

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 September 2002 (26.09.2002)

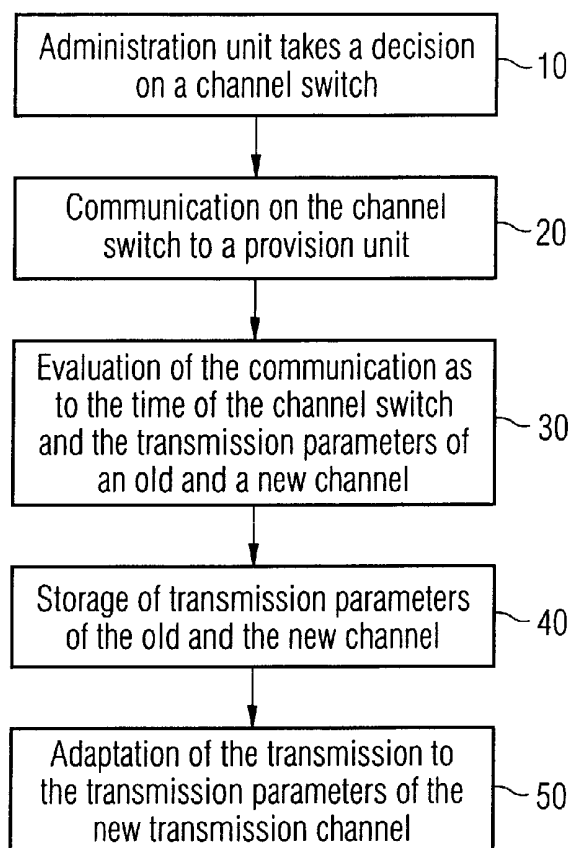
PCT

(10) International Publication Number
WO 02/076036 A1

- (51) International Patent Classification⁷: **H04L 12/56**, (74) Agent: **MOHSLER, Gabriele**; Ericsson Eurolab Deutschland GmbH, Ericsson Allee 1, 52134 Herzogenrath (DE).
- (21) International Application Number: PCT/EP02/02264
- (22) International Filing Date: 2 March 2002 (02.03.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
01106858.2 20 March 2001 (20.03.2001) EP
- (71) Applicant: **TELEFONAKTIEBOLAGET LM ERICSSON (publ)** [SE/SE]; SE-126 25 Stockholm (SE).
- (72) Inventors: **LUDWIG, Rainer**; Maubacher Str. 4, 52393 Hürtgenwald (DE). **MEYER, Michael**; Grossheidstr. 27, 52080 Aachen (DE).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent

[Continued on next page]

(54) Title: METHOD AND DEVICE FOR IMPROVING A DATA THROUGHPUT



(57) Abstract: The invention relates to a method and a device for improving a data throughput in a transmission via communication network, especially in a channel switch, where a newly assigned channel has a smaller transmission rate. A decision on the channel switch is thereby taken in an administration unit, which is communicated to a provision unit. Upon receipt of the communication, the transmission parameters of an old and a new channel are stored. On the basis of said information an adaptation to the transmission parameters of the new channel is performed. If the communication takes place prior to the performance of the channel switch, the number of data packets transmitted is reduced so as to achieve a better adaptation and to avoid abrupt changes in the transmission rate. If the communication on a channel switch takes place only after the performance thereof, the abrupt reduction of the transmission rate is compensated in that more data are temporarily transmitted in addition to the proportionally permitted transmission rate of the new channel.

WO 02/076036 A1



(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Declaration under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Method and device for improving a data throughput

- 5 The invention relates to a method, a device and a computer program for improving a data throughput during a transmission via a communication network, in particular, if a channel switch takes place on a part connection within the communication network.
- 10 The transmission of data between the communicating partner instances is based on standardized protocol stacks. A protocol stack is formed of several protocol layers, hereinafter also referred to as layers. The uppermost layer, the application layer, comprises applications directly used by a user. The internet use is, for example, one possible application. A transport layer with the
- 15 corresponding transport protocols such as the TCP (Transmission Control Protocol) is provided underneath the application layer. The Internet Protocol (IP) is an example for the network layer. The two lowermost layers, the link layer and the physical layer, may be summarized by the term network-oriented layer, since they are specifically defined in response to the underlying network.
- 20 The transport protocol TCP provides a reliable transport service of data between two communicating partners, whereby the TCP is responsible for an end-to-end transmission, possibly via a heterogeneous connection, which may section-wise be characterized by different network topologies, transmission
- 25 rates and transmission delays. Laborious mechanisms guarantee the adaptation of the transmission to the characteristics of the connection link as well as robustness against the occurrence of errors. The precise description of the TCP methods may be inferred from "TCP/IP. Illustrated, Volume 1" by W. Richard Stevens. In the following, only a few of said mechanisms will be
- 30 entered into in more detail.

For the time being a method is introduced, which relates to the detection of packet losses, which may take place if errors occur or in an overload situation during a transmission.

5 A mechanism for the detection of packet losses is based on the so-called cumulative acknowledgement scheme by TCP. If a packet loss occurs during a transmission, the receiver becomes aware thereof by receiving a data packet, which, according to the sequence number, was not expected as the next one in the sequence. As a result, an acknowledgement message is sent to the sender,
10 which includes the sequence number of the data packet not received. In this case, said acknowledgement messages are called duplicate acknowledgments, and the receipt of the third duplicate acknowledgment indicates the sender the loss of the corresponding data packet, whereupon a retransmission of the data packet is initiated. The receipt of three duplicate acknowledgments will be
15 interpreted by the sender as a temporary interruption. An acknowledgement message is sent to the sender once the receiver has received another data packet, and due to the fact that additional acknowledgements are received, the sender concludes that the data packets sent after the faulty data packet were actually received. This indicates that merely the explicit sending of the faulty
20 data packet suffices to eliminate the error. A permanent disturbance is recognized by another mechanism, which will be introduced in the following.

The TCP mechanism, which recognizes a serious loss of data packets, consists in the use of time limits. An expiration of a time limit is the so-called timeout. A
25 time limit is associated with a time margin provided for performing a certain task. If the task is not realized within said time margin, a timeout occurs. In the case of TCP, for example, a time margin is defined, the so-called RTT (Round Trip Time), which indicates the time passing between the delivery of a TCP packet to the IP layer and the receipt of a pertinent acknowledgement message
30 from a receiver, that said data packet was received correctly. This time margin enables an adaptation to different transmission conditions, as said value contains a message on the present performance of a transmission. On the

basis of the average value and the variance of the RTT, an RTO (Retransmission Timeout) is determined. An RTO designates the time point, at which a retransmission of a data packet takes place, for which it took longest to receive a corresponding acknowledgement message on the receipt thereof. The exact determination of the RTO can be inferred from RFC 2988, Transmission Control Protocol, November 2000. With timeout-based error recognition the exactness of the RTO calculation largely influences the data throughput. If the timeout has been scheduled too short, unnecessary data packets are retransmitted. If it is too long, excessive time passes until the loss of a data packet is noticed, which again results in time intervals during which one inactively waits for the continued transmission. The determination of an RTO takes place dynamically during a transmission so as to thereby record the RTT fluctuations in a transmission as precisely as possible. Despite said measures, the algorithm for calculating the RTO value entails problems, if abrupt fluctuations in the RTT occur. An improved algorithm has been developed, which provides more exact values for the RTO. Each time the RTO value is measured, a smoothed average value of the deviation between the previous average value of the RTT and the actually measured RTT values is additionally formed anew. If, the RTT changes abruptly, however, after merely small deviations had occurred for a longer time period in advance, also the calculated mean deviation cannot compensate the fluctuations immediately. If the RTT shows an essentially larger value than the RTO, it may be possible that the timeout for a sent data packet expires before an acknowledgement message for said data packet is received. The occurrence of timeout is evaluated as a data packet loss and, at the same time, as an indication for transmission problems. Consequently, although a data packet was transmitted and received correctly, the same is recognized as having been lost, and TCP mechanisms are started for eliminating the error and for avoiding the occurrence of additional errors. The basic idea behind said TCP mechanisms is to immediately reduce the transmission rate, at which the packets are forwarded to the communication network, when packet losses occur. Afterwards the rate is slowly increased so as to approach the actually existing network capacity.

The algorithms used for the flow control of the data packets, which avoid the occurrence of subsequent transmission errors, are called congestion avoidance and slow start algorithms. For the realization of said two algorithms, the so-called window technique is applied. With said technique, two kinds of windows are administered by the sender; a window whereof the size was suggested and confirmed by the receiver, the so-called advertised window, and the so-called congestion window. The number of packets which can be sent corresponds to the minimum of both windows. The determination of the size of the congestion window takes place step by step. Upon the initiation of a connection, a sender starts to send a packet. Upon receipt of an acknowledgement message from the receiver on the correct receipt the congestion window is increased by one, and, as a consequence, two packets are being sent. After the acknowledgement messages for the two sent packets are received, four packets and so on are being sent until a window size suggested by the receiver when the connection was set up is reached, or until a packet loss has occurred. In this connection, a packet loss is detected by a timeout or by the receipt of the third duplicate acknowledgment. Upon the detection of a packet loss by means of a timeout the size of the congestion window is reduced to one segment, whereby a slow start algorithm is initiated, and the timeout is doubled over the present value. Upon the initiation of a slow start algorithm the size of the congestion window at first grows exponentially. The exponential growth is maintained until half of the original window size is reached, i.e. the window size valid before the occurrence of a packet loss. Thereafter, the growth becomes linear until a packet loss is again recognized by a timeout or by the receipt of the third duplicate acknowledgment. In other words, if a permanent disturbance takes place, the window size is set at one packet maximum, and the RTO is set at double the value of the previous RTO value. Upon the restart of the transmission the window size is regulated by the slow start algorithm, and the RTO is calculated anew in response to the RTT.

- In the case where a packet loss is detected by means of duplicate acknowledgments, however, no slow start algorithm is started. Instead, the size of the congestion window is reduced by half. In this case, only the packet to which the duplicate acknowledgments refer is retransmitted, and the transmission of the following packets is continued subsequently. Said method is also known as the so-called Fast Retransmit and Fast Recovery. An exact description thereof can be found in "TCP/IP. Illustrated, Volume 1" by W. Richard Stevens.
- 10 In other words, a detected packet loss results in a retransmission of the lost data packet and in the application of TCP methods such as the Fast Retransmit or Slow Start, which have been provided to eliminate packet losses and to avoid the occurrence of subsequent errors. A retransmission of a data packet and the application of the TCP methods, which are started upon the detection of a data
- 15 packet loss, require resources and time. This has a negative effect on the transmission performance in those situations, in which a packet loss is recognized superfluously. For example, this happens if a transmission is delayed due to a short interruption, and the result may be that the receipt of an acknowledgement message takes such a long time that, in the meantime, a
- 20 timeout expires and consequently a packet loss is recognized. In reality, however, the packet is received. Since, however, a packet loss was detected, the data packets already sent and correctly received are unnecessarily retransmitted and TCP mechanisms associated therewith are unnecessarily started. The superfluous occurrence of a timeout expiration is called spurious
- 25 timeout, and the retransmission of already sent packets takes place according to the go-back-N algorithm.

This becomes particularly problematical in communication networks characterized by high fluctuations in the transmission rates. Part of said

30 networks are cellular networks such as the GSM (Global System for Mobile Communication), GPRS (General Packet Radio Service); UMTS (Universal

Mobile Telecommunication System), EDGE (Enhanced Data rates for GSM Evolution) or IMT-2000.

The spurious occurrence of a timeout expiration often takes place especially in
5 mobile networks, however, above all, if sudden delays in the transmission occur
due to a channel switch. A channel switch may occur due to a change between
communication networks, for instance, due to a change from UMTS to GSM or
from WLAN (Wireless Local Area Network) to GPRS. A channel switch may
also take place in a communication network, for example, due to a handover.
10 Another cause for the channel switch may be based on the priorities for the
treatment of different users. The users, who make a phone call, often have a
higher priority than users carrying out a data transfer. Therefore, if new users
wanting to make a phone call join in, the data users can be shifted to other
transmission channels. The channels are characterized by transmission
15 parameters such as the transmission rate or the packet size, and a channel
switch may, for example, result in different transmission rates, or at least lead to
other RTTs measured on the link layer and, thus, possibly to the expiration of
the timeout on the corresponding layers. A user may, for example, be shifted
from one channel having a high transmission capacity to a channel having a
20 smaller transmission capacity.

The changes associated with the channel switch on the physical layer are
noticed by the transport layer, especially the TCP, only after the channel switch
has taken place, namely because an abrupt change in the RTT is recorded.
25 Said changes may result in that a timeout is exceeded and thus in a
retransmission of a data packet and in the start of the corresponding methods
relating to the elimination of the occurred transmission error and to the
avoidance of the occurrence of additional transmission errors. The channel
switch commonly does not entail a data packet loss. For this reason, the
30 determination of the packet loss by a timeout expiration and the performance of
the TCP mechanisms started with said determination are superfluous. Said
mechanisms require resources and a long time, resulting in the reduction of the

transmission throughput. For example, given the slow start algorithm, the transmission capacity available is not optimally exploited due to the slowly increasing window size.

5 In accordance therewith it is the object of the present invention to provide a method and a device guaranteeing an optimized exploitation of the transmission capacity being available. The object of the present invention especially relates to an optimized use of the transmission capacity being available in a channel switch.

10

According to the invention this object is provided by the teaching of patent claim 1 and by the teaching of patent claims 11 and 15.

15

It is thereby an advantage that a superfluous adaptation of the protocols of the higher layers is avoided, so that the resources during a transmission are exploited in a more favorable manner.

Additional advantageous embodiments according to the invention are described in patent claims 2 to 10 and 12 to 14.

20

In the following, the invention will be explained in more detail by means of examples and figures, wherein

25

Fig. 1 shows a flow chart according to the present invention,

Fig. 2 illustrates a connection via a communication network,

Fig. 3a is a schematic illustration of the transmission rates of two channels,

30

Fig. 3b shows pre-scheduling,

Fig. 3c shows post-scheduling by using an additional channel,

Fig. 3d shows post-scheduling with a commonly used channel.

The invention will hereinafter be explained in more detail by means of figure 1
5 and patent claim 1.

A channel switch during a transmission, or an assignment of channels during a connection set-up takes place in a communication network by means of an administration unit for administering channels and a provision unit for providing,
10 distributing and sending data via the channels. The channels are characterized by transmission parameters. The term transmission parameter covers different quantities such as the transmission rate, the packet size, the so-called processing time, which designates the delay time with the sender and the recipient due to the data processing. Additional transmission parameter will
15 hereinafter be dealt with in more detail. The transmission parameters cause the delay time of the packets, and the adaptation of the transmission parameters influences the adaptation of the packet delay times. In this respect the packet delay time stands for the delays in the transmissions of data packets, which may, for example, be caused by the reduction in the transmission rate. For this
20 reason, the definitions introduced in this connection will in the following be used alternatively, unless it is explicitly mentioned as to which transmission parameters are concerned.

According to figure 1 the decision on the channel switch is taken by the
25 administration unit in a first step 10. Said unit has the necessary information and corresponding algorithms, which assist in taking said decision. This decision, which contains the time of the channel switch and the transmission parameters of a new and an old channel, is communicated to the provision unit,
20. The provision unit evaluates said communication, whereby it is especially
30 evaluated as to whether a channel switch is to be carried out immediately or delayed, 30. On the basis of said information an adaptation of the transmission parameters is realized either before or after the performance of the channel

switch. Another possibility provides for the realization of the adaptation both, before or after the performance of a channel switch. The adaptation 50 takes place on the basis of the determination of the data packets to be transmitted. For this purpose, the packet delay time caused by the transmission parameters of the new channel as well as of the old channel are taken into account. If it is, for instance, established that the packet delay time of the new channel is longer, the packet delay time of the old channel is deliberately increased prior to the performance of the channel switch, namely by providing and transmitting fewer data packets. In the case where a channel switch has already taken place, the packet delay time is reduced by temporarily transmitting more data packets, for example, via an additional data channel.

In order to realize an adaptation of the transmission parameters of channels having different packet delay times, a storage of the transmission parameters of an old or/and a new channel has to be performed, e.g. by means of storing them in the means provided for said purpose, 40. In the case where a channel switch has already taken place, the transmission parameters of the old channel are secured so as to guarantee a subsequent adaptation of the transmission parameters to the conditions of the new channel on the basis of the transmission parameters of the old and the new channel. In the case of a forthcoming channel switch, the storage of the transmission parameters of the new channel takes place such that an adaptation of the transmission parameters of the old channel, which, at this time, is still in use, however, to the transmission parameters of the new channel used after the performed channel switch is guaranteed.

A system comprising an administration unit for administering channels and a provision unit for providing data for the transmission via the channels is provided for the realization of the inventive solution to improve a data throughput of a transmission via a communication network. Said system has a decision means, which is, for example, disposed in the administration unit. The decision means takes a decision on a channel switch. The decision is

communicated to the provision unit by means of a communication means. Upon the receipt of the communication, the transmission parameters of an old channel and a new channels are stored in a memory. This is preferably performed in the provision unit. On the basis of the transmission parameters of the old channel and the new channel, an adaptation of the transmission of the old channel to the transmission parameters of the new channel is performed by an adaptation means.

In the following the invention will be explained in more detail by means of an embodiment, whereby the embodiment exemplarily relates to the UMTS.

For this purpose, a network system is used, which will here briefly be explained in a schematic manner. The UMTS network comprises user equipment being able to communicate via an UTRAN (UMTS Terrestrial Radio Access Network) with the CN (Core Network). The user equipment, hereinafter also referred to as mobile station, may be a UMTS terminal. The mobile station communicates with a node B via a radio connection. Said node B, again, is connected to an RNC (Radio Network Controller) being part of the UTRAN. The RNC is connected to the CN. Another unit, the so-called gateway GGSN is responsible for a communication with an external network, for example, another mobile network or the Internet.

In the following the network-oriented layers of the protocol stack of a mobile station and a UTRAN are explained in more detail by means of figure 2.

Figure 2 illustrates schematically a connection between a mobile station MS and a so-called server AS. As was described above, a UTRAN (UMTS Terrestrial Radio Access Network), a CN (Core Network) and a gateway are involved in said connection. Except for the CN, said units are illustrated by means of protocol stacks. As has already been described, a protocol stack consists of several protocol layers. The protocol layers are, according to figure 2, correspondingly designated in the respective protocol stacks of the units. The

protocol stack of the MS, for example, contains an application layer A.

Thereunder are provided the transport layer TCP, then the network layer IP and thereunder the network-oriented layers. The network-oriented layers contain a convergence protocol PDCP (Packet Data Convergence Protocol), a thereunder
5 integrated RLC (Radio Link Control) layer, an MAC (Medium Access Control) layer and a physical layer (PHY). A similar structure of the network-oriented layers is likewise contained in the UTRAN. The continuous connecting lines point to the logic connections between the respective layers. As an example it can be seen from figure 2 that a TCP connection is an end-to-end connection.

10 The network-oriented layers of the MS and UTRAN will be dealt with in more detail in the following. The network-oriented layers of the remaining units have merely been pointed out by the hints L1 and L2, as a detailed structure of said protocol layers is associated with the underlying network (CN) and is irrelevant in this respect.

15

The PDCP constitutes the interface between the network layer, e.g. the Internet protocol (IP) layer and the RLC layer. The PDCP's task hereby consists in the conversion, especially in the segmentation, of the IP packets into smaller data packets. With each connection set-up one or several RLC instances are
20 activated on the RLC layer, which is regulated by the RRM (Radio Resource Management). The RLC layer is, inter alia, responsible for the subdivision of the received data into smaller data packets. On the basis of said data packets the correction of the errors can take place.

25 The MAC layer is responsible for the mapping of the logical channels to the transport channels, which comprises the multiplexing of the logical channels, the distribution of the data onto the different logical channels as well as the change of the logical channels onto the available transport channels on the basis of the decision taken in the RRM. In this connection, different kinds of the
30 logical channels are distinguished in response to the used service. For example, a difference is made between the kind of information to be transmitted, user data or control data. The control data are transmitted by

control channels. The Broadcast Control Channel (BCCH), for instance, transmits network information to the terminals being present in a cell and not coupled to the network, in other words, the terminals dispose of a channel connection. The different control channels of the MAC layer will not be dealt
5 with in more detail at this point.

The so-called traffic channels transmit user data, whereby a difference is made between the so-called DTCH (Dedicated Traffic Channel) and the so-called CTCH (Common Traffic Channel). A DTCH is exclusively assigned to a
10 terminal, so that the user data are transmitted exclusively by the DTCH. In contrast thereto, the CTCH is destined for the transmission of the data to a plurality of users.

Another differentiation in the division of the channels is made in view of the
15 transmission direction. The data are either transmitted from the network to the user, which, in this case, is the so-called downlink, or the data are transmitted from the user to the network, in which case the transmission takes place on the so-called uplink. In this connection, there are channels being provided for either direction or for one direction only.

20 A similar differentiation with regard to the available channels in view of the transmission direction and the common or exclusive use of the transmission resources is likewise made between the transport channels, which are defined between the MAC layer and the physical layer. In this respect, a difference is
25 made between the exclusively assigned channels, the so-called DCH (Dedicated Channels) and the common channels, the so-called CCH. The DCH channels are used on both, the uplink and the downlink. A DCH is exclusively assigned to a user for a transmission, and the stability of said channel is permanently controlled during a transmission so as to guarantee a quality of the
30 radio transmission as constant as possible. This means, an exclusive assignment of the radio resources to a user takes place during a transmission.

In contrast thereto, the radio resources are shared between the different users as far as the CCH channels are concerned.

In the following, the different commonly used transport channels, which are
5 relevant for the present invention, will be introduced.

The so-called RACH (Random Access Channel) and the so-called CPCH (Common Packet Channel) are available on the uplink. The RACH is responsible for the initiating network access including the connection set-up on
10 the physical layer associated therewith. In addition, it is provided for supporting the transmission of the user data. Said task is likewise fulfilled by the CPCH.

On the downlink, the so-called FACH (Forward Access Channel) is responsible for the transport of both, the control data and the user data.

15 By means of the so-called DSCH (Downlink Shared Channel), the user data may additionally be transmitted to an exclusively dedicated channel, the DCH. A DSCH is proportionally used between the DCHs, for which it has been assigned.

20 The RRM's task, which is located in an RNC node, consists, inter alia, in the configuration of the radio resources in response to the characteristic and quality of the requested service. This means that the decision on a channel switch or on the channel assignment in a connection set-up is taken by the RRM. Said decisions relate to both, namely the transmission on the uplink as well as on the
25 downlink, and they serve the regulation of an efficient compensation between the available transmission capacity on the transmission links and the requests made by the users using the resources of the transmission links.

By means of different algorithms, which need not be discussed here in more
30 detail, the RRM determines as to when and in which form the channel switch has to take place. When applying said algorithms, for example, the performance budget of the mobile station and of the cell are taken into account. By using the

defined threshold values the RRM takes a decision on the channel switch. In this respect, the algorithms comprise a corresponding hysteresis, which avoids switching back and forth too frequently. The taken decision is reflected in the form of parameters, which, among others, are forwarded to the RLC layer. The
5 parameters in question are the transport formats such as the data rate, and the used algorithms such as the method for a retransmission of data packets.

In the following, the function of an MAC layer will be dealt with.

10 The MAC layer is responsible for the distribution of data packets received from the RLC layer to the physical connection by taking into account the permitted transmission rate. Said distribution will hereinafter be referred to as the so-called scheduling, and the function performing the same on the MAC layer will be referred to as MAC scheduler. The MAC layer, moreover, receives the
15 transmission parameters from the RRM. The transmission parameters are, among others, the so-called TFS (Transport Format Set), the so-called TFCS (Transport Format Combination Set), the priorities of the logical transmission channels and the mapping of the logical transmission channels onto the transport channels. The distribution of the data packets to the physical
20 connection is performed on the basis of said information.

In the following, the term TFS will briefly be dealt with more closely. On the MAC layer, one or more transport formats in the form of transport format set, the so-called TFS, is/are assigned to each transport channel. A transport format
25 contains, for instance, information on the channel coding type, the depth of the interleaving or the transmission rate for the data, or, more precisely, the number of the data which may be transmitted on the radio layer in a frame. For example, a TF set of the form (12,8,4,2,1) indicates that the transport channel, to which said set is assigned, can transmit either 12 or 8 or 4 or 2 or 1 data packet of the
30 RLC layer in a frame on the radio layer. This means, on one hand, that the MAC layer has the information on the number of the data packets which can be sent in a TTI, whereby the TTI (Transmission Time Interval) designates the time,

which was defined for the transmission of a frame. On the other hand, the MAC layer receives from the RLC layer the information on the data packets being available for the transmission, which are stored in an RLC buffer. The availability of both information causes the MAC layer to perform the scheduling
5 algorithm.

In the UMTS it is possible to perform the channel switch in different ways. It is provided to change the transmission of the user data from a DCH to another DCH. In this connection it may occur that the transmission rate and/or the TTI of
10 the newly assigned DCH differs from the DCH that was used last. Another possibility is to switch the transmission of the user data from a DCH to a CCH channel, whereby it has to be considered that the entire transmission rate of a CCH may be higher than that of a DCH. The user does not have, however, an exclusive possibility to use the channel, since said channel is shared with other
15 user with the result that the effectively used transmission capacity can be smaller than the transmission capacity of the previously used DCH had guaranteed. A sharing of the CCH between several users is specified by the scheduling algorithm used.

20 As was already described, the channel switch may result in performance losses of the higher layers, e.g. the TCP layer, especially if the switch from a channel having a higher transmission rate to a channel having a lower transmission rate is performed abruptly. If the transmission rate is low, fewer packets are transmitted, which entails a long time interval between the transmission of a
25 packet and the receipt of an acknowledgement message. This, again, results in that the time for the release of the packet from a higher protocol layer, e.g. the TCP, turns out to be long for a transmission.

The solution according to the invention suggests to more efficiently utilize the
30 information on the channel switch, e.g. the time of the channel switch and the transmission parameters. In this respect, the transmission parameters of a new channel may have a relative or an absolute value over an old channel. On the

basis of said information the data flow may be adapted to the properties of the new channel. If this is conferred to the introduced embodiment, the MAC layer is communicated the information on the channel switch by the RRM. On the basis of the transmission parameters, e.g. on the basis of the transmission rate
5 of the old channel, in other words, the previously used channel, and of the newly assigned channel, the number of data packets to be transmitted via the physical connection is defined. The data are transmitted into the TTI. The time of the performance of a channel switch is used for applying a corresponding method. The solution according to the invention makes a difference between
10 two embodiments, namely in response to the times at which the information flow takes place. If the MAC layer has the information on a forthcoming channel switch, the transmission parameters of the old channel are adapted to the expected packet delay time of the new channel. In the following, the term pre-scheduling will be used for this event. Another embodiment deals with the
15 event, when the information on the channel switch is utilized only after the performance thereof through the MAC layer. In this connection, the term post-scheduling is used.

The invention will hereinafter be explained in more detail by means of figures 3a
20 and 3b.

Figure 3a schematically illustrates the transmission rate of two channels. The left-hand square shows the high transmission rate of the old channel, and the left-hand square shows a new channel having a low transmission rate. The free
25 section between the squares indicates a channel switch during which no data can be transmitted, but an exchange of the signaling messages takes place. The course in terms of time is illustrated by a time axis T provided underneath the figure. The channel switch takes place between time t (110) and t (200).

30 Figure 3b illustrates a pre-scheduling. The striped part of the left-hand square points to the unused transmission rate of the old channel so as to thereby adapt the transmission rate of the old channel to the transmission rate of the new

channel. This is to say that as of time t (101) fewer data packets are transmitted. The data rate is reduced until time t (100) until the same is adapted to the transmission rate of the new channel. By means of the pre-scheduling, the number of the data packets passed on to the transport channel are

5 controlled. In the case where the transmission rate of the newly assigned channel is lower than the data rate of the presently used channel, the number of the data packets transmitted per time unit, e.g. TTI, is reduced. If, for example, a channel switch from a channel having the transmission rate of 384kbit/sec to one merely having 64kbit/sec for the transmission has to be performed, a slow

10 adaptation of the higher transmission rate to the lower transmission rate is initiated in accordance with the invention. In other words, the data are, for example, transmitted at 256kbit/sec at first, then at 128kbit/sec and eventually at 64kbit/sec. This, again, means, although the MAC scheduler could transmit more data packets, it provides fewer data packets for the transmission. In this

15 connection, the provision of the data packets constitutes a withdrawal of the data packets from an RLC buffer. A slower withdrawal of the data packets from the RLC buffer influences the TCP, as the fact that fewer RLC data packets are transmitted results in that the acknowledgement message for a TCP packet is generated later on the receiver's side, thereby also arriving later on the sender's

20 side. This entails that the RTT measurements, i.e. the time between the transmission of a TCP data packet and the receipt of an acknowledgement for the receipt of said data packet has a greater value. In this manner the TCP learns about the reduction in the transmission rate. The increased RTT especially effects an increase of the RTO. In this embodiment, the designation

25 "presently used channel" corresponds to the above-chosen designation "old channel". Said wording corresponds to the event of the pre-scheduling in a more favorable manner.

For the adaptation of the transmissions different methods may be applied,

30 which, for instance, reduce the transmission capacity differently fast. In case of said methods attention has to be paid that, on one hand, not too much transmission capacity remains unused and that, on the other hand, however, an

abrupt change in the transmission parameters is avoided so as to prevent the expiration of the timeout.

In the case where the MAC layer receives information on the forthcoming
5 channel switch, it can control the transmission in accordance with the invention, and slowly initiate the adaptation in that the MAC scheduler does not completely exploit the transmission rate of the used channel, as would be possible, but by slowly adapting the present transmission parameters to the transmission
10 parameters of the newly assigned channel before the channel switch takes place. This is achieved in that fewer data packets than would basically be possible are provided for the transmission. The information on the number of data packets, which can basically be transmitted, is – as was already described – contained in the TFS. With said steps, the entire transmission of the TCP is slowly adapted at the same time. This method requires that the transmission
15 parameters of the new channel are known in advance. Said information is, however, available to the RRM, the task of which it is to communicate said information to the MAC layer, so that the required steps are performed. It is the task of the MAC layer to store the corresponding transmission parameters. Due to the fact that the MAC layer receives said information before the channel
20 switch takes place, it has both, namely the transmission parameters of the presently used channel and those of the newly assigned channel. On the basis of said information, the slow adaptation of the transmissions prior to the performance of the channel switch is obtained.

25 Due to the fact that the packet delay time of the presently used channel is artificially manipulated, especially increased, the artificial generation of a longer packet delay takes place on the TCP layer by step-wise measuring longer RTTs. The latter are likewise step by step considered when calculating the RTO resulting in the avoidance of spurious timeouts. As a consequence, the
30 superfluous transmission of packets is avoided, where a packet loss is recognized erroneously, and the performance of mechanisms started upon the

recognition of a data packet loss is avoided. In this manner, transmission resources and time are saved.

In some situations the channel switch has to be performed immediately, so that the time to carry out a pre-scheduling is not sufficient. In this case it is

5 suggested by the invention to influence the scheduling of the data packets on the MAC layer after the performance of the channel switch. If the channel rate of the new channel is smaller, the abrupt reduction of the transmission rate must be compensated by transmitting more data in addition to the proportionally permitted transmission rate of the new channel. This can be inferred from
10 figures 3c and 3d.

Figure 3c shows an event, where more data are transmitted by means of an additional channel. The additional channel, or respectively the used capacity of the additional channel, is illustrated by the triangle 300. Said channel is used
15 until the transmission rate of the old channel is adapted to the transmission rate of the new channel, which happens at a time t (201). In other words, an adaptation of the TCP also takes place between times t (200) and t (201). This may, for example, extend to several TCP data packets. With said method it is required to store the transmission parameters of the old channel. During the
20 post-scheduling the MAC layer performs the channel switch, and only thereafter will the adaptation be initiated. In other words, the newly assigned channel is already being used, and the MAC layer, therefore, has the transmission parameters of said channel. The transmission parameters of the old channel, i.e. of the channel used prior to the performance of the channel switch, were,
25 however, stored. For this reason, the MAC layer has the parameters of the old and the new channel, and an adaptation of the transmissions is carried out on the basis of said information.

The forms of realization of the post-scheduling vary between the uplink and the
30 downlink as well as between the switching to a DCH or to a CCH. UMTS provides for the so-called DSCH channels. As was already described, the latter

relate to channels, which can proportionally be used in addition to a DCH so as to transmit data on the downlink.

The solution according to the invention provides to temporarily use said
5 channels for the transmission on the downlink, if a channel switch takes place, so as to achieve a better adaptation of the transmission rates. In this case, the sharing of the DSCH between the users, who used the channel at that given time, has to be controlled. As an example, the priority principle with the preference of those users, with whom a channel switch has taken place, may be
10 applied in this respect.

A similar principle may likewise be applied in the case of a channel switch from a DCH to a CCH. In this case, the entire rate of the CCH channel may be larger than the transmission rate of a DCH, however, due to the fact that the CCH is
15 shared between several users, the transmission rate assigned to a user after a channel switch may be essentially smaller than the rate, which was available to the user prior to the channel switch on the DCH used last. The solution according to the invention suggests to assign the user temporarily, i.e. until the transmission has adapted to the transmission rate of the newly assigned
20 channel, a larger share of the broadband in the CCH than was actually planned for the user. Said step requires the application of a priority principle, which prefers those users, with whom a channel switch has taken place, whereby the rule of a fair treatment for all users of a CCH is violated in the favor of one user. Due to the fact that the solution according to the invention guarantees, after a
25 channel switch, the avoidance of the unnecessary performance of TCP mechanisms started upon the recognition of a packet loss, the number of the retransmissions occurring with abrupt interruptions is reduced. This results in that the transmission capacity of the CCH channel assigned to a user is used more optimally, and for this reason the entire transmission capacity of the CCH
30 is used more effectively, which eventually results in higher transmission capacities being available to all users of the CCH. Said event is illustrated in figure 3d. Between times $t(200)$ and $t(202)$, an adaptation to the capacity

assigned to a user takes place starting at time t (202). In other words, until time t (202) more capacity is made available to a user than he is actually allowed to have.

- 5 A third embodiment of the inventive solution provides for a combination of the pre- and post-scheduling, i.e. given a punctual report on a forthcoming channel switch, the pre-scheduling is performed first and the post-scheduling is applied after the performance of the channel switch.
- 10 The pre-scheduling proves to be advantageous when taking into account the time delays associated with the channel switch, namely in the determination of the data quantity to be transmitted. During a channel switch between t (100) and t (200) signaling messages are exchanged, which require time and, therefore, result in time delays, which may be taken into account during the pre-
- 15 scheduling.

For the realization of the invention it is necessary that the RRM unit signals the channel switch to the MAC unit. In UMTS it is, however, feasible that the MAC unit is located in different network nodes, preferably in RNC or in a mobile

20 station. In the following, different cases in response to the location of the MAC layer will be described as examples.

If the MAC unit and the RRM unit are located in the same network node, the invention is guaranteed to function. If the MAC unit is, for instance, located in

25 the terminal of the user, a signaling to the MAC layer is required. It is likewise required to inform the MAC layer, if the same is provided in an RNC other than that where the RMM was located. In this case, a signal must be sent via the interface between the RNCs.

- 30 Furthermore, the invention relates to a computer program stored on a computer-readable medium or in a form, which can be loaded in a working

memory of a computer, whereby the computer program controls a device according to one of claims 1 to 10.

Patent Claims

1. Method for improving a data throughput of a transmission via a communication network comprising an administration unit for
5 administering channels having transmission parameters, and a provision unit for providing, distributing and transmitting data via the channels,
 - wherein the administration unit informs the provision unit about a channel switch from an old to a new channel, and
 - wherein the information on the transmission parameters of the new
10 and the old channel as well as on a time of the performance of the channel switch is provided to the provision unit, and
 - wherein, on the basis of said information, an adaptation of the transmission parameters of the old channel to the transmission parameters of the new channel is performed.
- 15 2. Method according to claim 1, wherein the adaptation is realized by controlling the number of data packets transmitted in a time unit.
3. Method according to claim 1 or 2, wherein the adaptation of the data is
20 carried out before and/or after the performance of a channel switch.
4. Method according to one of the preceding claims, wherein, with the adaptation to a channel, whereof the transmission parameters cause a long packet delay time, the number of data packets provided for the
25 transmission is reduced prior to the performance of the channel switch.
5. Method according to one of the preceding claims, wherein, with the adaptation to a channel, whereof the transmission parameters cause a long packet delay time, more data are temporarily transmitted in addition
30 to a proportionally permitted transmission capacity after the performance of the channel switch.

6. Method according to claim 5, wherein additional data are transmitted via one or more additional channels.
7. Method according to claim 5, wherein a higher transmission capacity is temporarily provided in a channel for the transmission of more data.
8. Method according to one of the preceding claims, wherein the transmission parameters are stored for performing an adaptation.
9. Method according to one of the preceding claims, wherein the channel switch is performed on an exclusively used channel.
10. Method according to one of the preceding claims, wherein the channel switch to a channel commonly used by several users is performed.
11. Device for improving a data throughput of a transmission via a communication network comprising an administration unit for administering channels having transmission parameters, and a provision unit for providing, distributing and transmitting data via the channels, comprising
 - decision means in the administration unit for taking a decision on a channel switch, and
 - communication means for communicating the decision to the provision unit, and
 - a memory for storing the transmission parameters of an old channel and/or a new channel, and
 - adaptation means for adapting the transmission to the transmission parameters of the new channel.
12. Device according to claim 11, wherein the administration unit and the provision unit are located in a network node.

13. Device according to claim 11, wherein the administration unit and the provision unit are located in different network nodes.
- 5 14. Device according to claim 13, wherein a signaling between the provision unit and the administration unit is provided for transmitting the communication.
- 10 15. Computer program stored on a computer-readable medium or in a form, which can be loaded in a working memory of a computer, wherein the computer program controls a device according to one of claims 1 to 10.

FIG. 1

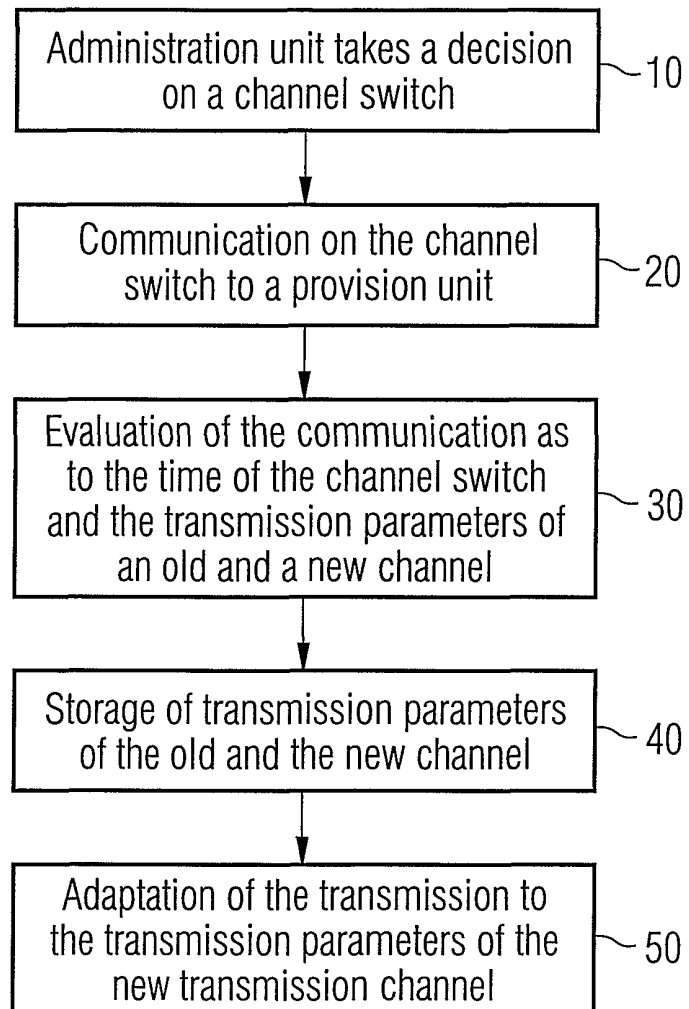


FIG. 2

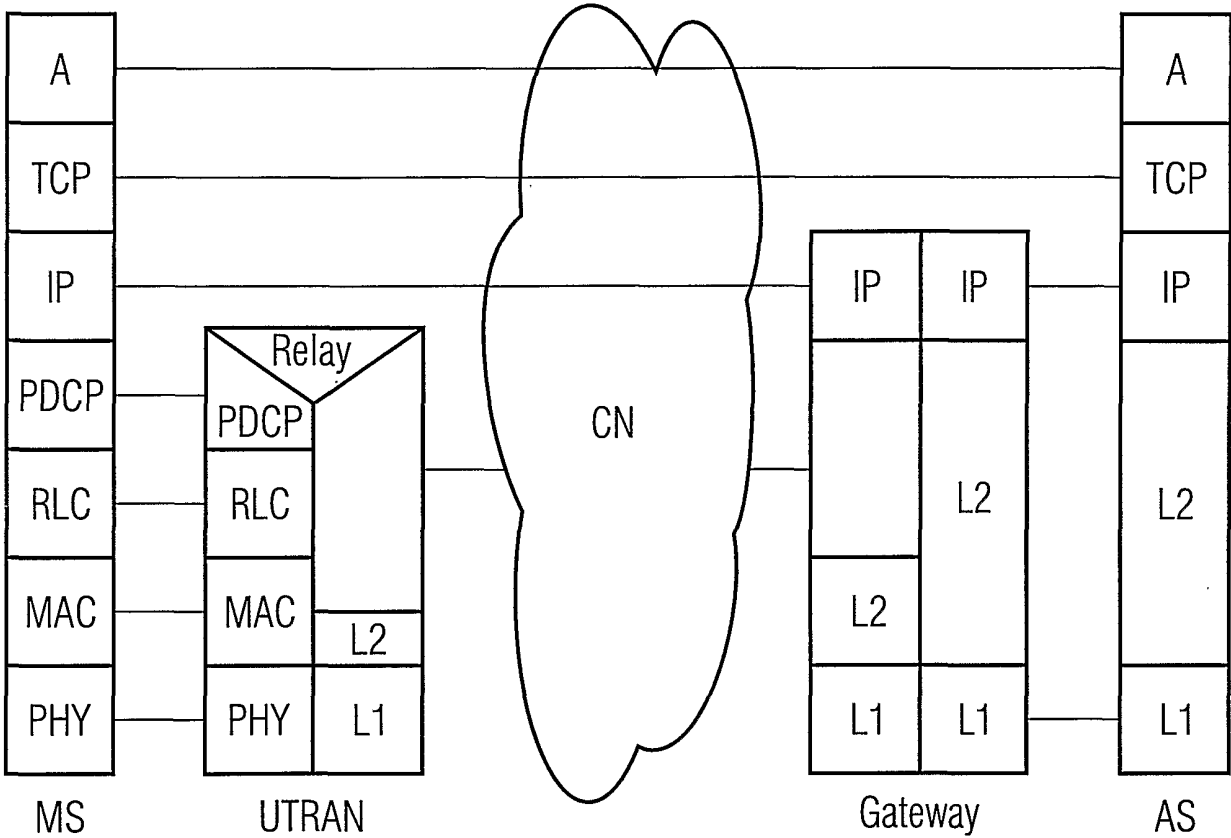


FIG. 3a

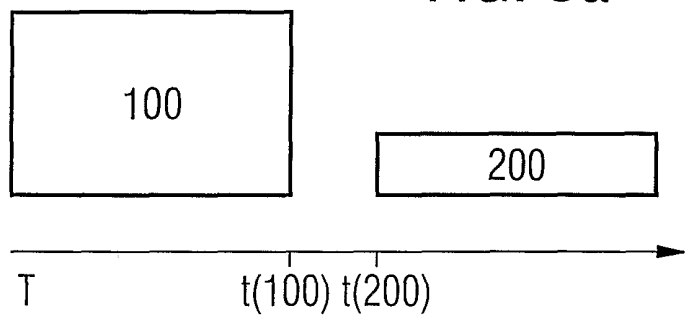


FIG. 3b

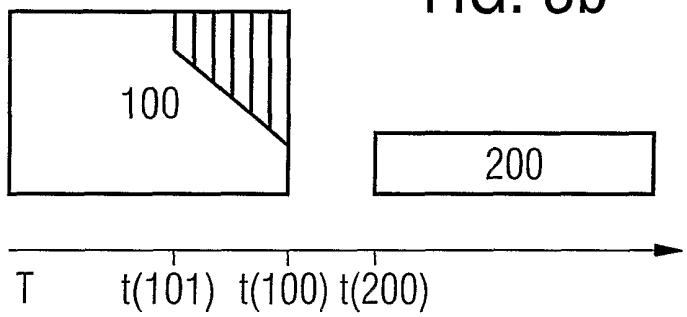


FIG. 3c

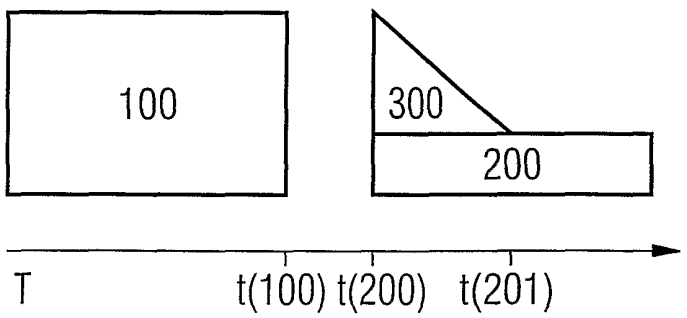
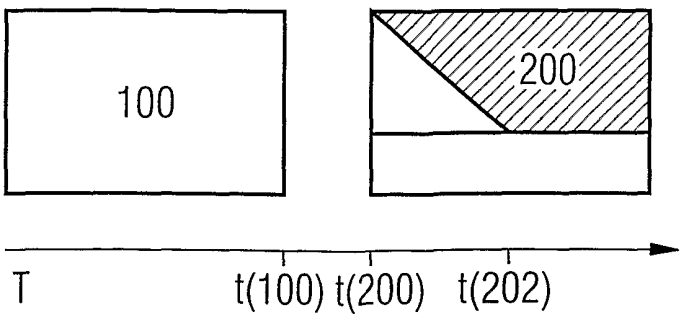


FIG. 3d



INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/02264

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04L12/56 H04Q7/38 H04L12/28

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)

IPC 7 H04L H04Q

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 practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 66748 A (ERICSSON TELEFON AB L M) 23 December 1999 (1999-12-23) page 2, line 19 - line 22 page 3, line 11 - line 26 page 19, line 4 -page 20, line 16; figures 10,11 page 11, line 6 - line 16 -----	1-15
A	US 5 920 547 A (WERTH GEORG) 6 July 1999 (1999-07-06) column 1, line 56 -column 2, line 45 -----	1-15



Further documents are listed in the continuation of box C.



Patent family members are listed in 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 document 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

16 August 2002

Date of mailing of the international search report

23/08/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Gregori, S

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 02/02264

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9966748	A	23-12-1999	US 6347091 B1	12-02-2002
			AU 1447999 A	05-01-2000
			BR 9815914 A	20-02-2001
			CN 1345518 T	17-04-2002
			EP 1088463 A1	04-04-2001
			JP 2002518958 T	25-06-2002
			WO 9966748 A1	23-12-1999
			TW 424377 B	01-03-2001
US 5920547	A	06-07-1999	DE 19629899 C1	21-08-1997
			JP 3143084 B2	07-03-2001
			JP 10145871 A	29-05-1998