e. as a computer executable program code, which when loaded into the instruction memory of a computer, enables the computer to execute the respective method steps or parts of the algorithm. As such the software may be incorporated with any means capable of storing permanently or temporarily computer program code or related data.

It should be noted that in the same way at least some of the steps and parts may as well be implemented in hardware, e.g. in electronic circuitry and/or logic devices of any kind. Such hardware may especially comprise equipment for packet classification, queuing and scheduling and their respective control.

Consequently and obviously, any system and device capable of or intended to be used for executing the method, or the underlying algorithm, or any part of any of these, is preferably equipped, has to be equipped or is at least with respective means, i.e. a computer (processing device with respective memory and input/output capabilities, etc.) and/or other respective hardware means.

A person skilled in the art understands that the principles of the invention as disclosed herein and illustrated based on the example of a 3GPP based system architecture are as well applicable to variants of this, or other architectures of mobile and/or fixed service networks employing the same or a similar type of bearer based traffic control in the access area, and will easily be able to apply these principles accordingly.

The aspects and exemplary embodiments defined above and further aspects of the invention are apparent from the examples of embodiment to be described hereinafter and are explained with reference to these examples of embodiment.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Fig. 1 schematically shows a mobile service network.

Fig. 2 schematically shows another kind of mobile service network.

Fig. 3 schematically shows a mobile service network indicating possible bottlenecks.

Fig. 4 showing schematically effective buffer filing vs. discard probability.

Fig. 5 schematically depicts the monitoring of downlink load.

Fig. 6 schematically depicts basic principle of eTE for an example of two greedy sources.

Figs. 7A and 7B schematically illustrate the redistribution during which the eTE takes into account sources with low activity.

Fig. 8 schematically illustrates a functional view of eTE.

Fig. 9 schematically depicts a whole eTE algorithm in a sequence view.

Fig. 10 schematically depicts the eTE algorithm of Fig. 9 in a time view.

DETAILED DESCRIPTION OF THE INVENTION

The illustrations in the drawings are schematic. In the following a detailed description of exemplary embodiments is given. In the beginning in the context of Fig. 1 some general remarks concerning a method or mechanism operating a network in particular a virtualized network are given.

In the following some general information concerning general traffic management is given which may be helpful for understanding the invention.

In LTE networks dedicated resources are reserved for Guaranteed Bit Rate (GBR) bearers, e.g. using strict priority queues and Admission Control (AC) procedures. This raises the issue of how to distribute the remaining bandwidth among non-GBR bearers. In case of congestion the bottleneck bandwidth should be distributed in such a way that the bandwidth share of a non-GBR bearer does not depend on the root-cause of congestion (air interface overload or MBH overload). To give an example: If a Mobile Operator decides that a Business User should get three times the bandwidth of an Economy User and configures the schedulers for the air interface accordingly, it should not happen that a Business User is constrained to only get the same or even less bandwidth than the Economy User, when the MBH is overloaded. Aligning the bandwidth allocation and the QoS delivered by the air interface and in the MBH is not easy, due to the principal differences between the related mechanisms applied at the air interface and in the MBH. The air interface scheduler acts on individual Radio Bearers while the MBH uses class based traffic management and is not aware of Radio Bearers.

For example, considering an eNodeB serving a number of p cells typical values of p could e.g. be 3 or 6. Each of the cells has its own Downlink Packet Scheduler, which allocates bandwidth to non-GBR bearers using weights (one weight per QCI) and further parameters which reflect radio conditions.

For a cell k, $1 \le k \le p$, a UE i in cell k, $1 \le i \le m_k$, and a bearer j of UE i, $1 \le j \le n_{k,i}$ let $w_{k,i,j}$ denote the weight associated with the QCI of bearer j. If all active non-GBR bearers in cell k face the same radio conditions, then the downlink scheduler of cell k considers

$$\frac{w_{k,i',j'}}{\sum_i \sum_{j, j\,active} w_{k,i,j}} BW_{air}^k \qquad (1)$$

as the fair bandwidth share of bearer j' of UE i'. Here BW_{air}^k denotes the bandwidth available for non-GBR connections in cell k. When the different UEs in a cell face different radio conditions (which is a more realistic case), the individual bandwidth shares actually allocated to active non-GBR bearers by the downlink air interface schedulers (i.e. the (true) fair shares of the air interface bandwidth) can deviate significantly from the (ideal) fair bandwidth shares defined above.

To ensure a service differentiation according to the user and service categories agreed upon with subscribers connected through the air interface (e.g. business and economy users and/or gold, silver, and bronze services), traffic management and control mechanism applied to other potential bottlenecks in the system should not interfere with the bandwidth distribution and the respective shares allocated by the air interface schedulers. Especially, the distribution of transport bandwidth in case of a congestion in the MBH should not contradict to or jeopardize the distribution of bandwidth to individual bearers as specified for the air interface.

This, however, cannot be achieved by applying the standard QoS mechanisms available in the MBH. Commercial mobile backhaul services typically offer only a very limited number of service classes (e.g. between two and four, as e.g. defined by the Metro Ethernet Forum in respective implementation agreements such as MEF 22.1). Throughput guarantees can be given for these service classes as a whole (e.g. using weighted fair queuing), but not for individual non-GBR bearers.

The bandwidth allocated to each service class is fixed, which can lead to further problems, when e.g. the traffic mix of different types of users (e.g. Business and Economy) cannot be predicted exactly. Even worse, since the service usage on the air interface frequently changes in time, the weights defining the bandwidth shares of Weighted Fair Queuing (WFQ) schedulers in the MBH will usually not fit with the actual traffic mix. It thus may happen that a Business User, though he should clearly be preferred against an Economy User (and actually receives this preferred service at the air interface), gets less bandwidth in the Mobile Backhaul

Network than the Economy User, because e.g. the traffic mix contains more business users than predicted.

In many cases, only a single service class is available for non-GBR bearers. Hence, all non-GBR bearers receive equal treatment in the Mobile Backhaul Network, even when congestion occurs. However, this does not imply that they are also receiving the same throughput, because the bandwidth allocation is now determined by TCP mechanisms. All TCP sessions, which use the same traffic class in the MBH, will lose packets at the bottleneck in the MBH with a similar (or even the same) probability. Therefore, in times of congestion, the Mobile Backhaul Network has the tendency to equalize the throughput of TCP sessions using the same QoS class. This is implied by the way, the fairness mechanisms of TCP are defined and implemented. As a consequence, the throughput of non-GBR bearers, assigned to the same QoS class in the MBH, is proportional to the (arbitrary) number of TCP sessions contained therein. In other words, in times of MBH congestion two non-GBR bearers, even if associated with the same bandwidth at the air-interface, may come out of the MBH with completely different bandwidths, if they contain a different number of TCP sessions.

In uplink direction the radio transceiver (e.g. an eNodeB of an LTE system) can easily do an individual shaping of all bearers sharing its resources. By doing so it can control and limit the amount and the mix of traffic according to the resources available for it in the MBH before the traffic enters the MBH.

In downlink direction the bearers sharing the air interface resources of a radio transceiver may pass through different gateways. Even AMBR shaping in the different gateways (as it has no ”common view“) cannot avoid potential traffic congestion in the MBH with consequences as described above. The result is a completely different behavior of the network depending on the location of a potential traffic congestion (air interface or MBH) and a respectively inconsistent service experience for the users.

In the following embodiments and details of embodiments of the present invention will be described.

As a first section basics of an enhanced Traffic Enforcement Algorithm (eTE) will be described which can be used in a method of the present invention.

Fig. 3 schematically depicts a basic concept of eTE. In particular, Fig. 3 shows a mobile service network 300 similar to the one depicted in Figs. 1 and 2 comprising at least one eNodeB 301, connected via a Mobile Backhaul Network (MBH) 302 to at least one gateway (Gateway, e.g. an SAE-gateway) 303, provides through a radio interface connectivity between a plurality of (in most cases mobile) subscriber devices (user equipment, UE) 304 among each other and with other devices, e.g. servers 306, or other components or devices (not shown), reachable from the least one gateway via the Internet 305, for example.

Furthermore, some potential transport bottlenecks for LTE downlink (DL) traffic are indicated in Fig. 3 by the dotted lines 307. As exemplary shown in Fig. 3 any part of the MBH (i.e. any link and/or node) could be congested at a given point in time, i.e. forms a bottleneck. It should be noted that different DL traffic streams destined for the same UE may pass different bottlenecks. Furthermore, DL per bearer shaping in the SAE-gateway may limit the traffic which can be sent over a bearer.

In addition in the upper portion of Fig. 3 some information and details concerning an eTE enforcement of “fair bandwidth share” for radio bearers are given. In particular, missing acknowledge messages at the TCP destination (e.g. UE) due to eTE packet discard trigger TCP congestion control as indicated by arrow 310 leading from the UE to a TCP server. The TCP congestion control is schematically indicated by the arrows 311.

In particular, the bearer knowledge available at an eNB is leveraged to manage bearer bandwidth allocation for the eNB in case of backhaul congestion (applicable to LTE and LTE-A). The basic concept of eTE is illustrated in the upper portion of Fig. 3. eTE requires a few basic building blocks which are described in more detail below:

1) A mechanism to estimate the overall transport bandwidth available for non-GBR bearers and the fair share of each non-GBR bearer (taking into account the QoS parameters attached to the QCI).
2) A mechanism for redistributing bandwidth if some non-GBR bearers do not need their fair bandwidth shares completely. They may send low traffic volumes due to air interface congestion or other reasons.
3) A mechanism to control and enforce the transport bandwidth allocated to each individual non-GBR bearer. Throttling a bearer may be done by discarding incoming IP packets at the PDCP layer and thus triggering TCP congestion control.

The eTE algorithm is relying on TCP congestion control. In mobile networks more than 90% of the traffic volume of non-

GBR bearers is using TCP (compare: J. Erman, A. Gerber, K.K. Ramakrishnan, S. Sen, O. Spatscheck. Over The Top Video: the Gorilla in Cellular Networks. IMC'11, November 2-4, 2011, Berlin, Germany).

The eTE algorithms uses the steps described in the following subsections. These steps are performed once per "cycle", this means once per interval of x ms (where x could e.g. have a typical value of 20 or 50, but may be selected from a larger range, e.g. between 1 and 1,000).

In the following the enhanced Traffic Enforcement Algorithm (eTE) (above described in general) will be described in detail in five sections. However, in the beginning some basics of an eTE algorithm will be given in form of a listing, which is then elaborated in greater detail afterwards in the five sections:

Functions of eTE Manager

(performed once per cycle, e.g. once per 20 ms)

1) Receive for each bearer from eTE Agents:

- traffic volume arrived in last cycle
- Estimate of BW available at air interface (optional)

2) Estimate total DL TRS BW available for all bearers in next cycle (based on traffic volumes measured in recent cycles)

3) Calculate for each bearer a fair share of DL TRS BW, applying methods which are consistent with BW allocation at the air interface
(Inactive bearers may be excluded from assigning a fair BW share)

4) Estimate for each bearer the load in the next cycle (using the measurements for the bearer)

5) Identify „non-greedy“ bearers not fully exploiting their fair share of DL TRS BW

- comparing load estimate and fair share of DL TRS BW
- potentially taking into account known bottlenecks e.g. at air interface

6) Calculate for each non-greedy bearer how much of the fair share of DL TRS BW is not used (left BW of a bearer)

7) Calculate the „Total Left BW“ (sum over all bearers).

(Total Left BW can be redistributed to „greedy bearers“)

8) Calculate for each bearer an „Enforced Rate“ taking into account

- its fair share of DL TRS BW
- „Total Left BW“, which can be redistributed

Send for each bearer its „Enforced Rate“ to eTE Agent

Functions of eTE Agents

(one agent per radio cell operating in the user plane)

Actions performed once per cycle:

- Send for each bearer to eTE Manager:
- traffic volume arrived in last cycle
- Estimate of BW available at air interface in next cycle (optional)
- Receive for each bearer from eTE Manager: „Enforced Rate“

Actions done on packet arrival:

- Update counter for traffic volume
- If „Enforced Rate“ is exceeded and action required:

Ask Traffic Sources using the bearer to throttle their rates (e.g. triggering TCP congestion control by discarding a packet or usage of Explicit Congestion Notification)

Section 1 Estimating Available Transport Bandwidth:

Fig. 5 schematically depicts the monitoring of downlink load arriving at eNodeB by the eTE for different cycles and different bearers. In particular, the DL load arriving at eNodeB (number of bytes) is continuously measured for each individual non-GBR bearer which is indicated for three bearers in Fig. 5 by the lines 501. The total DL load arriving at eNB, i.e. the sum of all non-GBR bearers, is continuously measured and indicated by line 503. In addition the DL load expected during time interval (t,t+x) estimated on the basis of measurements up to t is indicated for the three bearers by lines 502, while the estimation of the total DL load arriving at eNB for interval (t,t+x) including a safety margin is indicated by line 504.

Per cycle the eTE algorithm measures the instantaneous traffic load comprising all non-GBR bearers (in Mbit/s, counting the bits which arrive in the interval (t,t+x)

$$TotalCarriedLoad_{inst}(t) = \frac{\sum_{cycle(t,t+x)} Pktsizes}{x}. \qquad (2)$$

The measurements of the instantaneous traffic load in recent cycles are used at time t in order to estimate the transport bandwidth BW_{TRS}(t) available for non-GBR traffic in the next cycle (t,t+x).

If it is noted at the end of the interval (t-x,t), that the estimate $BW_{trs}(\mathrm{t-x})$, which was done at the beginning of the interval, does not exceed the measured load (i.e. $TotalCarriedLoad_{inst}(t-x)$) at least by a predefined margin (which allows for traffic growth), then the estimate will be increased at least by the margin. Otherwise, this means the estimate $BW_{trs}(t-x)$ overestimated the $TotalCarriedLoad_{inst}(t-x)$ by more than the predefined margin, the estimate $BW_{trs}(t)$ will be reduced using an exponential filter formula. The mathematical representation of the above description is as follows:

$$Margin(t-x) = \max(1\ Mbit/s\ , BW_{trs}(t-x) * 0.1)$$

$$if\ \left[BW_{trs}(t-x) \leq TotalCarriedLoad_{inst}(t-x) + \text{Margin}(t-x)\right]$$

Then: $L(t-x) = max\,(TotalCarriedLoad_{Inst}(t-x), BW_{trs}(t-x))$ (3)

$$BW_{trs}(t) = \max(L(t-x) + 1\ Mbit/s,\ 1{,}1 * L(t-x))$$

Else: $BW_{trs}(t) = (1 - 1/T_{total})BW_{trs}(t-x) + 1/T_{total} * TotalCarriedLoad_{inst}(t-x)$,

where T_{total} has the default value 10.

Section 2 Active User Detection:

eTE distributes $BW_{trs}(t)$ among all active non-GBR bearers. In order to decide which non-GBR bearers are active during the interval (t,t+x), eTE estimates for each non-GBR bearer which traffic load it will generate in the next cycle (t,t+x). This estimate is denoted as $CarriedLoad_{est}^{k,i,j}(t)$. Let $CarriedLoad_{inst}^{k,i,j}(t)$ denote the measured instantaneous load of bearer j in cycle (t,t+x), which is determined by counting the number of bits arriving in cycle (t,t+x)

$$CarriedLoad_{inst}^{k,i,j}(t) = \frac{\sum_{cycle\ (t,t+x)} Pktsizes\ of\ bearer\ j}{x}. \quad (4)$$

The measured instantaneous load in cycle (t-x, t) is used to estimate the carried load in cycle (t,t+x):

$$CarriedLoad_{est}^{k,i,j}(t) = \left(1 - \frac{1}{T_{bearer}}\right) CarriedLoad_{est}^{k,i,j}(t-x) + \left(\frac{1}{T_{bearer}}\right) CarriedLoad_{inst}^{k,i,j}(t-x). \quad (5)$$

where T_{bearer} has the default value 5.

Non-GBR bearer j is considered as active during cycle (t,t+x), if

$$CarriedLoad_{est}^{k,i,j}(t) > Active\_minimum\_rate$$

where *Active_minimum_rate* is a small rate, e.g. 2 kbit/sec.

Section 3 Calculating Fair Shares of Transport Bandwidth:

Consider a cell k, $1 \le k \le 3$, a UE i in cell k, $1 \le i \le m_k$, and a bearer j of UE i, $1 \le j \le n_{k,i}$.

Let $w_{k,i,j}$ denote the weight associated with the QCI of bearer j and define

$$W_{k,i,j}(t) = \begin{cases} w_{k,i,j} & if\ bearer\ j\ is\ active\ at\ time\ t \\ 0 & otherwise \end{cases}. \quad (6)$$

The fair share of the transport BW for a non-GBR bearer j’ of UE i’ in cell k’ for cycle (t,t+x) is estimated as

$$FS^{trs}_{k',i',j'}(t) = \frac{w_{k',i',j'}(t)}{\sum_k \sum_i \sum_j w_{k,i,j}(t)} \cdot BW_{trs}(t) \quad (7)$$

It should be noted that the fair share of transport BW is 0 for inactive non-GBR bearers.

It is crucial that BW_{trs}(t), which is calculated at time t provides a good estimate for the unknown bottleneck bandwidth in cycle (t,t+x). The eTE algorithm distributes the bandwidth BW_{trs}(t) among all active non-GBR bearers. If the totally available bandwidth in cycle (t,t+x) is underestimated, then eTE will throttle the traffic more than necessary. On the other hand, if the totally available bandwidth is overestimated, then bearers using too much bandwidth will not be throttled and the desired QoS differentiation will not be reached.

Problems can arise when the actual measured throughput is low, because in this case the total amount of bandwidth available for redistribution is very limited, and may be insufficient to allow even a single bearer to ramp up its throughput quickly. Therefore it may be necessary to define a minimum value for BW_{trs}(t), which is always distributed among the non-GBR bearers independently of the measured instantaneous load.

<u>Section 4 Redistributing Bandwidth and Calculating Enforced Rates:</u>

The eTE algorithm monitors non-GBR bearers and ensures that a non-GBR bearer j of a UE i in cell k does not exceed an

$EnforcedRate_{k,i,j}(t)$

during cycle (t,t+x).

The eTE algorithm allows that each non-GBR bearer exploits its fair share of the transport bandwidth $FS^{trs}_{k,i,j}(t)$. However, if a bearer is not using its fair BW share completely for t_{redist} milliseconds (default t_{redis}= 120 milliseconds, but other values may be envisaged, depending e.g. on the size of the "cycles"), eTE assumes that the bearer is unable to ramp up to its fair share and redistributes unused bandwidth to other bearers allowing them Enforced Rates which are higher than their fair BW shares.

There are various reasons why a bearer may not be able to fully exploit its fair bandwidth share. A bearer may simply not have enough data to send, its traffic may be shaped at the egress of the Serving Gateway, its traffic may pass a bottleneck somewhere between source and eNodeB (which is not seen by other bearers), or a bearer may suffer from air interface congestion. Bottlenecks at the air interface are known in the eNB. In contrast to other bottlenecks they can be taken into account explicitly.

Let $FS^{air}_{k,i,j}(t)$ denote the fair share of the air interface bandwidth for a bearer j of UE i in cell k in the cycle (t,t+x). The air interface scheduler of cell k provides this value to eTE once per cycle. In case the air interface BW is not fully used during cycle (t,t+x) the air interface scheduler defines $FS^{air}_{k,i,j}(t)$ as a very large number. It is described in Section 4 how the air interface schedulers calculate $FS^{air}_{k,i,j}(t)$. eTE uses $FS^{air}_{k,i,j}(t)$ as an upper limit for the bandwidth allocated to bearer j. In case $FS^{trs}_{k,i,j}(t) > FS^{air}_{k,i,j}(t)$, the difference $FS^{trs}_{k,i,j}(t) - FS^{air}_{k,i,j}(t)$ is available for redistribution to other bearers.

Bandwidth redistribution and calculation of Enforced Rates may be done with the following algorithm.

Define *List(t)* as the set of all non-GBR bearers

//List (t) contains the bearers for which the Enforced Rate is still to be defined.

Define $T_{redistr}$

// default value 120 milliseconds

$LeftBW(t) = 0$

//- ----------- Processing non-greedy users ---------

Loop 1: DO for all bearers j in List(t)

$$Proposed\_Bandwidth^{k,i,j}(t) - \min\left(\max_{u\in(t-Tredistr,t)} CarriedLoad^{k,i,j}_{est}(u)\ ,FS^{air}_{k,i,j}(t)\right)$$

//eTE allocates the proposed bandwidth to a bearer, unless this is not possible due to congestion. In the code an approximation of $\max_{u\in(t-Tredistr,t)} CarriedLoad^{k,i,j}_{est}(u)$, is used: The time axis is divided into slots of y ms (default y=40 ms). For each of the last n-1 slots (where n= Tredistr/y) the maximum carried load estimate is stored. The maximum taken over these n-1 values and the carried load estimates observed in the current slot so far is used as approximation for $\max_{u\in(t-Tredistr,t)} CarriedLoad^{k,i,j}_{est}(u)$.

IF 1 $Proposed\_Bandwidth^{k,i,j}(t) \leq FS^{trs}_{k,i,j}(t)$

THEN

$$EnforcedRate_{k,i,j}(t) = min\ (FS^{trs}_{k,i,j}(t), FS^{air}_{k,i,j}(t))$$

Remove bearer j from List(t)

$$LeftBW(t) = LeftBW(t) + FS^{trs}_{k,i,j}(t) - Proposed\_Bandwidth^{k,i,j}(t)$$

END IF 1

END Loop 1.

//- ----- Processing greedy users -----------------

Loop 2: DO for all bearers j in List (t)

$$BW^{trs}_{k,i,j}(t) = FS^{trs}_{k,i,j}(t)$$

//$BW^{trs}_{k,i,j}(t)$ is the BW preliminarily reserved for bearer j. It may be increased during later iterations, if bearer j profits from BW redistribution

END Loop 2

List_Changed = TRUE

Loop3 : DO WHILE (List_Changed = TRUE)

List_Changed = FALSE

Loop 4: DO for all bearers j in List(t)

$$BW^{trs}_{k',i',j'}(t) = BW^{trs}_{k',i',j'}(t) + \frac{W_{k',i',j'}(t)}{\sum_k \sum_i \sum_{j,j \epsilon List(t)} W_{k,i,j}(t)} LeftBW(t)$$

END Loop 4

$$LeftBW(t) = 0$$

Loop 5: DO for all bearers j in List(t)

IF 2 $Proposed\_Bandwidth^{k,i,j}(t) \leq BW^{trs}_{k,i,j}(t)$

// the proposed bandwidth is smaller or equal to the BW preliminarily reserved for bearer j

THEN

// Enforced Rate can be set for j

$$EnforcedRate_{k,i,j}(t) = Proposed\_Bandwidth^{k,i,j}(t)$$

// reduce BW allocated for j

$$LeftBW(t) = LeftBW(t) + (BW^{trs}_{k,i,j}(t) - Proposed\_Bandwidth^{k,i,j}(t)$$

END IF 2

Remove bearer j from List(t)

List_Changed = TRUE

END Loop 5

END Loop 3

Loop 6: Do for all bearers j in List(t)

// For the bearers left in List(t) the proposed *bandwidth* is greater than $BW^{trs}_{k,i,j}(t)$, however there is no additional left BW which could be distributed

. $EnforcedRate_{k',i',j'}(t) = BW^{trs}_{k,i,j}(t)$

// These bearers may suffer from eTE packet discard in (t,t+x).

END Loop 6

<u>Section 5 Throttling Sources Exceeding their Enforced Rates:</u>

The eTE algorithm maintains for each non-GBR bearer a virtual queue, which is used to calculate the packet discard probability to be applied in the interval (t,t+x), if a bearer violates its enforced rate.

When an IP packet of bearer j arrives, then the virtual queue is incremented by the length of the packet (counted in bits).

At the beginning of every cycle (this means at time t in the cycle (t,t+x)) the virtual queue of an active non-GBR bearer is served at its $EnforcedRate_{k,i,j}(t)$, the virtual queue of an inactive user is served at the *Active_minimum_rate*, which may be defined e.g. as 2 kbit/s.

A packet discard function is activated for an active non-GBR bearer j during the interval (t,t+x) if its virtual queue exceeds a configurable threshold after serving it at time t.

The discard probability for an incoming packet is defined with the aid of the Effective Buffer Filling (EBF), which is measured in seconds and calculated by dividing the current virtual queue length by the $EnforcedRate_{k,i,j}(t)$

$$EBF_{k,i,j}(t) = \frac{Virtual\_queue\_length_{k,i,j}(t)}{EnforcedRate_{k,i,j}(t)} \quad . \qquad (8)$$

The discard probability for an incoming packet of bearer j is 0 if the Effective Buffer Filling (EBF) is below a configurable threshold. For greater values the discard probability grows with increasing EBF as indicated in Fig. 4 showing schematically the EBF vs. the discard probability.

Either random discard using this discard probability or periodic discard (every m^{th} packet, with 1/m equal to discard probability) are applied to incoming IP packets.

When eTE discards an IP packet from a bearer, which has exceeded its Enforced Rate, the TCP layer at the source will react when noting the packet loss. It will reduce its sending rate (typically by 50%) and then start to increase the rate again (typically by one TCP segment size per round trip time). In this way bearers exceeding their Enforced Rates are throttled and the bandwidth becomes available for other bearers.

In the next section some details concerning the fair share of air interface bandwidth will be given

The Fair Share of Air Interface Bandwidth

BW limits resulting from air interface congestion are known in the eNodeB and can be taken into account explicitly.

Let $PRBs_k^{inst}(t)$ denote the number of Physical Resource Blocks (PRBs) used in TTI t for all kind of radio bearers in cell k. Then

$$PRBs_k^{avg}(t) = {}^{1}/_{T_{PRBs}} \cdot PRBs_k^{inst}(t) + \left(1 - {}^{1}/_{T_{PRBs}}\right) \cdot PRBs_k^{avg}(t - TTI) \qquad (9)$$

is the average number of PRBs used during recent TTIs. The filtering constant T_{PRBs} has the default value $T_{PRBs} = 10$. If

$$PRBs_k^{avg}(t) \Big/ PRBs_k^{avail} > 0{,}95, \qquad (10)$$

where $PRBs_k^{avail}$ denotes the number of PRBs available per TTI (e.g. 50 for 10 MHz), then cell k is considered as congested at time t and the calculation of the fair share of the air interface bandwidth is relevant.

A person skilled in the art can easily find various ways to determine a respective value for the fair share $FS^{air}_{k,i,j}(t)$ of the air interface bandwidth for a bearer j of UE i in cell k in the cycle (t,t+x). A simple way could be to use the ideal fair shares as specified by formula (1) above. Respective options are not described in detail.

In the following an exemplary embodiment will be described with reference to the figures.

In particular, Fig. 6 schematically depicts basic principle of eTE for an example of two greedy sources. Starting from the DL transport bandwidth (BW) for non-GBR bearers which is measured at eNB (601) the eTE calculates or estimates for each UE the fair BW share of measured transport BW (602). During the calculation air interface weights (QCI weights of the DL packet schedulers) are used which is schematically indicated by arrow 603. Furthermore, the eTE monitors whether radio bearers exceed their fair shares, which is schematically indicated in Fig. 6 by the column 604 showing the enforced rates for UE1 and UE2.

Fig. 7 schematically illustrates the redistribution during which he eTE takes into account sources with low activity. In particular, Fig. 7A starts from the same point as Fig. 6. Starting from the DL transport bandwidth (BW) for non-GBR bearers which is measured at eNB (701) the eTE calculates or estimates for each UE the fair BW share of measured transport BW (702). During the calculation air interface weights (QCI weights of the DL packet schedulers) are used which is schematically indicated by arrow 703. In the case of Fig. 7A UE1 has a carried load estimation which is lower than the fair share of UE1 while UE2 has a higher carried load estimation than its fair share as can be seen in Fig. 7A in column 704 wherein the estimations of the carried load are based on per bearer measurements as indicated by arrow 705. The next column 706 indicates the enforced rate of UE1 in case no redistribution is performed, i.e. equals the fair share of UE1, while column 707 represents the difference between the fair share or enforced rate of UE1, i.e. represents left BW which can be redistributed to UE2. The resulting enforced rate, i.e. the sum of the fair share of UE2 and the left BW of UE1 is depicted as the last column 708 in Fig. 7A. In general, Fig. 1A shows a portion of the redistribution in case UE1 is not using its fair share. In this case UE2 may use BW left over by UE1, resulting in an enforced rate for UE2 which is the sum of the fair share for UE2 and the BW left over by UE1. However, it should be noted that eTE does not prevent UE1 from ramping up to its fair share.

The ramping up process is schematically illustrated in Fig. 7B. A new cycle starts from the DL transport bandwidth (BW) for non-GBR bearers which is measured at eNB (711) the eTE calculates or estimates for each UE the fair BW share of measured transport BW (712). During the calculation air interface weights (QCI weights of the DL packet schedulers) are used which is schematically indicated by arrow 713. In the case of Fig. 7B UE1 has a carried load estimation which is lower than the fair share of UE1 while UE2 has a higher carried load estimation than its fair share as can be seen in Fig. 7B in column 715 wherein the estimations of the carried load are based on per bearer measurements as indicated by arrow 715. However, the carried load estimation for UE1 is higher than the one for a former cycle which is indicated by the delta 719 of carried load estimation in the column 715. The next column 716 indicates the enforced rate of UE1 in case no redistribution is performed, i.e. equals the fair share of UE1, while column 717 represents the difference between the fair share or enforced rate of UE1, i.e. represents left BW which can be redistributed to UE2. The resulting enforced rate, i.e. the sum of the fair share of UE2 and the left BW of UE1 is depicted as the last column 718 in Fig. 7B. However, the left over of UE1 and thus, the enforced rate for UE2 is lower in the case of Fig. 7B than in the one of Fig. 7A, i.e. UE1 is ramping up and the BW left over by UE1 decreases. Thus, the enforced rate of UE2 is reduced and packets for UE2 are discarded since carried load of UE2 exceeds its enforced rate. As a consequence TCP congestion control will reduce carried load for UE2.

Fig. 8 schematically illustrates a functional view of eTE. The eTE algorithm can be implemented using two functional components as show in Fig. 8: an eTE manager or managing entity 801 in the Transport SW, which runs the eTE algorithm, and eTE agents 802 or agent entities in the user plane at PDCP layer, which count the incoming packets per bearer, maintain the virtual queues, and accept/discard incoming packets. The eTE manager exchanges information with eTE agents and eTE agents get the required radio information from DL schedulers 803.

Fig. 9 schematically depicts the whole eTE algorithm (taking into account BW limits at the air interface and distinguishing active and inactive users) represented as a sequence diagram with the following steps.

List of eTE processing steps:

Step 901 DL schedulers provide $FS^{air}_{k,ir,jr}(t)$ for each non-GBR bearer.

Step 902 eTE Agents perform for each non-GBR bearers the following tasks:

- o increment the virtual queue length (according to the bytes received in the interval (t-x,t)
- o send the following information to the eTE Manager:
  - number of bytes which have arrived in interval (t-x,t)
  - $FS^{air}_{k,i,j}(t)$

Step 903 eTE Manager calculates the bandwidth $BW_{trs}(t)$ available in total for all non-GBR bearers

Step 904 eTE Manager calculates for each non-GBR bearer $CarriedLoad^{k,i,j}_{est}(t)$ and decides whether the bearer is active or not.

Step 905 eTE Manager calculates for each active non-GBR bearer the fair share $FS^{trs}_{k,i,j}(t)$ of $BW_{trs}(t)$

Step 906 eTE Manager performs BW redistribution

Step 907 eTE Manager calculates for each active non-GBR bearer $EnforcedRate_{k,i,j}(t)$

Step 908 eTE Manager gives for each non-GBR bearer the following information to the eTE agents:

- o status of the bearer (active or inactive)
- o $EnforcedRate_{k,i,j}(t)$ if the bearer is active

Step 909 eTE Agent receives status information and $EnforcedRate_{k,i,j}(t)$ and performs the following tasks for each non-GBR bearer:

- o If the bearer is inactive: Serve the virtual queue according to the *Minumum_Service_Rate.*
- o If the bearer is active: Serve the virtual queue with $EnforcedRate_{k,i,j}(t)$. Calculate the discard probability function and apply it during (t,t+x).

In addition to the sequence view of Fig. 9, Fig. 10 shows the same steps as Fig. 9 but in an eTE timing view.

In the claims, any reference signs placed in parentheses shall not be construed as limiting the claims. The word "comprising" and "comprises", and the like, does not exclude the presence of elements or steps other than those listed in any claim or the specification as a whole. The singular reference of an element does not exclude the plural reference of such elements and vice-versa. In a device claim enumerating several means, several of these means may be embodied by one and the same item of software or hardware. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

CLAIMS

1. A method for a consistent allocation of bandwidth shares for non-guaranteed bit rate bearers on air interfaces and in a mobile backhaul network section of a packet based mobile service network, the method comprising:

- estimating an overall transport bandwidth available for non-guaranteed bit rate bearers in the mobile backhaul network;

- estimating a fair share of the overall transport bandwidth for each non-guaranteed bit rate bearer and its respective traffic rate; and

- in case of a potential congestion throttling the respective traffic rate of a non-GBR bearer to its respective fair share transport bandwidth.

2. The method according to claim 1, wherein the throttling is done by discarding incoming packets and/or by using Explicit Congestion Notification.

3. The method according to claim 2, wherein respective TCP traffic control mechanisms are invoked for the throttling.

4. A method for a consistent allocation of downlink bandwidth shares for non-guaranteed bit rate bearers on an air interfaces and in a mobile backhaul network section of a packet based mobile service network, the method comprising:

- estimating an overall transport bandwidth available for non-guaranteed bit rate bearers in the mobile backhaul network;

- estimating a fair share of each non-guaranteed bit rate bearer and its respective bandwidth; and

– in case of a potential congestion throttling a respective traffic rate of a non-GBR bearer to its respective fair share bandwidth;

– wherein the throttling is done by discarding incoming packets to a radio transceiver, thereby invoking respective TCP traffic control mechanisms.

5. The method according to any one of the claims 1 to 4, wherein the estimating of the fair share of the overall transport bandwidth a weighting factor is considered for each non-guaranteed bit rate bearer.

6. The method according to any one of the claims 1 to 5 further comprising:

determining whether an actual traffic rate of respective non-guaranteed bit rate bearer is below its respective fair share of transport bandwidth; and

in case it is determined that the actual traffic rate of the respective non-guaranteed bit rate bearer is below its respective fair share of transport bandwidth redistributing the difference between the estimated fair share traffic rate and the actual traffic rate to other bearers.

7. The method according to any one of the claims 1 to 6, wherein the estimation of the fair share of the overall transport is performed for a given time interval.

8. The method according to any one of the claims 1 to 7, further comprising:

deciding whether a non-guaranteed bit rate bearer is an active or inactive bearer.

9. The method according to claim 8, wherein the estimation of the fair share of transport bandwidth is only performed for bearer determined to be an active bearer.

10. A managing entity for a mobile service network, the managing entity comprising:

a calculation unit configured to estimate an overall transport bandwidth available for non-guaranteed bit rate bearers and configured to estimate a fair share of the overall transport bandwidth for each non-guaranteed bit rate bearer and its respective traffic rate; and

a communication interface configured to communicate information indicative of the estimated fair share.

11. The managing entity according to claim 10, wherein the calculation unit is further configured to decide for each non-guaranteed bit rate bearer whether it is an active bearer or not.

12. The managing entity according to claim 11, wherein the estimation of a fair share is only performed for active bearers.

13. The managing entity according to any one of the claims 10 to 12, wherein the managing unit is configured to perform a bandwidth redistribution.

14. The managing entity according to claim 13, wherein the calculation unit is further configured to calculate an enforcement rate for at least one active bearer.

15. An agent entity for a mobile service network, the agent entity comprising:

a counting unit configured to count incoming data packets per non-guaranteed bearer; and

a communication interface configured to communicate with a managing entity according to any one of the claims 10 to 14.

16. The agent entity according to claim 15, wherein the agent entity is configured to perform a throttling by discarding packets and/or by using Explicit Congestion Notification.

17. A packet based mobile service network comprising a managing entity according to anyone of the claims 10 to 14 and an agent entity according to anyone of the claims 15 or 16.

18. A computer program product comprising software, which when loaded into the memory of a computer enables the computer to execute any of the steps of the method of any one of the claims claim 1 to 9.

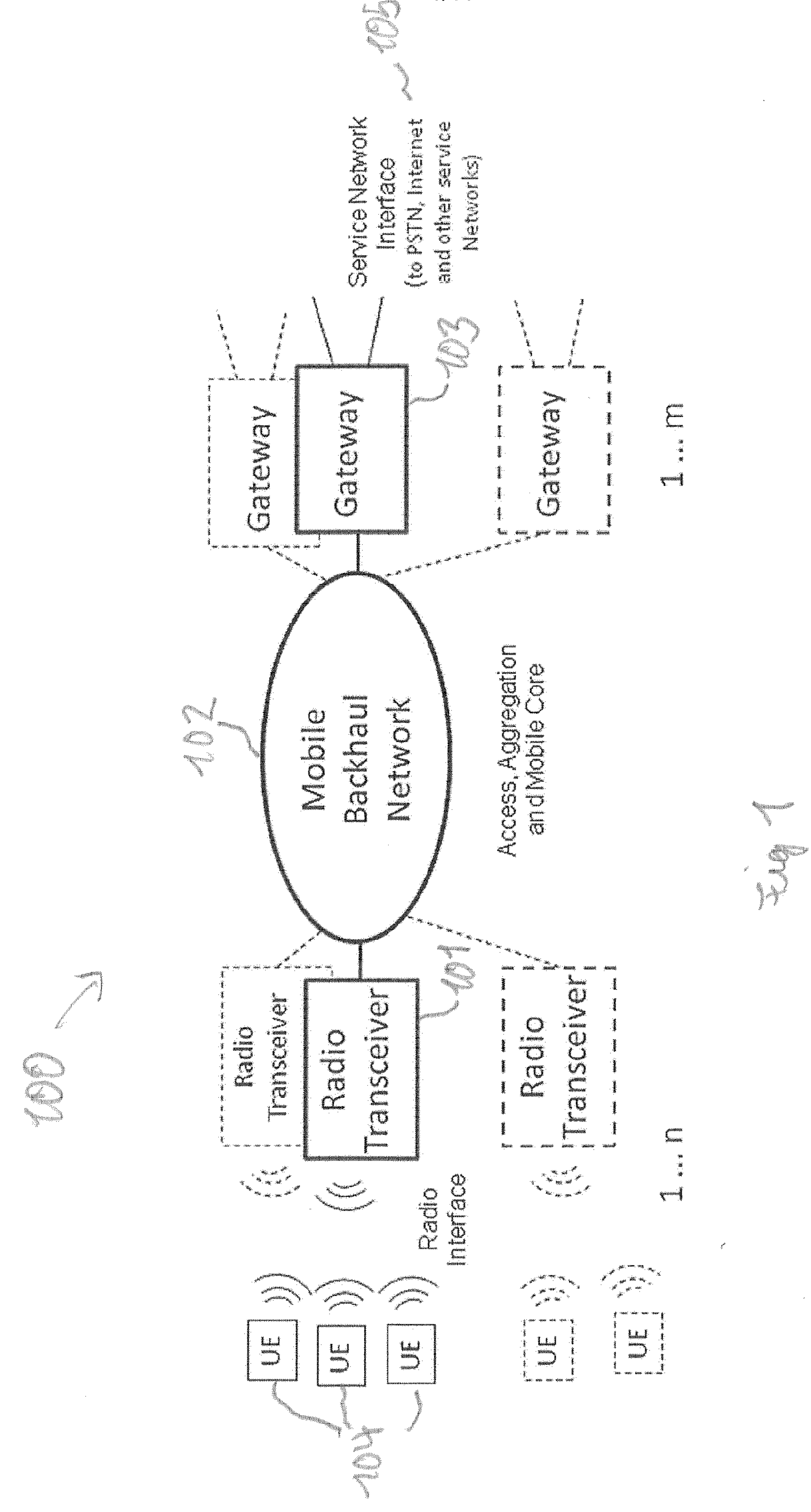

100
UE
UE
UE
UE
UE
104
Radio Interface
Radio Transceiver
Radio Transceiver
Radio Transceiver
101
1 … n
Mobile Backhaul Network
102
Access, Aggregation and Mobile Core
Gateway
Gateway
Gateway
103
1 … m
Service Network Interface (to PSTN, Internet and other service Networks)
105

Fig 1

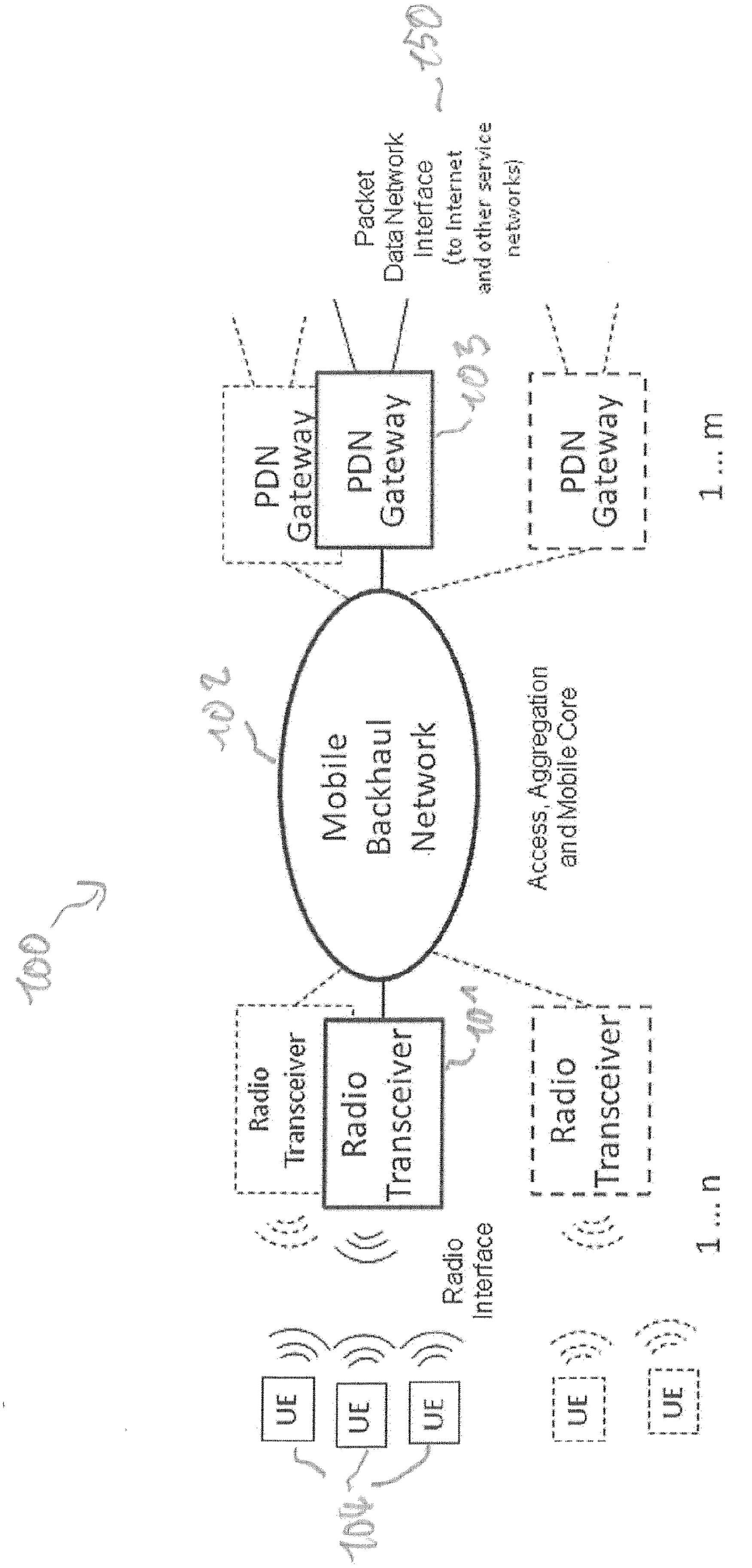

100
Packet Data Network Interface (to Internet and other service networks)
150
PDN Gateway
PDN Gateway
103
PDN Gateway
1 … m
Mobile Backhaul Network
102
Access, Aggregation and Mobile Core
Radio Transceiver
Radio Transceiver
101
Radio Transceiver
1 … n
Radio Interface
UE
UE
UE
UE
UE
104

Fig 2

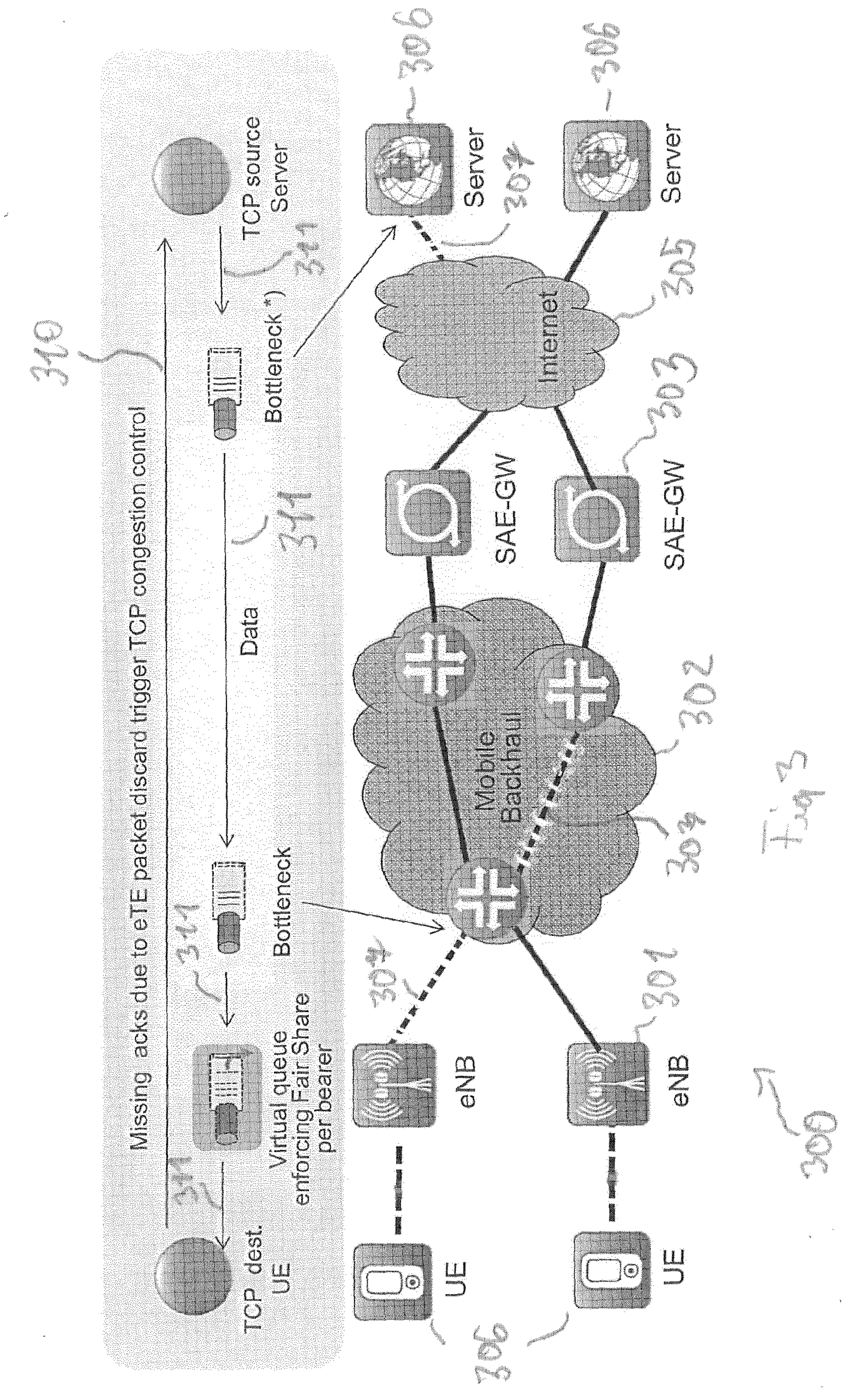

Missing acks due to eTE packet discard trigger TCP congestion control
TCP source Server
Bottleneck *)
Data
Bottleneck
Virtual queue enforcing Fair Share per bearer
TCP dest. UE
Server
Server
Internet
SAE-GW
SAE-GW
Mobile Backhaul
eNB
eNB
UE
UE
310
311
306
307
305
303
302
301
300

Fig 3

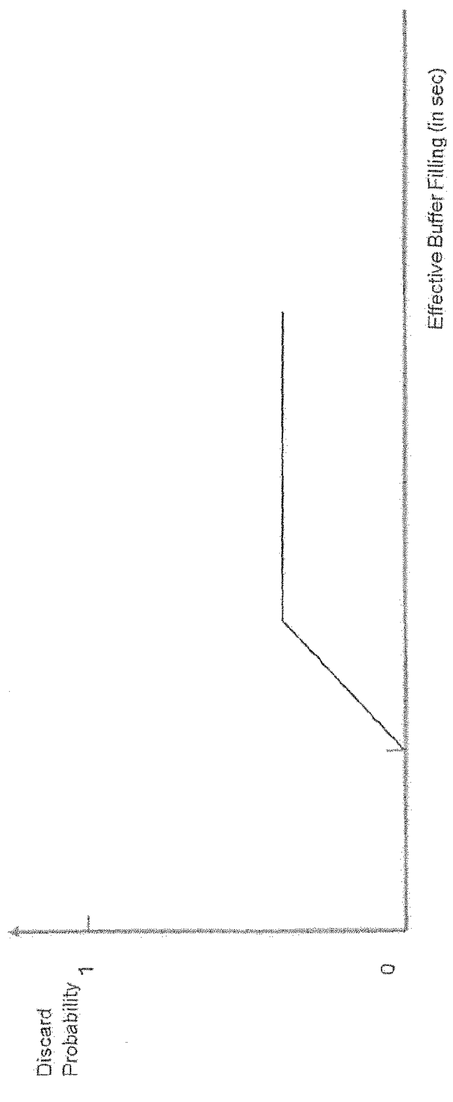
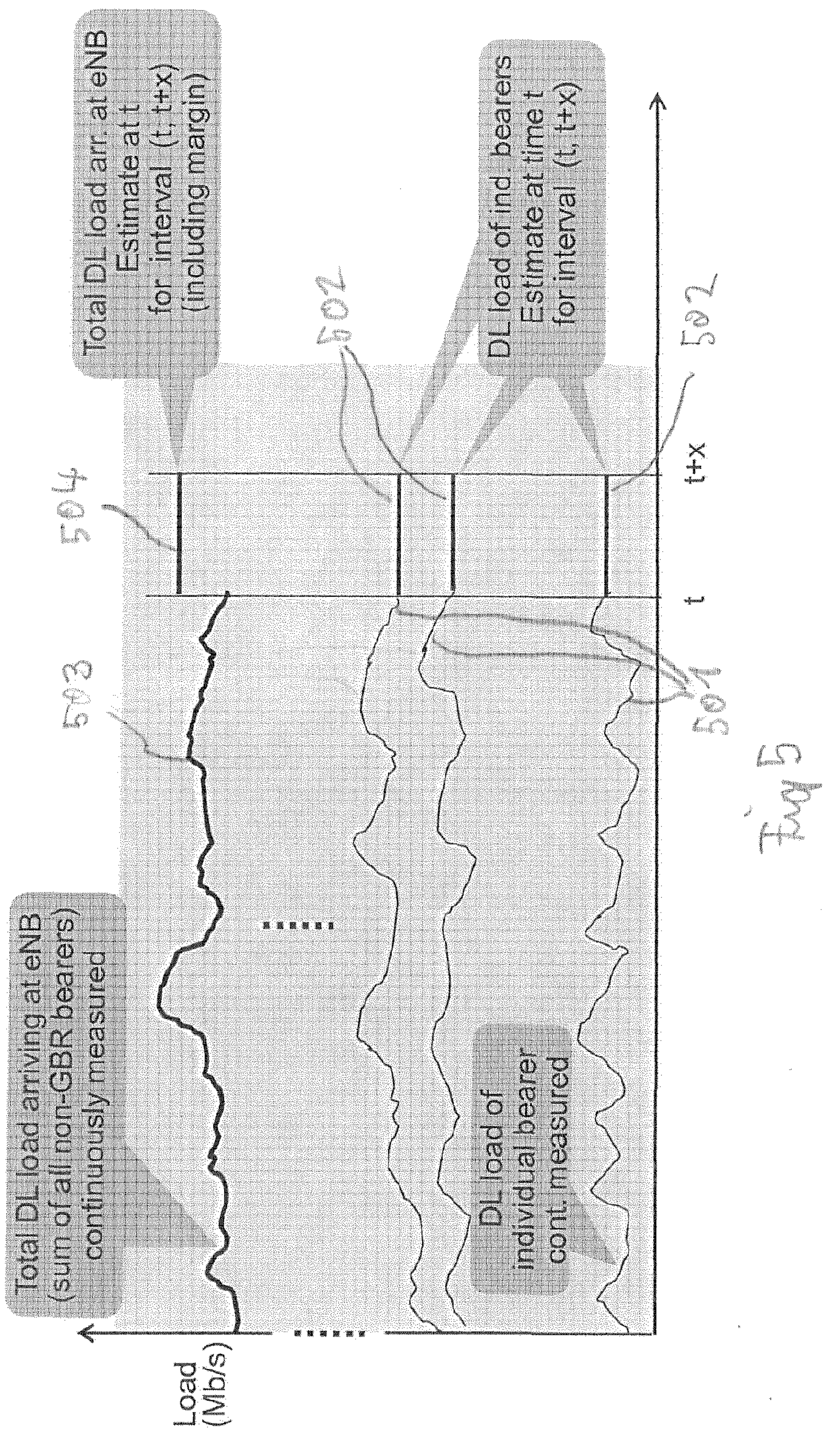


Fig 4


Total DL load arriving at eNB (sum of all non-GBR bearers) continuously measured
Total DL load arr. at eNB Estimate at t for interval (t, t+x) (including margin)
DL load of ind. bearers Estimate at time t for interval (t, t+x)
DL load of individual bearer cont. measured
Load (Mb/s)
t
t+x
501
502
503
504
Fig 5

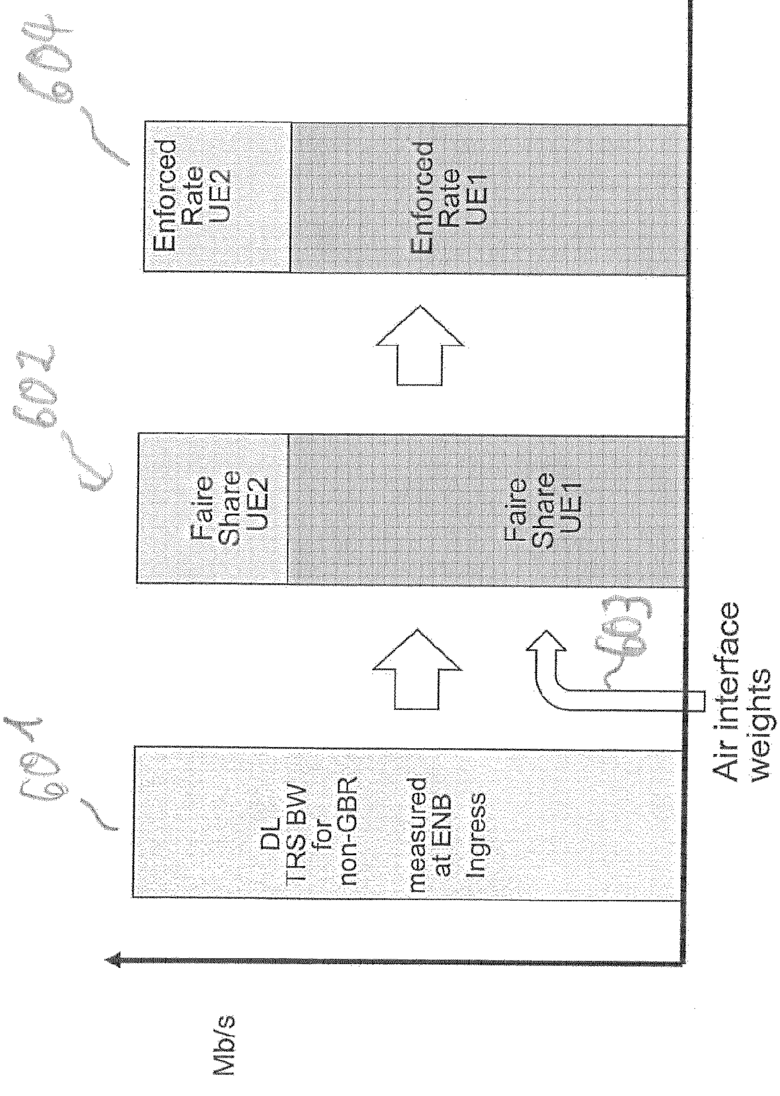

604
Enforced Rate UE2
Enforced Rate UE1
602
Faire Share UE2
Faire Share UE1
601
DL TRS BW for non-GBR
measured at ENB Ingress
603
Air interface weights
Mb/s

Fig 6

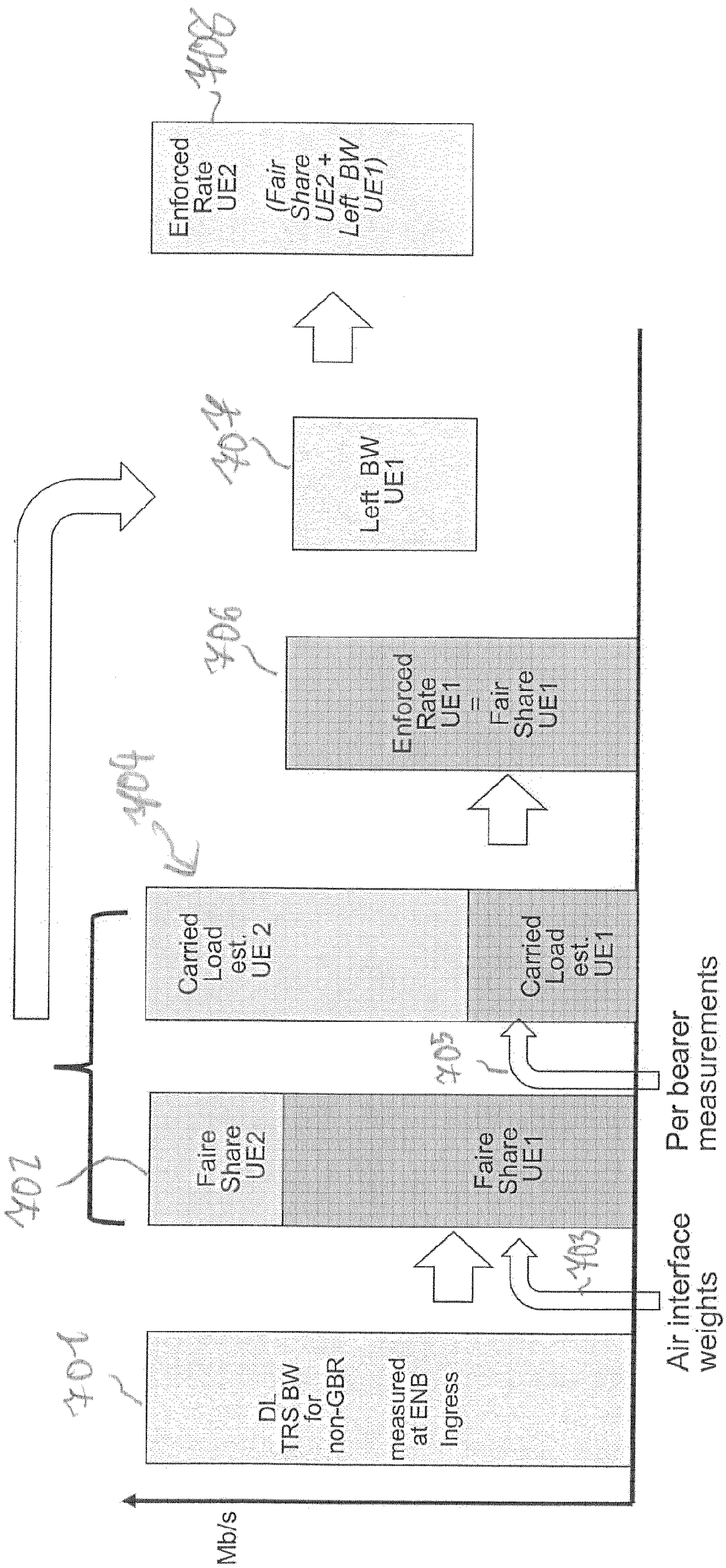

Mb/s
DL TRS BW for non-GBR measured at ENB Ingress
701
Air interface weights
703
Faire Share UE1
Faire Share UE2
702
Per bearer measurements
705
Carried Load est. UE1
Carried Load est. UE 2
704
Enforced Rate UE1 = Fair Share UE1
706
Left BW UE1
707
Enforced Rate UE2 (Fair Share UE2 + Left BW UE1)
708

Fig 7A

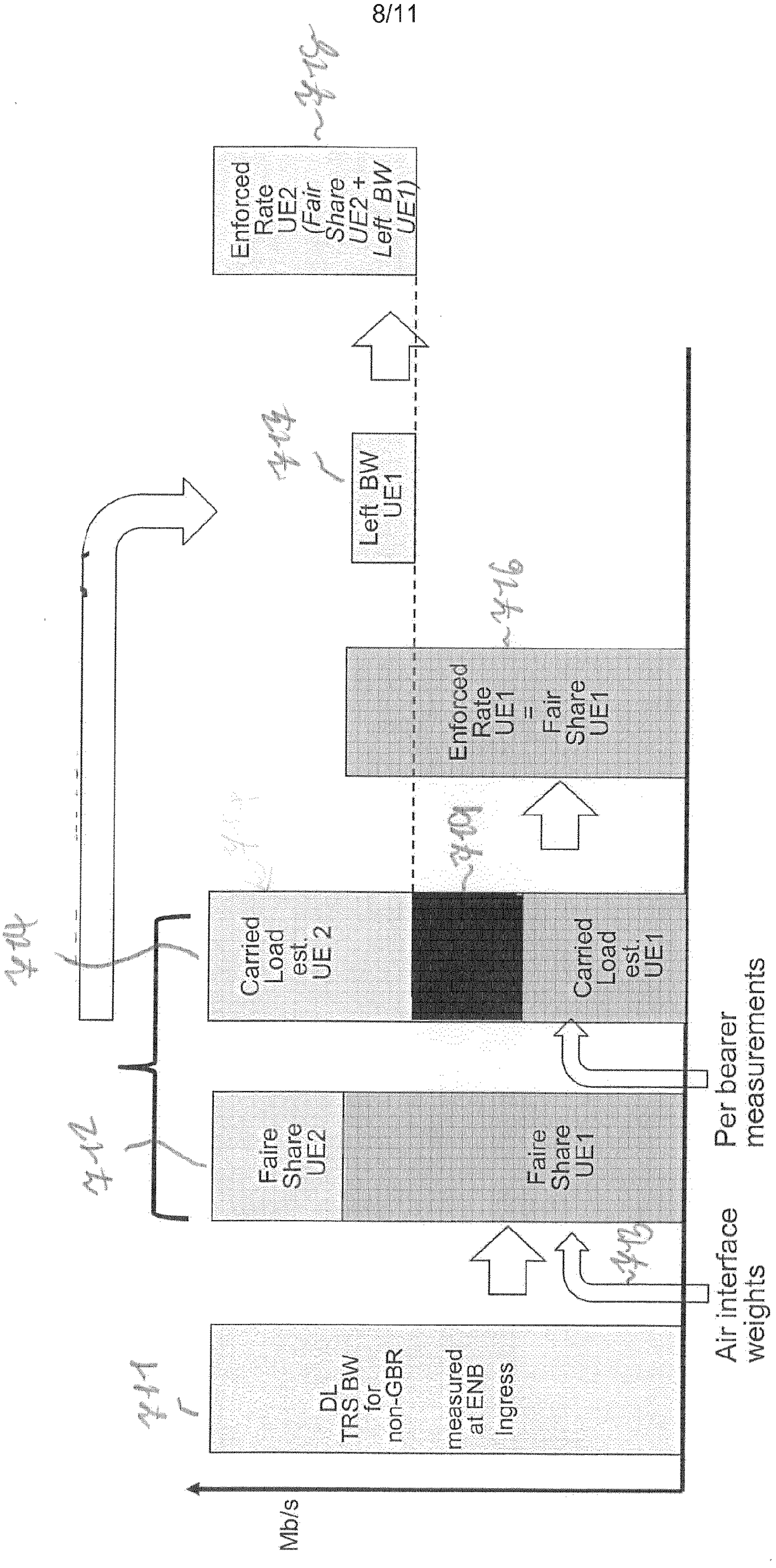

Mb/s
DL TRS BW for non-GBR measured at ENB Ingress
711
Air interface weights
713
Faire Share UE2
Faire Share UE1
712
Per bearer measurements
Carried Load est. UE 2
Carried Load est. UE1
701
714
Enforced Rate UE1 = Fair Share UE1
716
Left BW UE1
717
Enforced Rate UE2 (Fair Share UE2 + Left BW UE1)
718

Fig 7B

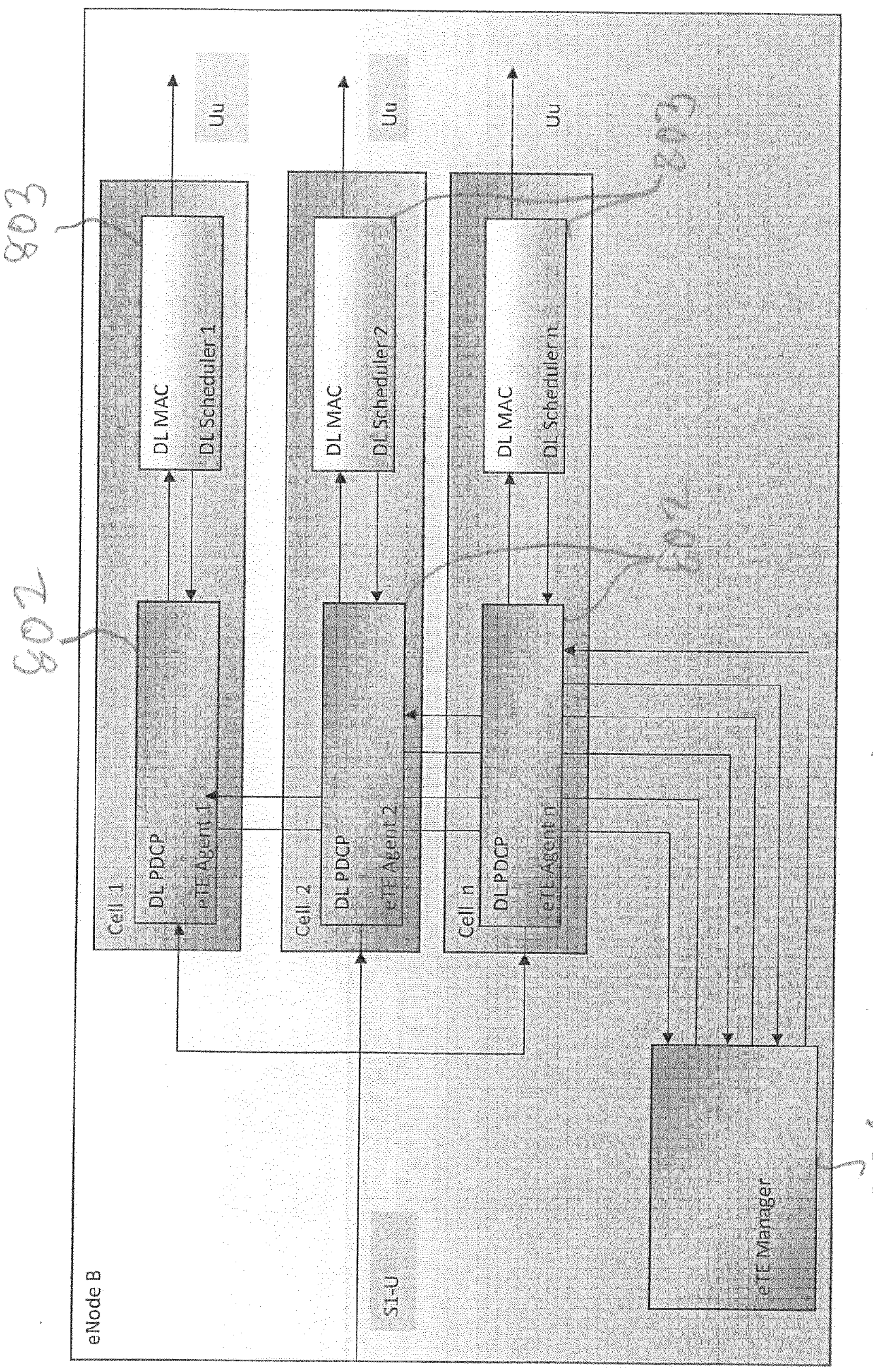

eNode B
Cell 1
Cell 2
Cell n
DL PDCP
eTE Agent 1
eTE Agent 2
eTE Agent n
DL MAC
DL Scheduler 1
DL Scheduler 2
DL Scheduler n
Uu
S1-U
eTE Manager
801
802
803

Fig 8

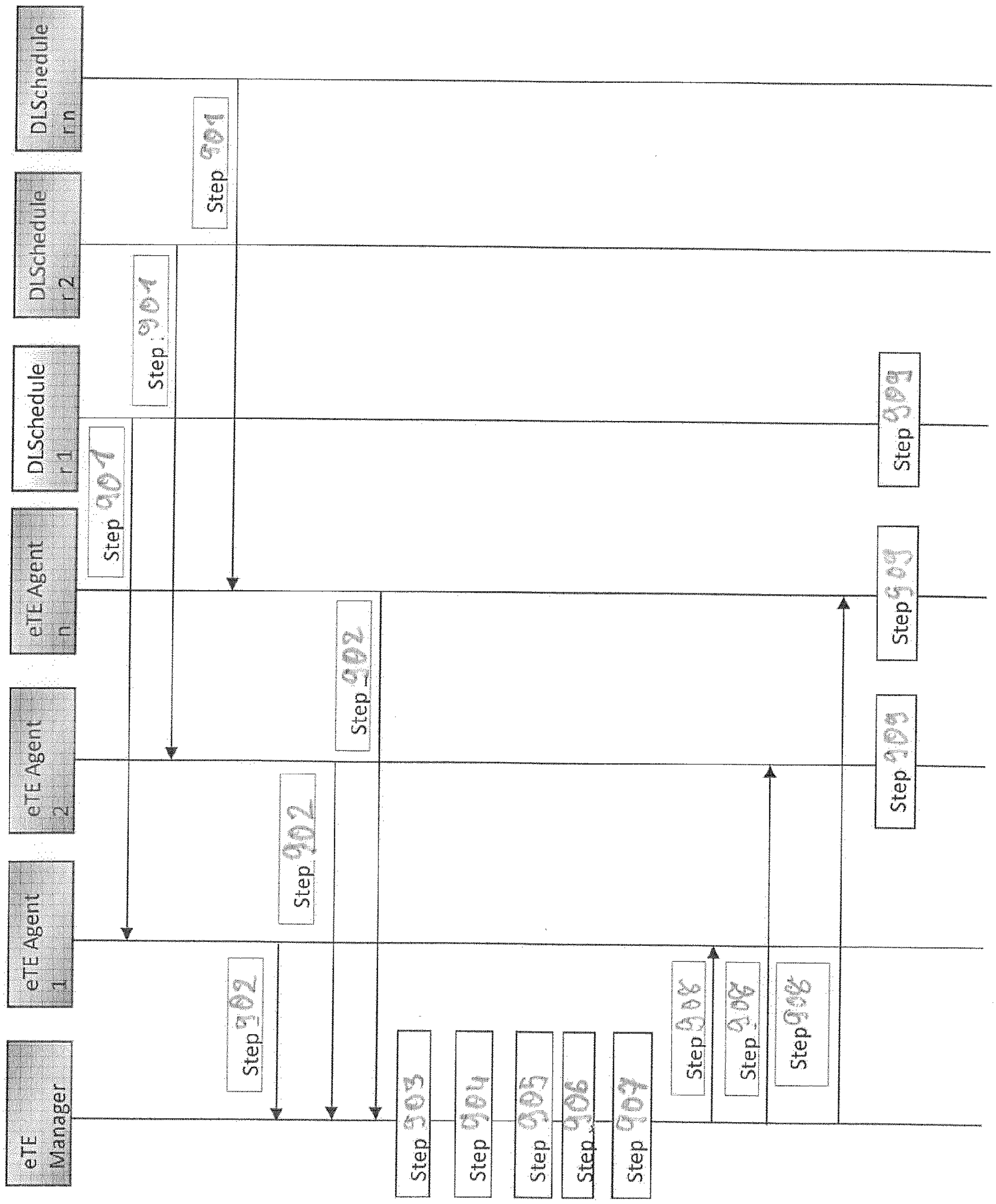

eTE Manager
eTE Agent 1
eTE Agent 2
eTE Agent n
DLSchedule r 1
DLSchedule r 2
DLSchedule r n
Step 901
Step 901
Step 901
Step 902
Step 902
Step 902
Step 903
Step 904
Step 905
Step 906
Step 907
Step 908
Step 908
Step 908
Step 909
Step 909
Step 909

Fig 9

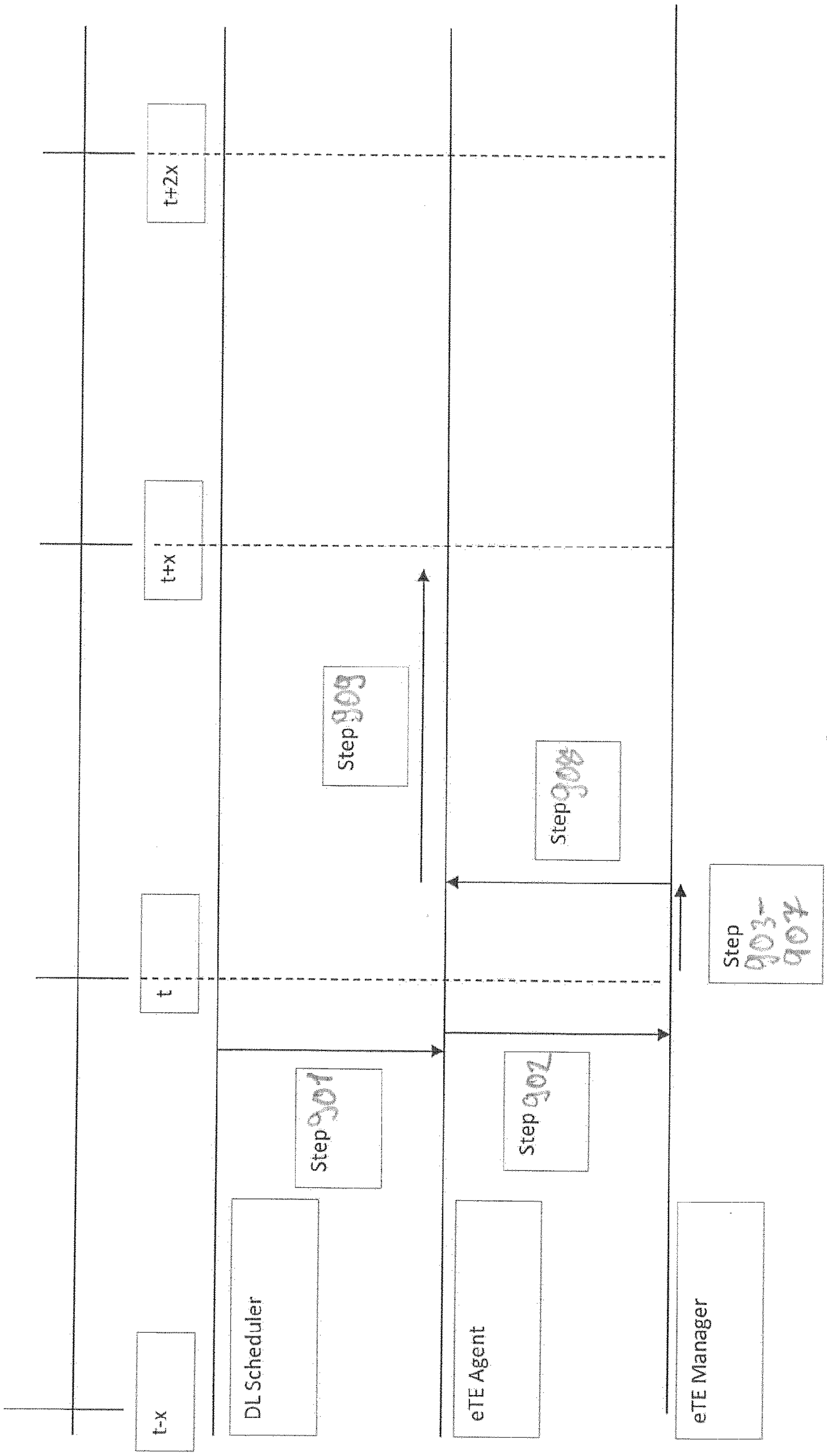

t-x
t
t+x
t+2x
DL Scheduler
eTE Agent
eTE Manager
Step 901
Step 902
Step 903-907
Step 908
Step 909

Fig 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/053388

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04W28/02
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04W H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	US 2012/218892 A1 (KOTECHA LALIT R [US] ET AL) 30 August 2012 (2012-08-30) abstract figures 4,8-10 paragraphs [0001], [0015], [0016], [0020], [0021], [0024], [0028], [0041], [0045], [0049], [0053] - [0061], [0075], [0090] -----	1-5, 7-12, 15-18 6,13,14

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 May 2014	22/05/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Manuel, Grégory

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EKSTROM H: "QoS control in the 3GPP evolved packet system", IEEE COMMUNICATIONS MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, US, vol. 47, no. 2, 1 February 2009 (2009-02-01), pages 76-83, XP011280742, ISSN: 0163-6804, DOI: 10.1109/MCOM.2009.4785383 the whole document -----	1-18
A	US 6 069 872 A (BONOMI FLAVIO [US] ET AL) 30 May 2000 (2000-05-30) the whole document -----	1-18
A	US 2007/127419 A1 (SAPEK ADAM [US]) 7 June 2007 (2007-06-07) the whole document -----	1-18
A	WO 2009/008817 A1 (ERICSSON TELEFON AB L M [SE]) 15 January 2009 (2009-01-15) the whole document -----	1-18
A	WO 2008/055936 A1 (ERICSSON TELEFON AB L M [SE]; ERIKSSON ANN-CHRISTINE [SE]) 15 May 2008 (2008-05-15) the whole document -----	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2014/053388

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
US 2012218892	A1	30-08-2012	NONE			
US 6069872	A	30-05-2000	NONE			
US 2007127419	A1	07-06-2007	NONE			
WO 2009008817	A1	15-01-2009	CA	2695010	A1	15-01-2009
			CN	101796777	A	04-08-2010
			EP	2165480	A1	24-03-2010
			JP	2010532952	A	14-10-2010
			US	2009067335	A1	12-03-2009
			WO	2009008817	A1	15-01-2009
WO 2008055936	A1	15-05-2008	EP	2080326	A1	22-07-2009
			JP	5279717	B2	04-09-2013
			JP	2010508744	A	18-03-2010
			US	2009310487	A1	17-12-2009
			WO	2008055936	A1	15-05-2008