

Source: Motorola

## Individual control of downlink data flow to multiple devices connected to a single UE: discussion paper

### 1 Overview of problem domain

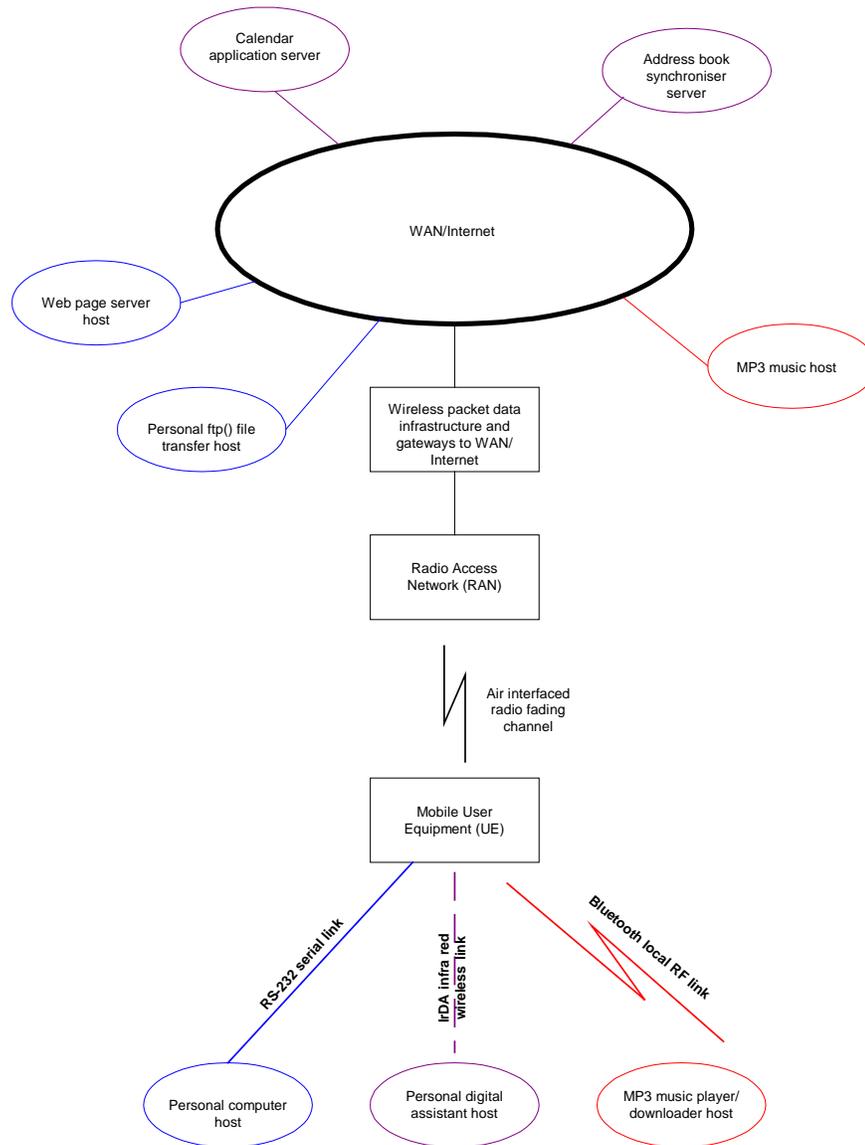
The General Packet Radio Service (GPRS) and Enhanced Data for Global Evolution (EDGE) for the Global System for Mobile Communication (GSM) have introduced the capability of the user data interchange into mobile wireless products. GPRS, and its superset, EDGE, along with their 3G counterparts, GPRS for Universal Terrestrial Radio Access (UTRA) and High Speed Downlink Packet Access (HSDPA), permit efficient use of radio and network resources when data transmission characteristics are i) packet based, ii) intermittent and non-periodic, iii) possibly frequent, with small transfers of data, e.g. less than 500 octets, or iv) possibly infrequent, with large transfers of data, e.g. more than several hundred kilobytes. User applications may include Internet browsers, electronic mail and so on, [1,2,4].

The applications that make use of the data transfer capabilities of user packet services may be greatly distributed over the user's operational environment, e.g. some may be internal to the user equipment device, such as an internal browser applications, while others may reside on a remote host, such as a personal computer (PC) or personal digital assistant (PDA).

Furthermore, the interconnection scheme employed to move data between the mobile wireless device and the host on which the application resides may vary substantially and exhibit distinctly differing characteristics from one user host to another. A user may have an electronic mail (email) application resident on a PC host connected to the mobile terminal by a physical serial data connection while a calendar application resides on a PDA "connected" to the mobile terminal using an Infra Red Data Association (IrDA) interface and its associated link control logic. Other types of mobile terminal to host interface may be utilized as well, e.g. a radio frequency (RF) wireless local link such as HomeRF or Bluetooth, Universal Serial Bus (USB), Ethernet, each of which have distinctly differing data transfer characteristics including data rates and isochronicity.

Consider an operational environment in which the mobile user equipment (UE) is treated as a "wireless data modem" of sorts having multiple host connections, each of which may be associated with one or more applications, as in **Figure 1**. Each user host (PC, PDA and MP3 music player) is logically connected to its corresponding server host via the Internet and physically connected to the UE "data modem" by different types of interfaces, via different interface methods, each having their own set of data transfer peculiarities, as shown in **Figure 2**:

- The RS232 connection is hard-wired to the PC host and operates at 115 kb/s or more, but also has the need to reflect "hardware flow control" information back to the mobile user equipment, effectively treating it like a "data modem" of sorts.
- The IrDA connection is wireless and incorporates its own link control protocol and channel adaptation methods which include the ability to select among many modulation and coding schemes depending on optical fading channel conditions, including local interference from ambient lighting. Not only does this type of connection suffer from the periodic outage caused by changes in the optical channel, but its maximum data rate may dynamically change with link adaptation. In addition, there is a price to be paid in terms of time regarding the link signalling and channel adaptation, during which periods user data are not able to flow.
- The Bluetooth connection is wireless RF, and may compete, and be subject to competition from interference caused by electromagnetic devices in the environment, including the mobile user equipment itself. It also has an integral link control protocol and is likewise subject to the local fading channel conditions.



**Figure 1: Operational environment example of mobile UE used as a "data modem" over which multiplexed data are routed between multiple clients and servers via wireless data network.**

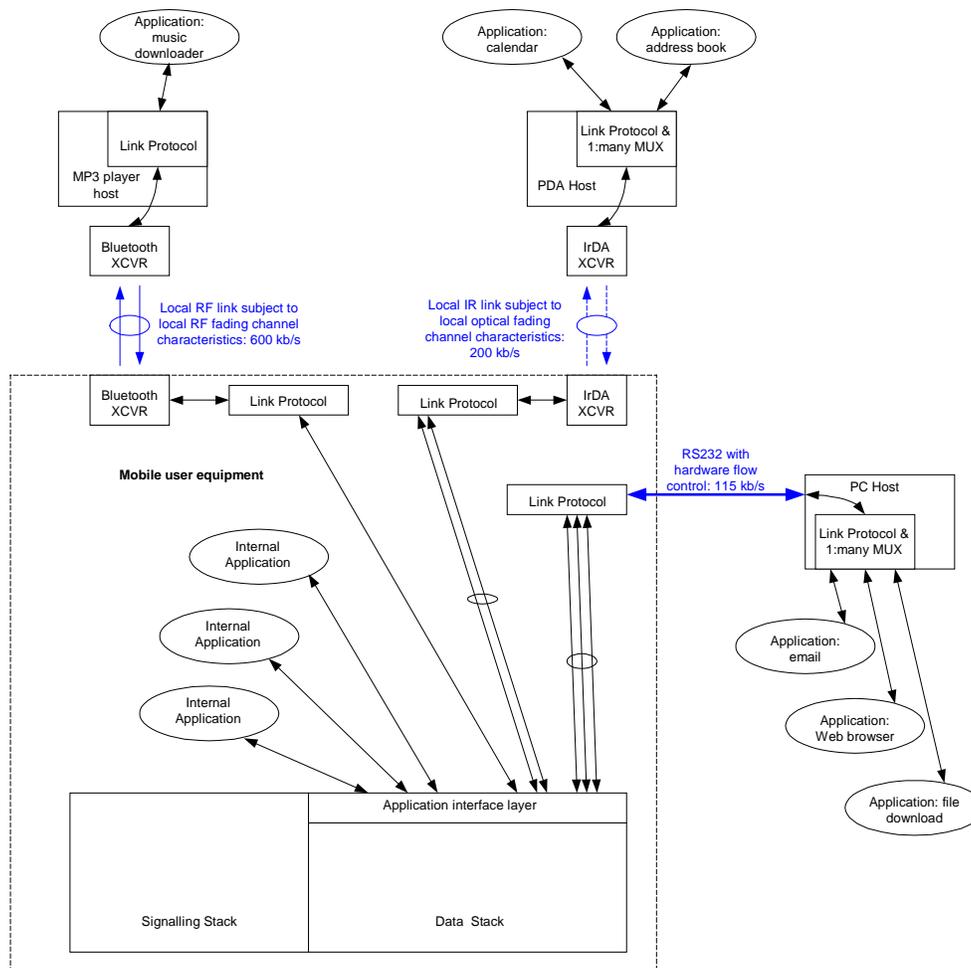
Furthermore, each user shown in **Figure 2** is associated with one or more applications, each of which may require the logical connection to one or more servers residing over the wide area network (WAN) or Internet.

## 2 Problem description

Due to i) the variance in actual data rates over multiple interface types between mobile user equipment and user hosts and ii) the requirement for mobile user terminals to support higher data rates in the near future, the mobile user equipment has no efficient method to allow it to incrementally control the flow of downlink data on an external interface-by-interface basis, running the risk of repeatedly exhausting internal memory resources.

If the mobile user equipment does not address this problem in the future, then the progressively higher downlink data transfer rates combined with the extreme variability of the external device interfaces is likely to cause the internal memory resources to be repeatedly exhausted on the mobile user equipment.

Consider the following:



**Figure 2: Example of UE treated as a "data modem" over which multiple application data streams are de-multiplexed and sent over individual heterogeneous interfaces, each having different transmission characteristics, including local optical and RF fading channels.**

1. Repeated exhaustion of memory resources results in a cascading effect of data protocol timeouts, resetting of transport protocol congestion window sizes and the initiation of controlled transmission roll-back and re-start. The repetition of these unnecessary procedures seriously impacts the downlink data throughput.
2. It is undesirable that a discontinuity of data flow to one external device impacts the data throughput or resource availability to all other external devices, and/or internal applications to which the mobile user equipment is connected. If a general type of flow control were used, then if any one device interface suffers link perturbation, then all applications served by the UE (including the internal ones, which have no device interfaces) would be impacted. It would be desirable that each set of streams associated with an interface be controlled independently from one another.
3. Each time memory resources are exhausted, downlink data are discarded, resulting in a waste of radio resources which (if some sort of fine level of flow control is used) could be utilised to i) service other applications associated with the UE or ii) other wireless users and other UE's.

While these problems manifest themselves even today in the current GPRS network environment, the progressively higher data rates being specified for future systems, including GPRS and Universal Mobile Telephone Service (UMTS), will magnify this problem in the future. Consider the downlink data transfer rates supported, or expected to be supported, by GSM 2+ Generation and 3G data services in the near future:

Service	Data Rates
GSM General Packet Radio Service (GPRS)	9.05 – 21.4 kb/s per timeslot with up to usually 4 timeslots in a single direction, resulting in composite maximum throughput of 36.2 - 85.6 kb/s, [1,5].
GSM Enhanced Data for Global Evolution (EDGE)	10.6 – 61.85 kb/s per timeslot, with up to usually 4 timeslots in a single direction, resulting in composite maximum throughput of 42.4 – 247.4 kb/s, [5].
UTRAN GPRS	Up to 384 kb/s, determined by variable spreading factor on downlink, [3].
UTRAN High Speed Downlink Packet Access (HSDPA)	Up to 3 Mb/s average and 10 Mb/s peak, determined by variable spreading factor on downlink, [4].

The fact that this problem is not addressed by the current GPRS/EDGE specifications for GSM raises the following questions:

### How is the problem being addressed in current GPRS implementations?

Today's GPRS implementations do not address this problem at all, but rather relies on the host application to provide feedback to the application or server at the other end in order to control the flow of downlink data, e.g. some applications provide repeated acknowledgements at the transport layer when operating in a "stream-oriented" mode with Transmission Control Protocol (TCP).

### Why is this solution inadequate?

The self-clocking TCP acknowledgements serve to regulate the flow somewhat, but may also cause other problems with regard to the TCP congestion window, the operation of which is designed for the wireline environment. Any delay in acknowledgement beyond a certain reasonable amount makes the sending TCP think that there is congestion on the network, i.e. that the routers between the two hosts have run out of packet queues. TCP then resets its window size to 1 stops sending and waits a pseudo-random time period to enable the router queues to flush. This may severely impact data flow, e.g. by more possibly than one and at times 2 orders of magnitude, [7].

In the case of an unacknowledged transport layer protocol, such as the User Datagram Protocol (UDP), the mobile user equipment would not "clock" the sender at all, and any excess data beyond the available resources of the mobile user equipment are simply thrown away, thereby wasting radio resources that may have been used for other purposes such as the interchange of data for other applications served by the same UE.

## 3 General requirements for solution

Due to i) the variance in actual data rates over multiple interface types between mobile user equipment and user hosts and ii) the requirement for mobile user terminals to support higher data rates in the near future, any flow control mechanism between the mobile and wireless data network should meet the following general requirements:

- Should enable the instantaneous control over i) multiple, logical data streams or sets of streams, each of which may be associated with a specific interface between the mobile user equipment and a local user host or ii) all downlink user data sent to a particular UE.
- Fine control is required over downlink data arriving from the network, so that buffering resources on the mobile user equipment side may be conserved and protected against overflow. This reasonably rapid reaction time suggests that the mechanism may be logically situated near the radio interface. Such control would prevent data from being thrown away should buffering resources be all consumed in attempts to buffer large bursts of downlink data at peak transfer rates and also enable the mobile manufacturer to conserve memory resources by operating close to the lower limit of available memory when necessary. The operator would benefit from an increase in downlink capacity, since

all downlink data would in effect be “wanted” data and not intentionally discarded due to lack of mobile resources. This aspect is not treated by existing system simulation tools, i.e. all data arriving at a specific UE is treated as “application-usable”.

- The signalling information sent from the UE to the RAN must be as economical as possible in terms of radio resource usage.

## 4 General concept direction

The foregoing requirements for a solution to the problem of handling multiple user devices on a single UE suggests the use of the following techniques:

- 1) Rapid, fine control suggests the direct control of radio bearers at, or logically near, the physical layer.
- 2) Control over multiple logical data streams, one or more of which may be associated with a single external device interface or internal application, requires that information identifying such device interface or internal application be in some manner logically associated to a radio bearer. At present, this is not the case, as there is no reason for the radio bearer to have knowledge of the type of information it is carrying, its ultimate destination or how it is to be used.
- 3) In order for the UE to inform the Node-B when to enable or disable downlink data flow, and specifically what stream to enable or disable, a suitable uplink signalling method must be developed.

### 4.1 Architectural direction

Consider the following simplified reference architecture that illustrates how the various components are logically situated in the UE and network in Figure 3. The components and flows in blue are logical representations of new elements. A brief explanation of each component’s purpose and operation follows:

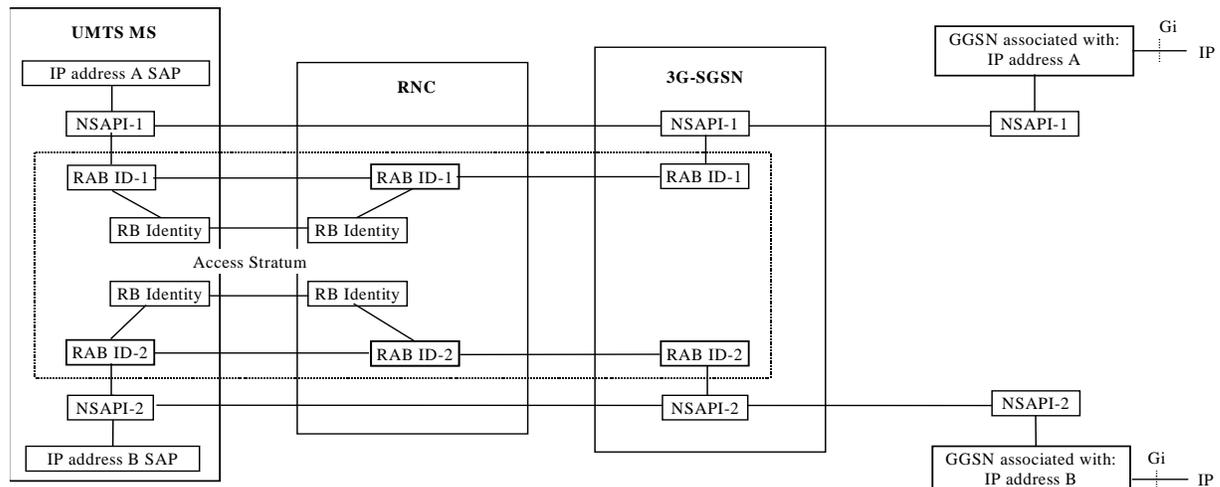
**PDP Context/Radio Bearer Identity (RB-ID) Logical Binding (Mobile/Network):** This component associates the PDP Context with the identifier by which the Node B and/or Radio Network Controller (RNC) would understand it, i.e. by its RB-ID. For simplicity, this diagram shows information going in a single direction in the mobile station to components requiring such information. In an actual mobile user equipment implementation, this component may simply be an information store and associated logic to maintain the logical binding between PDP Context and RB-ID, which could then be queried, or accessed by any interface processing component that requires it. In the UTRAN/GERAN environment, the PDP Context identification is bound to the Radio Bearer Identity, which indirectly identifies both the network service access point identifier (NSAPI) and the radio bearer associated with the data identified by the PDP Context, [1].

Refer to the following diagram from [1] that demonstrates how the identities resident at higher network layers are indirectly associated, and therefore available to the radio layers over which control is desired.

Using this association, the mobile user equipment associates the data stream information associated with a certain PDP Context with its corresponding Radio Bearer Identity (RB-ID) information.

The RB-ID is mapped to its corresponding position in a flow control bit map sent to the network. Standardised knowledge of how such a bit map associates each RB-ID to each PDP Context allows the network to apply fine flow control, starting and stopping data flow associated with each PDP Context in the radio layers without the necessity of signalling back over multiple interfaces to the upper network elements, such as the Serving GPRS Support Node (SGSN) or Gateway GPRS Support Node (GGSN).

Note that while the same basic concept may be used for both i) UTRAN/GERAN 3G and ii) GSM GPRS/EDGE Phase 2+ system environments, the concept of a PDP Context identifier is bound to different parameter sets in each environment. Therefore, the PDP Context is associated with the “Packet Flow Identifier”, as this serves an analogous function in the GSM GPRS/EDGE environment as does the RB-ID in the UTRAN environment.



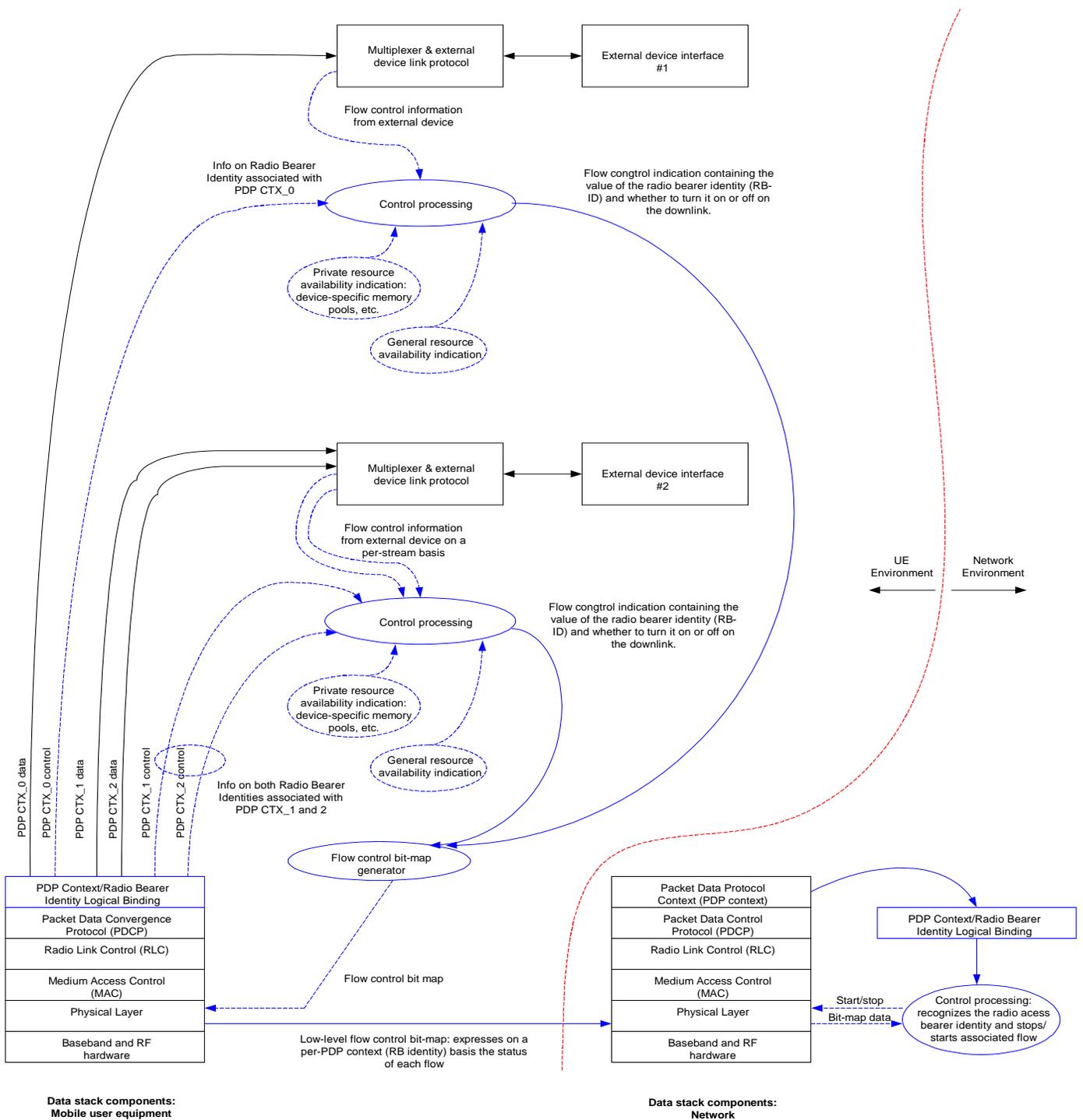
**General Resource Availability Indication (Mobile):** This represents an indication that system memory (or process creation) resources have been exhausted and reached a logical “low-water mark”.

**Private Resource Availability Indication (Mobile):** This represents an indication that some private resources, such as a per-PDP Context or per-external interface memory pool, have been exhausted and reached a logical “low-water mark”.

**Control Processing (Mobile):** This component would have two primary sets of inputs: i) the association of PDP Context to RB-ID and ii) all status information that would cause the mobile station to either stop or start the flow of information from the network. Status information includes the availability of i) both private and ii) system resources, flow control information forwarded from the external device interface (such as IrDA, Bluetooth, or RS-232 serial hardware flow control information).

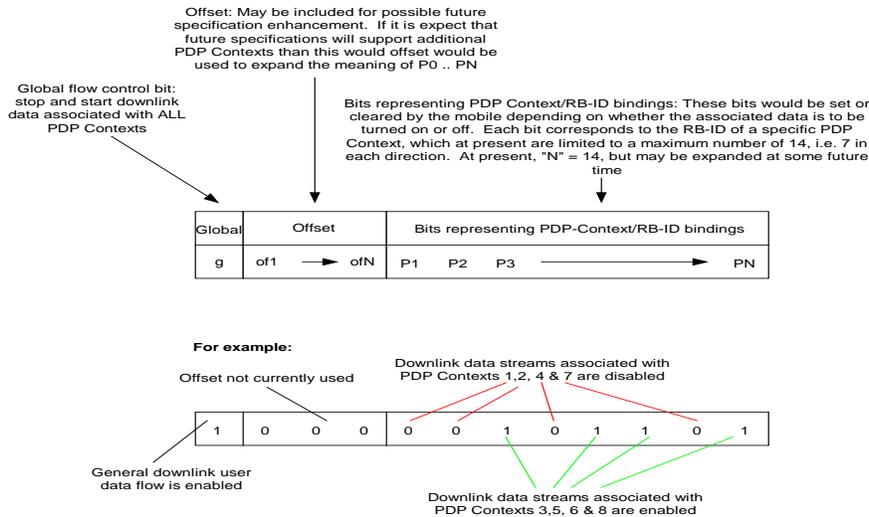
**Control Processing (Network):** This component represents the logical entity that would i) interpret the bit-map values arriving from the mobile and ii) enable or disable the downlink data flow to the mobile based on their position (to what RB-ID they refer) and their value (Boolean: off/on).

**Flow Control Bit-Map Generator (Mobile):** This component is responsible for creating a bit-map that properly reflects the mobile’s control processing intentions and making available for transmission to the network.



**Figure 3: Conceptual reference architecture as represented in UTRAN/GERAN environment (flows and components in blue represent new concepts)**

**Low Level Flow Control Bit-Map (Information Element):** This requires definition and standardisation. In UTRAN, it may be sent on an uplink dedicated control channel over which other control data are multiplexed. In GERAN or GSM GPRS/EDGE, it may be sent over a Packet Data Channel (PDCH) as a “Control Block” in the uplink direction. Conceptually, the bit map may be represented as follows:



### 4.2 Issues regarding uplink signalling

While the concept of a “flow control bit-map” conveniently illustrates our intention of individual radio access bearer control, it presents some practical challenges in, e.g. the development of an acceptable uplink signalling method for HSDPA.

Problems related to uplink signalling methods for HSDPA were discussed in detail at RAN1 #20 in Busan, Korea during May, 2001, [8, 9, 10, 11]. Of these contributions, the Qualcomm [8] and Ericsson [11] papers both discuss the time division multiplexing (TDM) and code division multiplexing (CDM) methods for adding uplink signalling to the dedicated physical channel (DPCH) that is assumed to exist in association with the high-speed downlink shared channel (HS-DSCH). Sony [9] points out that the rate of uplink carrier-to-interference (C/I) measurements transmitted on the uplink channel may be effectively slowed down to a reasonable rate so as to minimize impact to capacity, while Panasonic [10] disagrees with the concept of C/I reporting by the UE in general.

Both Qualcomm and Ericsson converge on CDM as a preferred solution. Ericsson considers the possibility of transmitting only the automatic retransmission query (ARQ) acknowledgement (ACK) field, emphasizing that it is only transmitted when the UE is actually addressed by the network on the HS-DSCH, and suggests a possibility for a transmission duration of a single timeslot in order to reduce delay, even though the actual transmission time interval (TTI) is 3 timeslots. Qualcomm meanwhile suggests the transmission of the ACK/NAK bit using a length-256 code, recognizing that a length-512 code is also possible, which the inclusion of the C/I measurement report on a second length-256 code. Both of these additional codes would be transmitted on the Q carrier phase reference.

If we were to develop an uplink signalling method for multiple stream flow control using the CDM method as proposed by Ericsson and Qualcomm, then we must consider i) the effective information rate required, ii) the required probability of error and iii) the amount of uplink capacity consumed.

Consider a flow control bit-map comprising 15 bits total, i) one for “global” control of all downlink user data to a single mobile and ii) 14 for the current maximum number of PDP contexts (and therefore downlink streams) permitted. If this entire bit map were to be sent once per TTI, at a rate of 500 times per second, then each mobile would require consume ( 14 bits \* 500 times/sec ) = 7000 bits/sec. If there were 8 mobiles addressed on a TTI, then we would require ( 7000 bits/sec/mobile \* 8 mobiles ) = 56,000 bits/sec additional uplink capacity for flow control signalling. While this approach would ensure that the Node-B would receive continuous flow control basis updates, it may be prohibitive to consume this amount of uplink capacity in a real system.

Now, consider an approach in which the UE, rather than sending periodic flow control bit-map updates every TTI, would send a flow control status update aperiodically, i.e. only when the flow control state of any flow would change. Even if the entire bit-map (15 bits in our example) were to be sent each time any of its values changed, the amount of actual uplink capacity required may be significantly less than the periodic update case, due to the probability that all mobiles requiring a bit-map update at the same time is likely to be fairly low in a real operational environment. Further optimizations may include the transmission of fewer bits of information, as in our example the UE would have 14 flows at most associated with it.

A problem introduced by this second approach is that it introduces an additional point of failure, viz. the problem of what happens when the network fails to properly decode the flow control information on the uplink. If this happens on a “on-to-off” transition, then the UE would be subjected to the reception of more information than it may reasonably handle, and the problem of having to discard downlink data then may resurface, along with its associated undesirable effects. Conversely, if the “off-to-on” transition is not properly decoded, then the UE would not be receiving its downlink user data when it was expected, thereby negatively impacting the throughput to an individual application or device.

Since the first approach, in which the flow control information is sent periodically, may be viewed as a logical form of repetition encoding, we may consider simply using bi-orthogonal encoding for the flow control information. If we have N flows to control, e.g. 14 plus a global one = 15 total, then we would associate each flow with one of N orthogonal words of length-N. Then, since the data on the uplink are coherently demodulated, we would binary modulate these words with the associated flow control bit value (either 1 or -1). A Walsh function may be used to permit the application of the efficient Fast Hadamard Transform (FHT) at the Node-B for demodulation.

The resulting signalling stream could then be i) repetition code and ii) spread using a specific length-256 code on the Q-channel of the uplink DPCCCH in a manner identical to the ACK/NACK and C/I reporting fields. This repetition code would be devised in such a manner as to ensure a reasonable degree of maximum bit-error probability (BEP). If the maximum BEP for the flow control information on the uplink were e.g.  $1e-3$ , and the network were to fail to decode the first one, the UE could re-transmit the flow control bit-map state update on the next TTI if it is still receiving information on the downlink. The probability of the Node-B missing two in a row then becomes the “coin-flipping probability problem”, or  $(10^{-3} * 10^{-3}) =$  a probability of  $10^{-6}$ . This may represent a fairly reasonable rate at which the undesirable effects of losing flow control effectiveness could be tolerated, although further study would be required to substantiate this hypothesis.

If this concept were to be re-used in the GSM GPRS/EDGE environment, then uplink signalling would present a different set of issues.

## 5 Standardisation issues

Standardisation of the described multiple stream flow control method requires work in many domains. It is also recognized that there are some aspects of the proposed solution that require standardisation and some that do not. The following summarizes the basic issues:

### What requires standardisation?

- 1) **The methods of binding a PDP-context and a Radio Access Bearer to an application data stream and/or external device interface.** Procedures need to be created to define what happens in the network and UE when a PDP-context is i) activated and ii) deactivated. The reason that all three data need to be associated (PDP-context, Radio Access Bearer ID (RAB-ID), Device interface) is because at different points in the operation of an actual system, it is expected that the PDP-context would be known, but the RAB-ID would be wanted, or the Device Interface-ID would be known and the RAB-ID would be wanted.

- 2) **Channel and/or data definition for uplink basic physical signalling information from UE to Network.** As previously discussed, the specification of the means by which these data are sent from the UE to the network is still fairly fluid for HSDPA. For such a flow control method to be effective in the GSM GERAN environment, it is expected that we would have to use existing physical channel structures.
- 3) **Signalling aspects.** The organization of flow control information and how it is to be transmitted from the UE and received by the network requires specification.

#### What does not require standardisation?

- a) **How the flow-control mechanism shall be used.** Although we suggest the solution of multiple-stream flow control for the purpose of reducing the undesirable effects relating to multiple external devices interfaces, the use of the underlying mechanism shall not be limited to solving this problem. If in the future other issues are discovered for which this mechanism provides a useful solution, then manufacturers should be allowed to make of this mechanism for other than its original intended purpose.
- b) **Control processing by the UE and/or Network.** This includes the manner in which the UE treats its resource boundaries or other conditions that would cause it to exercise flow control on the downlink, e.g. the algorithm or intelligence that is used to determine when to stop or start downlink user data. This includes specifying resources such as memory (and associated "high/low watermarks"), processing resources, detection of external interface perturbations. It also includes the manner in which downlink data are queued by the network.

The preceding issues, a and b, are implementation specific, and therefore their standardisation is not needed. It is expected that manufacturers would act reasonably in their implementations to optimize their usage of flow control. Any additional specification in the cited areas would likely serve to limit the latitude of manufacturers' implementation decisions.

In summary, it is our objective to standardise the multiple-stream flow control mechanism and its operation, and not the manner in which it would be used.

## 6 Summary

This brief concept proposal is intended to i) identify a fairly serious problem as the trend toward higher-speed mobile user data continues and ii) make suggestions as to the requirements for its solution, as well as to highlight some of the challenges expected. Due to the unique aspects of the use of mobile user equipment as a "data modem" of sorts serving multiple external user applications, possibly transmitting user data over additional optical and local RF fading channels at different rates, this problem is unusual in that it is not confined to a single domain, "layer" or component.

The fact that the problem domain itself spans multiple sub-domains suggests that a fair amount of additional cooperation between multiple research and standardisation groups is needed to arrive at an optimal solution. We welcome the input and suggestions from other manufacturers and network operators to help us establish a constructive way forward in addressing this issue.

## 7 References

[1], GSM-02.60, "Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service Description; Stage 1", (European Telecommunications Standards Institute, (ETSI) Global System for Mobile Communications (GSM) specifications).

[2], GSM-03.60, "Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service Description; Stage 2", (European Telecommunications Standards Institute, (ETSI) Global System for Mobile Communications (GSM) specifications).

- [3], 3GPP 25.301, "3<sup>rd</sup> Generation Partnership Project; Technical Specification Group Radio Access Network; Radio Interface Protocol Architecture", (3<sup>rd</sup> Generation Partnership Project (3GPP); Technical Specification (TS)).
- [4], 3GPP 25.848, "3<sup>rd</sup> Generation Partnership Project; Technical Specification Group Radio Access Network; High Speed Downlink Packet Access (HSDPA)", (3<sup>rd</sup> Generation Partnership Project (3GPP); Technical Report (TR)).
- [5], GSM-05.01, "Digital cellular telecommunications system (Phase 2+); Physical Layer on the Radio Path; General Description", (European Telecommunications Standards Institute, (ETSI) Global System for Mobile Communications (GSM) specifications).
- [6], 3GPP 23.060, "3<sup>rd</sup> Generation Partnership Project; Technical Specification Group Services and Systems Aspects: General Packet Radio Services (GPRS); Service Description; Stage 2", (3<sup>rd</sup> Generation Partnership Project (3GPP); Technical Specification (TS)).
- [7], "TCP performance issues over wireless links", (G. Xylomenos, G. Polyzos; Athens University of Economics and Business, Greece; Petri Mahonen, Mika Saaranen; University of Oulu, Centre for Wireless Communications, Finland; *IEEE Communications*, April, 2001).
- [8], "UL structure in support for HS-PDSCH", (Qualcomm; Temporary document (Tdoc) #477; Technical Standards Group (TSG) Radio Access Network (RAN) working group 1 #20; Busan, Korea; May 21<sup>st</sup> – 25<sup>th</sup>, 2001).
- [9], "Follow up on variable C/I feedback rate proposal", (Sony Corporation; Temporary document (Tdoc) #512; Technical Standards Group (TSG) Radio Access Network (RAN) working group 1 #20; Busan, Korea; May 21<sup>st</sup> – 25<sup>th</sup>, 2001).
- [10], "On the MCS selection for HS-DSCH", (Panasonic; Temporary document (Tdoc) #534; Technical Standards Group (TSG) Radio Access Network (RAN) working group 1 #20; Busan, Korea; May 21<sup>st</sup> – 25<sup>th</sup>, 2001).
- [11], "Uplink signalling for Hybrid ARQ", (Ericsson; Temporary document (Tdoc) #571; Technical Standards Group (TSG) Radio Access Network (RAN) working group 1 #20; Busan, Korea; May 21<sup>st</sup> – 25<sup>th</sup>, 2001).