Flat Cellular (UMTS) Networks

Peter Bosch, Louis Samuel, Sape Mullender, Paul Polakos, Gee Rittenhouse

Bell Laboratories

Abstract— Traditionally, cellular systems have been built in a hierarchical manner: many specialized cellular access network elements that collectively form a hierarchical cellular system. When 2G and later 3G systems were designed there was a good reason to make system hierarchical: from a cost-perspective it was better to concentrate traffic and to share the cost of processing equipment over a large set of users while keeping the base stations relatively cheap. However, we believe the economic reasons for designing cellular systems in a hierarchical manner have disappeared: in fact, hierarchical architectures hinder future efficient deployments.

In this paper, we argue for completely *flat* cellular wireless systems, which need just one type of specialized network element to provide *Radio Access Network* (RAN) functionality, supplemented by standard IP-based network elements to form a cellular network.

While the reason for building a cellular system in a hierarchical fashion has disappeared, there are other good reasons to make the system architecture flat: (1) as wireless transmission techniques evolve into *Hybrid ARQ* systems, there is less need for a hierarchical cellular system to support spatial diversity; (2) we foresee that future cellular networks are part of the Internet, while hierarchical systems typically use interfaces between network elements that are specific to cellular standards or proprietary. At best such systems use IP as a transport medium, not as a core component; (3) a flat cellular system can be self scaling while a hierarchical system has inherent scaling issues; (4) moving all access technologies to the edge of the network enables ease of converging access technologies into a common packet core; and (5) using an IP common core makes the cellular network part of the Internet.

I. INTRODUCTION

We argue that future cellular networks must be *flat*: A *Radio Access Network* (RAN) must be confined to a single network element type, while all other components in the cellular back-haul network are derived from off-the-shelf IP components. By replacing the complex and specialized network elements typically found in a cellular systems by standard IP-based network elements, (future) cellular networks become more cost-effective and perform better than current, hierarchical (3G) cellular systems.

We define a flat cellular system as a system where Layer 1, Layer 2 and Layer 3 coincide in the base station we call this base station an *integrated base station*. Layer 1 is concerned with the transmission and reception of data over the wireless channel, Layer 2 carries out link-layer fragmentation/reassembly, error-detection and retransmission and Layer 3 deals with the IP-layer or wireless voice circuits. Further, while this paper focuses primarily on issues of integrated base stations, we envision a flat system to be completed with an Authentication Center (AuC) or Home Subscriber System (HSS) for authentication and other issues related to security and an IP anchor (e.g. a Mobile IP Home Agent) for helping with IP back-haul routing. In the discussion that follows, we use the UMTS architecture as the example system; in fact, we have already built a completely flat UMTS system [10], but the arguments are equally applicable to other cellular standard.

By integrating Layers 1, 2 and 3 in an integrated base station, we achieve two vital optimizations. At a micro-level, flat systems reduce transmission delays. Vertical protocol layer integration and application aware scheduling make the system respond better to application demands.

At a macro-level, integrating all protocol processing in base stations reduces the overall complexity of the wireless system when compared to current 3G systems. It is essential to observe that a flat system's base station can operate independently of central access-specific controllers: an integrated base station receives an IP packet for transmission through the IP-based back-haul/core network, processes the packet through the air-interface dependent protocol stack and transmits the packet in the access-dependent format over the wireless channel.

When a flat UMTS is compared to a hierarchical UMTS system, for example, a Node B cannot operate without connecting to a central *Radio Network Controller* (RNC) over the Iub protocol [9], and the RNC cannot operate without a Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). The latter two network elements are really just IP routers with 3G specific extensions sprinkled on top. For a UMTS integrated base station, we integrate functions from the GGSN, SGSN, RNC and Node B into a new IP-based base station that runs *all* protocols for sending and receiving IP packets over a UMTS wireless channel to a terminal. A collection of such new base stations can then form a cellular network. Note that while this paper primarily describes the implications for UMTS, similar arguments hold for a DO EvDO network, as well as WiMAX and other cellular standards.

Cellular systems became hierarchical for two reasons. When cellular systems were first devised, sharing the expensive vocoders over a (large) number of users led to considerable savings when deploying such cellular systems [38]. The savings result from not having to deploy such expensive vocoders in all the cell sites. Secondly, since wireless voice transmissions are compressed, fewer bits needed to be sent over the backhaul connecting the core network to the base stations, so more voice calls could be supported on a T1 or E1.

Later, with the introduction of CDMA systems, the hierarchy had an additional benefit for performing diversity transmission and reception [14]. Here, downlink data is prepared by a central anchor and then distributed to a number of base stations for simultaneous transmission over the wireless link. A mobile can thus combine the information from multiple *legs* before decoding the information. This type of transmission is particularly helpful in combatting fast-fading radio channels. Similarly, in the uplink, a central controller, such as the UMTS RNC, can select the best voice uplink packet before transmitting the received packets to the vocoders. Moreover, all protocol processing can be performed centrally to divide the at the time expensive processing resources over a large group of users.

We argue that the three fundamental reasons to build cellular systems in a hierarchical manner have disappeared. First, advances in electronics have made the cost argument disappear there is no reason the cost of electronics needs to dictate the cellular system architecture: every base station today can be equipped with cost effective processing environments to perform all access specific functions (including protocol processing). Second, by switching from circuit-switched voice to voice over IP, voice streams are already compressed over the backhaul between the end-points or end-point and media gateways and, with the increase of data usage, voice streams are expected only comprise a small part of the overall bandwidth. Lastly and more importantly: instead of using spatial diversity to combat fast fading, time diversity (i.e. fast retransmission from *one* base station [11], [5]) can be used instead. This means that there is no

reason to anchor a call higher up in the network for supporting diversity. Lastly, while spatial diversity is important for current/legacy 3G macro networks, evolved 3G networks are positioned for high-SNR environments with less frequency re-use [6], [8]. This implies that cell sizes are about to shrink, and overlap between such mini-cells is likely to reduced. The latter implies that diversity through soft handover will have diminishing returns.

So, in integrated base stations all protocol particulars are embedded at the cell site and a diverse set of air interfaces are easily integrated into a single converged IP-based cellular network. We argue that by encapsulating the parts specific to the air interface inside the integrated base stations themselves, different wireless standards, such as WiMAX, GSM, and/or UMTS, can easily co-exist in one infrastructure and accommodating new standards becomes much simpler. When the protocol anchors are well encapsulated and do not rely on air-technology dependent state in their network, we can envision inter-air-technology mobility. Within a flat cellular network, the interfaces to the AuC/HSS, inter-base station mobility procedures and IP-anchor must be standardized. We argue that all these interfaces are to be based on IP functions, such as DIAMETER over AAA and Mobile IP.

Using IP networks between the integrated base stations and the Layer-3 anchor is a reasonable approach, both from economic and performance viewpoints: IP routers are relatively inexpensive because of their abundance and they are fast as their functionality is mostly limited to routing IP packets. We observe that by keeping the interfaces between the access agnostic core and access-specific edge as small as possible, the access integration above described becomes realizable.

While we argue that most cellular functions can easily be replaced by the equivalent functions in IP networks, there are particular functions that require special attention. These functions are micro- and macro mobility, paging dormant mobiles, authorization and accounting, security functions and lawful intercept. We describe the implications of these functions later in this paper.

II. RELATED WORK

Truly hierarchical cellular systems are a relatively new development in cellular systems. In 1972, Joel [32] introduced a centralized *Base Station Controller* (BSC) that allows for paging a number of cells to find mobiles while the base stations patched directly into the *Mobile Switching Center* (MSC) for connecting to the land line telephony system.

In IS-136 and IS-95 the BSC functionality executed inside the MSC [27], while CDMA2000 used a separate base station controller. While DAMPS introduced a separate BSC, AMPS did not use a separate BSC. The DAMPS' BSC is the digital version of AMPS and the BSC performed audio compression and relocation functionality [14]. Note that both CDMA2000 and IS-95 are CDMA systems with soft handover, so strictly speaking there is no (technical) reason to introduce a separate Layer-2 anchor in the cellular system, but a central component is helpful for the stringent power control requirements [14] and to provide for soft handover.

GSM uses a BSC for call processing, mobility support and data forwarding and it also provides a vocoder network element that can be integrated in the central office [38]. The vocoder converts 64 Kb/s ISDN voice down to 16 Kb/s compressed voice and signaling streams. The reasons for the separate vocoder network element are that (1) the GSM consortium wanted to re-use existing ISDN switches as a MSC, and (2) by placing the vocoder in the central office, saving on operational expenditures for transmitting the voice over the wireless backhaul to the base station. Since UMTS was targeted as a successor to GSM it is no surprise that the central Layer-2 anchor remained

in place. In current 3G systems a network element called the *Media Gateway* is set to replace the GSM-like vocoders.

To summarize: The hierarchy originated for both technical and economic reasons. At first it was to offload the cross-bar switch or MSC, then vocoding helped build the hierarchy and later soft handover for CDMA systems further enforced the system hierarchy.

While cellular systems are mostly hierarchical, IEEE-802.11 is a typical example of a flat system. Here all wireless protocol processing is performed in the access point. While mobility issues have been at the forefront of cellular system design [20], IEEE-802.11 only implements the equivalent of a break-before-make hard handoff [36]. IEEE-802.11 fast handovers are handovers that do not require re-authentication after association with a new Access Point [29]. By using techniques such as SyncScan [40], the impact of the periodic scan for beacons can be limited; this happens by carefully synchronizing the transmission of the AP's beacon with the mobile's scan functionality. With SyncScan, the mobile only needs to authenticate and re-associate itself with an AP measured earlier.

When we consider currently known next-generation cellular network systems, a few fundamental themes are clear. Since IP has become the accepted access protocol for the majority of data networks and since a significant portion of future wireless traffic will be data based, it is not difficult to understand why there is an effort in converting the backhaul network to be based on IP. IEEE-802.11 networks already carry IP traffic in their backhaul, even Bluetooth networks carry IP traffic. Cellular equipment makers are also introducing *Radio Access Networks* (RANs) based on IP (Nokia's IPRAN [26], Ericsson's cellular IP based radio network [17], Flarion's all-IP network [24], NTT-Docomo's wireless IP networks [21] etc).

There exist alternate architecture proposals from the IP community to introduce and establish IP protocols in the cellular networks (e.g. Hawaii [41], Cellular IP [17], IDMP [37], hierarchical Mobile IP [18], etc). However these architectures primarily restrict themselves to efficient protocol designs to support handoff that minimize packet loss.

While iHSPA as proposed by Nokia [39] seems to correspond well to our flat architecture, especially in the traffic plane, we argue that a true flat architecture is farther along the path of 'flattening'. The main difference between our approach and that of iHSPA is that iHSPA still requires central network elements for signaling traffic.

III. ECONOMICS

In economic terms we postulate that combining Layers-3, -2 and -1 processing in the base station makes more sense than placing some of the functionality centrally. A central component such as the RNC needs to divide its processing resources between a possibly large number of Node Bs (e.g. Ericsson's GSM solution uses 512 base stations per BSC [38]), serving a very large number of users.

However, as described before, the reason this central component came about was to be able to share the expensive resources for a large number of base stations and thus users. Since cheap processing resources have become abundant, the economic reason for centralization and sharing has disappeared: in fact, every base station can now be made to process all Layer-3, -2, and -1 functions at the cell site cost effectively. A side effect of this grouping is that while central components can suffer from queuing delays for having to share a limited number of processing elements, decentralized processing can avoid such queuing delays altogether.

Hierarchical systems are not self-scaling: advances in wireless transmission technology that boost the performance of a set of base stations may carry serious performance issues for the central components. If those central components do not have enough resources to deal with the advanced base stations, it becomes a new bottleneck in

the cellular system. Another way of looking at this particular problem is that when advances in base station transmission technology alter anchor protocol processing, a central component needs to support both older and newer technologies at the same time. In a flat system, older and newer base stations would just co-exist as they are not inter-dependent a flat system is inherently self-scaling.

A serious consideration for central components is that failures in such central components can lead to serious disruptions of the cellular service. A central solution, thus needs to be provisioned to survive (partial) failures, which likely lead to costly solutions - e.g. active fail-over solutions. Typically, such solutions are one-of-a-kind solutions and therefore expensive and error prone. Instead, by distributing the processing to the edge of the network, a failure in one integrated base station cannot bring down the system as a whole.

A hierarchical system solution with central controllers and base stations implies that access-specific functionality is split over multiple network elements. Since central controllers need to operate in unison with base stations, interfaces need to be defined, maintained and tested before a cellular system can be deployed. Worse: since the connection between a central controller and base station is yet another networking protocol, communication between the central controller and base station is subject to typical networking problems. This implies that failure recovery for lost, duplicated and delayed and proprietary messages is required in both the central controller and base stations. This adds complexity to the system. The complexity manifests itself by longer transmission times, higher development costs for the central controller and base stations, but also in higher operational costs as there are more interfaces to control. By removing the specialized backhaul protocol altogether and replacing the inter-connect by a simple IP-based network, the complexity of the system reduces and thus is likely to perform better.

While it is true that operating and maintaining central components is cheaper than lots of decentralized components, we argue that a flat system is easier to operate and maintain than a hierarchical system. Even with today's hierarchical system, each base station needs to be maintained separately in any event. In short: by removing the central component from the system, there are fewer interfaces to operate and maintain.

IV. CONVERGENCE

Convergence of access technologies is a hot topic for wireless and cellular system design. For instance, 3GPP is considering a system architecture where a WiFi network can easily be integrated in a GPRS core network [6]. Similarly, mobility between systems such as WiMAX [30] and EvDO [11] is considered advantageous. Lastly, it is not surprising that operators are contemplating the convergence between fixed solutions, such as DSL, and mobile solutions.

We argue that only by keeping the interface between the various access technologies minimal, access technologies can successfully be integrated into a common infrastructure. One of these common infrastructures is an IP packet core. In its simplest form, such an IP core provides IP packet routing only. This means that an IP packet destined for a particular mobile is routed from the Internet to the base station serving that mobile by its address label.

A cellular system requires more than just packet routing for packet delivery. For instance, a nomadic user can relocate between base stations and for this, a Mobile IP anchor [15], [33] is required that receives the packet destined for the mobile and routes it to a particular base station for delivery. To establish a mobile IP session through a packet, additional IP functions are required to distribute Mobile IP keys through the cellular network and mobile [19], [16]. Generally, an accounting, authorization and authentication scheme based on IP
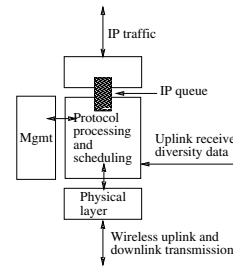

Fig. 1.   *Flat architecture components*

technologies is necessary to build an access-agnostic packet core [23], [13].

Once all access specific functions have been removed from a packet core, converging the various access technologies into a single logical system is a straightforward operation. Such convergence enables leveraging the common packet core over a (potentially) large number of access technologies.

V. PROTOCOL PROCESSING IN FLAT ARCHITECTURES

Protocol processing in flat architectures is not different, in principle, from protocol processing in hierarchical systems the interface to the mobile does not change by changing the infrastructure in the core network. The main difference in protocol processing lies in the absence of protocol layers that are only meant for access-specific inter-core-network element communication. For example, for an integrated UMTS system with macro-mobility through Mobile IP, there are no GTP tunnels and there is no Iub protocol layer [7], [9].

Figure 1 shows a typical software component architecture for a flat architecture base station. IP traffic arrives over the asynchronous backhaul into the base station and a queue is maintained to bridge the asynchronous IP backhaul reception to synchronous wireless transmissions. The majority of the protocol processing is performed in the middle component: whenever a mobile is scheduled for service, the protocol stack for that mobile is initiated, it de-queues IP packets from the IP queue and the protocol stack prepares transmission frames for transmission over the wireless link. The transmission blocks are sent through the physical layer. Similarly, uplink transport blocks are received by the physical layer and pushed into the protocol stack for reassembly. When an IP packet has been re-assembled, the IP packet is injected into the backhaul. If uplink receive diversity is supported, neighboring base stations may forward received uplink packets into the protocol stack through a side (IP) door.

To support macro mobility, the IP component supports IP mobility by, for example, Mobile IP [15]. This component then maintains communication with a Mobile IP Home Agent as is shown in Figure 2.

A control component controls the communication paths through the protocol stack and physical layer. This component sets up and tears down communication paths through the protocol processing layer into the physical layer. Additionally, the control component aids in the authentication and mobility management of mobiles.

VI. MOBILITY IN A FLAT ARCHITECTURE

Mobility in hierarchical systems is provided for by plumbing explicit routes (or tunnels) through the hierarchy of the system to one or more base stations that maintain a wireless connections with the mobile. In the UMTS packet service for example, the GGSN forwards the IP packets to the SGSN which then tunnels the data to the RNC. Next, the RNC processes the IP packet into transport blocks and forwards those transport blocks over ATM to the Node Bs.

While flat architectures can re-use existing GGSNs and SGSNs for backhaul routing, there is no requirement to do so. Instead, for mobility, a flat cellular architecture can use any macro-mobility mechanism when there is no inter-operability requirement for interacting with existing systems. Compared to the relatively heavy-weight protocols used by SGSNs and GGSNs, the use of Mobile IP (v4 or v6) for macro-mobility is appealing for its simplicity [15], [28]. Moreover, when compared to GPRS mobility, Mobile IP is an active area of research. In our example flat architecture, we assume that most SGSN and GGSN functions are encapsulated in the base station, and IP routing is provided by Mobile IP.

One of the strengths of Mobile IP is also one of the weaknesses of Mobile IP: in principle Mobile IP does not concern itself with micromobility. This means that when a physical channel is relocated from serving cell to serving cell, care must be taken to recover pending transmission and reception state to avoid the loss of already received but not yet re-assembled IP packets and to avoid the loss of untransmitted partial IP packets. The loss of any these data leads to an application layer retransmission and typically such retransmission leads to service interruptions. For example, the loss of a single TCP packet implies the TCP congestion window to be (dramatically) reduced [25].

While proposals for generalized micro- and macro-mobility with Mobile IP procedures are available [35], these algorithms are typically standard specific. In this paper we review an integrated micro- and macro-relocation procedure for a UMTS relocation procedure. Even though this procedure is specific to the standard at hand, we can abstract from this specific procedure to understand the underlying issues. We have implemented such a combined mobility procedure for a UMTS flat system and reported on the performance separately [12]. Independently, our approach has been adopted by SAE/LTE [22].

Note that even the standard SGSN approach has limited support for seamless anchor relocation [2].

Micro mobility

To understand the issues of micro mobility in flat and hierarchical systems, we present four different system architectures in Figure 2: a strictly hierarchical, traditional 3G system using SHO, then a 3.5G hybrid system with *Soft Handover* (SHO) in the uplink and H-ARQ in the downlink, a flat system with uplink and downlink SHO and a flat system with uplink SHO and downlink H-ARQ. Depending on the configuration, a different type of action is taken when the mobile reports the RF conditions for a number of measured cells [3]. The major difference between hierarchical and flat is the location of the protocol processing anchor: in flat systems the anchor resides in the base station, in hierarchical systems, the anchor resides in the central controller.

For the discussion that follows, an active set of radio legs is a group of radio channels (containing signaling or data channels or both) transmitted from a set of base stations. The mobile combines the radio channels from the base stations participating in the active set before decoding the information. This type of mobility is typically used to combat fast fading in CDMA systems.

Typically, in 3G cellular systems, mobility is triggered by the mobile: as soon as the mobile measures a good or better cell by way of the cell's pilots, it reports this through the transmission of a measurement report to the RAN. For hierarchical, non-H-ARQ-based systems, a measurement report either triggers the addition of a new leg to the active set, or acts as reinforcement that the leg remain part of the active set. This is the default scenario for hierarchical UMTS systems. For this type of operation *all* downlink data is multi-cast to *all* legs in the active set and this scenario

requires tight synchronization between the RNC and all of the base stations participating in the active set. Note that the leg with the most delay/jitter in the backhaul delays the end-to-end transmission most.

For hierarchical, H-ARQ based systems the course of action depends on the relative strength of the reported leg. If the leg was not part of the active set, it is added to the active set by signaling the base stations and terminal. If the report indicates that the selected leg is the strongest leg to the terminal, the system relocates the H-ARQ scheduler to the cell hosting that strongest leg. Since hierarchical systems do not exchange information of the last packet transmitted from the old leg, some packet or capacity loss is expected during the relocation as old and new serving base station do not synchronize their transmission queues.

For flat, non-H-ARQ systems the course of action depends on supporting soft handover. If soft handover is supported in a flat architecture, then a similar action is performed as in the hierarchical system. This implies that data is multi-cast from the Layer-2 anchor up into the backhaul to the other members of the active set. As is shown in the figure, the hierarchical backhaul is replaced by an IP backhaul. Given that there is much less control of what goes over the backhaul compared to hierarchical systems, predicting the transmission times of packets in downlink soft handover is much harder. This implies that compared to hierarchical soft handover systems, flat soft handover systems are likely to perform less in terms of transmission latency.

If soft handover is not supported in the flat system, then an anchor relocation is performed if the reported strongest pilot has changed. The mobile is commanded to perform a hard-handover and all of the state associated with the Layer-2 anchor is relocated to the base station serving the strongest cell. Since, protocol processing anchors are integrated in base stations, this anchor relocation implies that a macro-mobility procedure is initiated as well. Note that care needs to be taken to perform a seamless mobility procedure here: ideally no transport blocks will need to be retransmitted, no link-layer resets are required [1] and no transmission opportunities are lost.

For flat, H-ARQ based systems the same course of action is followed as the flat, non-H-ARQ based, non-SHO systems: the Layer-2 anchor follows the strongest downlink. The major difference between this situation from the non-H-ARQ, non-soft handover flat system is that an active set can be supported in this scenario. Here, data is never transmitted in soft-handover mode (not even signaling channels), which implies that even though an active set is maintained, no two base stations require tight synchronization. The active set is only maintained for RF synchronization for when the anchor is relocated from cell to cell to avoid costly over-the-air synchronization procedures [4].

If the measurement report indicates that a leg is not strong enough anymore in the active set, it is discarded by cleaning up radio resources and removal of data paths through the system.

Macro mobility

As shown in Figure 2, mobility in hierarchical systems can be performed without changing the anchor, while in flat system the anchor needs to be relocated. An implication of anchor relocation is that macro-mobility needs to be initiated as well: the protocol processing endpoint is relocated from base station to base station.

If Mobile IP is used for macro-mobility, its procedures need to be triggered to reroute the GRE tunnel from the source base station to target base station when the anchor is relocated. To avoid packet loss during this migration, we use a triangle route between source and destination base station for the duration of the mobility procedure. Once the relocation has been performed, the triangle route is discarded.
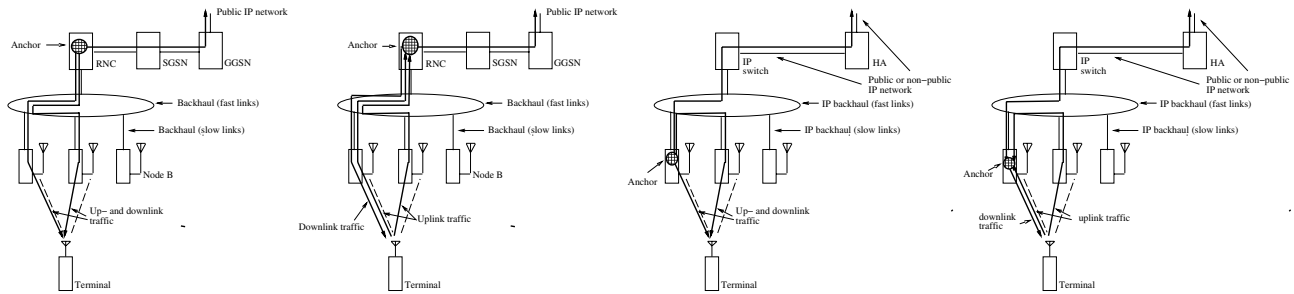
Fig. 2. *From left to right: hierarchical with* SHO*, hierarchical with* H-ARQ*, flat with* SHO*, flat with* H-ARQ

VII. PAGING

Paging is the procedure to find dormant mobiles in a particular paging area. In 3G systems, such as UMTS, the mobile does not maintain radio channels when there are no packets to transmit (e.g. voice or IP packets); this conserves battery power. Periodically, the mobile listens to the paging channels to find out if it needs to re-establish radio channels with the RAN to set up communication paths for an incoming voice call or IP session.

In a hierarchical system, the central node (e.g. RNC, MSC or SGSN) maintains which mobiles are available in which paging area. An incoming call for a dormant mobile first addressed the central location, which then initiates access-specific paging functions to the set of base stations that collectively form a paging area in which the mobile is registered. All base stations in the paging area then transmit a page over the paging channel in their cell to find the mobile. When the mobile hears the page, it wakes up and re-establishes contact. Once communication paths have been re-established, the call or IP packet is delivered on the mobile.

In flat systems, there is no central anchor to receive the call or IP packet. Instead, one of the base stations assumes the role of the central node. Typically, the base station of last attachment for the mobile becomes the anchor point for the mobile. An incoming call or IP packet now arrives in the last point of attachment. From this point, like in hierarchical system, the paging area is addressed by an n-point uni-cast from the anchor. Each of the base stations addressed page the mobile and the mobile responds to one of the base stations. Next, if the mobile responds on a different base station than the anchor, the state associated with the mobile is relocated from base station to base station before the initial call or IP packet delivery completes.

In a hierarchical system, the paging state can be kept highly available, for example, by replicating it. If the anchor fails, another node can take over the role of anchor to avoid making the mobile unreachable. Similarly, in a flat system, the state associated with the mobile must be replicated.

VIII. SECURITY FUNCTIONS

A 3G cellular system maintains security state associated with the mobiles in physically secure central nodes. For instance, a UMTS system maintains the secret keys it shares with the mobile's uSIM inside its authentication center. In addition, in the SGSN and RNC, UMTS maintains ciphering and integrity state that is used to, respectively, encrypt data that is exchanged with mobile and to sign the signaling messages. Such state must be kept secure to prevent eavesdropping and hijacking of sessions. RNCs and SGSNs are considered to be physically secure machines.

In addition to the state that must be kept secure, there are several other functions in the cellular network that need to be kept under lock and key. For instance, charging records for the actual data usage may not be altered without authorization and should preferably be created by network data usage need not be altered with and preferably created by network elements that are under full control of the cellular provider.

We argue that those functions that are access specific need to be collapsed in the cell site, provided these functions can be secured. Those functions that are really access-technology independent need to be subsumed in the core network. So, by this principle, all UMTS keys must be secured inside the cell site, while functions such as charging can be integrated into the core network: the former functions cannot be shared between the various access technologies, while the latter function is a clear example of a function that can be shared between multiple access technologies.

Having stated that we need to collapse functions at the cell site, the next question is how to secure the cell site. Within 3GPP SAE/LTE it is believed that the cell site cannot be secured, although we believe this decision is not based on considering all available processing technologies that are currently available [34]. We argue that cost-effective tamper-resistant hardware exists. With this tamper-resistant hardware at the cell site we construct a *cell-site vault*. The idea of the vault is that all functions that need to be secured run inside the vault and all functions that do not need to be secured can be run outside the vault.

The vault inside the cell site is a processing environment that is hard to break in to. It provides the processing capacity for performing all functions related to data re-ciphering and to accept and respond to sensitive control messages that are exchanged both with the core network and with the mobile. In UMTS terms, this means that the vault provides functionality to cipher data with Kasumi and it provides the functions to accept the NAS messages that are destined to the SGSN in typical hierarchical systems.

In fact, base stations need a secure computing environment for other reasons as well: providers do not want their base stations to be tampered with, so they want control over software upgrades, authentication and several other things. In addition, the keys for Mobile IP need to be maintained so control messages can be signed to prove to the packet core that the base station is part of the cellular network.

IX. VARIOUS

Accounting, authentication and authorization are typically extraneous functions provided by an access network to count the user's usage, to authenticate users to the network and for the network to authorize the user to use the network. In current cellular networks, these functions are intimately tied to the central network elements. In UMTS for example, accounting is performed in the SGSN and authorization is done through the AuC and SGSN/MSC. In flat systems, these functions need to be performed as well. We believe these functions are access-agnostic core functions and thus they should

be performed by the shared infrastructure, that is, the IP packet core. For those functions that are access specific, the integrated base station provides translation functionality into the common and shared functions. The IETF has standardized a number of procedures that can host AAA functionality [16].

There are more functions that are part of a typical packet core, such as legal intercept, QoS management, etc. Again, these functions can either be implemented easily in an access-agnostic packet core, or they can be provided for through auxiliary functions such as ITU's RACF [31].

X. Summary

In this paper we argue for flat cellular networks connected through IP networks. While H-ARQ is an enabler for this development, flat systems provide benefits in terms of reduced overall system complexity, reduced end-to-end latencies, self scaling opportunities, and the deployment of converged cellular networks with heterogeneous air interfaces. When converting between a hierarchical and flat system with an IP core, care must be taken to support seamless mobility, security, accounting, authorization and authentication, paging and QoS. In this paper we show that such functions can be supported through existing IP functions.

References

[1] 3GPP. Radio link control (rlc) protocol specification (rel. 5). Technical Report TR 25.322, 3GPP, september 2002.
[2] 3GPP. Utran functions, examples on signalling procedures. Technical Report TR 25.931 V5.0.0, 3GPP, March 2002.
[3] 3GPP. Rrc protocol specification. Technical Report TR 25.331, 3GPP, 2003.
[4] 3GPP. Synchronization in utran stage 2. Technical Report TR 25.402, 3GPP, December 2003.
[5] 3GPP. High speed downlink packet access (hsdpa): Overall description. Technical Report TR 25.308 V5.4.0, 3GPP, December 2004.
[6] 3GPP. 3gpp system architecture evolution: Report on technical options and conclusions (rel 7). Technical Report TR 23.882, 3GPP, 2006.
[7] 3GPP. General packet radio service (gprs): Service description, stage 2 (release 7). Technical Report TS 23.060, 3GPP, June 2006.
[8] 3GPP. Requirements for evolved utra (e-utra) and evolved utran (e-utran) (release 7). Technical Report TR 25.913, 3GPP, 2006.
[9] 3GPP. Utran iur and iub interface data transport and transport signalling for dch. Technical Report TS 25.426, 3GPP, June 2006.
[10] Markus Bauer, Peter Bosch, Nidal Khrais, Louis Samuel, and Peter Schefczik. The umts base station router (bsr). *(to appear in) Bell Labs Technical Journal*, 2006.
[11] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayanna, and A. Viterbi. CDMA/HDR: A Bandwidth Efficient High-Speed Wireless Data Service for Nomadic Users. In *IEEE Communications Magazine*, july 2000.
[12] Peter Bosch, Sape Mullender, and Louis Samuel. Mobility in umts packet ip networks. In *WCNC'06, Las Vegas*, 2006.
[13] Peter Bosch and Louis Samuel. Gprs versus mobile ip. Technical report, Bell Laboratories, Lucent Technologies (white paper), 2006.
[14] Neil J. Boucher. *The Cellular Radio Handbook: A Reference for Cellular System Operation*, Jan. 1995.
[15] ed. C. Perkins. Ip mobility support for ipv4. Technical Report RFC3344, IETF: Network Working Group, August 2002.
[16] P. Calhoun, T. Johansson, C. Perkins, T. Hiller ed., and P. McCann. Diameter mobile ipv4 application. Technical Report RFC4004, IETF: Network Working Group, August 2005.
[17] Andrew T. Campbell, Javier Gomez, and Andrs G. Valk. An overview of cellular IP. *IEEE Wireless Communications and Networks Conferance 1999 (WCNC'99)*, 2:606–611, 1999.
[18] Claude Castelluccia. A hierarchical mobility management scheme for ipv6.
[19] Betsy Covell (chair). Fast handoff for hrpd. Technical report, 3GPP2, X31-20060731-xxx, 2006.
[20] R.A. Chaney. Automatic mobile radio telephone switching. Technical Report US Patent 3,355,556, Bell Laboratories, Nov. 28, 1967.
[21] Docomo. Openran: A new architecture for mobile wireless internet radio access networks. *IEEE Communications Magazine*, 40(5), May 2002.
[22] Ericsson. Requirements on s1 and x2 user plane protocols. Technical Report R3-060720, 3GPP TSG-RAN WG3 52, May 2006.
[23] P. Eronen, T. Hiller, and G. Zorn. Diameter extensible authentication protocol (eap) application. Technical Report RFC4072, IETF: Network Working Group, August 2005.
[24] Flarion. The benefits of a packet-switched, all-ip mobile broadband network. Technical Report Flarion White Paper, Flarion, February 2004.
[25] Sallu Floyd and van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):297–413, August 1993.
[26] Yile Guo, Z. Antoniou, and S. Dixit. Ip transport in 3g radio access networks: an mpls-based approach. *IEEE Wireless Communications and Networking Conference*, 1:11–17, March 2002.
[27] Lawrence Harte, Richard Levine, and Steve Prokup. *Cellular and PCS*, 1997.
[28] Christian Huitema. Ipv6: the new internet protocol. Technical Report ISBN 0-13-850505-5, Prentice Hall, 1997.
[29] IEEE. Draft: Recommended practice for multi-vendor access point inter-operability via an inter-access point protocol access distribution systems supporting ieee 802.11 operation. Technical Report P802.11f/D3, IEEE, January 2002.
[30] IEEE. Ieee standard for local and metropolitan area networks, part 16: Air interface for fixed and mobile broadband wireless access systems. Technical report, IEEE, 2006.
[31] ITU-T. Ngn draft tr-racf rev. 2, functional requirement and architecture for resource and admission control in next generation networks. Technical report, ITU.
[32] A.E. Joel. Mobile communication system. Technical Report US Patent 3,663,762, Bell Telephone Laboratories, May 16, 1972.
[33] D. Johnson, C. Perkins, and J. Arkko. Mobility support in ipv6. Technical Report RFC3775, IETF: Network Working Group, June 2004.
[34] 3GPP SAE LTE/SA3/RAN2. 3gpp sae lte decision against flat architectures, Oct. 2005.
[35] P. McCann. Mobile ipv6 fast handovers for 802.11 networks. Technical Report 80211fh, IETF, October 2004.
[36] Arunesh Mishra, Minho Shin, and William Arbaugh. An empirical analysis of the ieee 802.11 mac layer handoff process. Technical report, University of Maryland, 2002.
[37] Archan Misra, S. Das, A.Dutta, and Anthony Mcauley. Idmp: An intra-domain mobility management protocol using mobility agents. Technical report, IETF, July 2000.
[38] Michel Mouly and Marie-Bernadette Pautet. The gsm system for mobile communications. Technical Report ISBN 0-94559-215-9, Telecom Pub, 1992.
[39] Nokia. Considerations for system architecture evolution. Technical Report SRJ050020/S2-051260, 3GPP TSG-SA WG2 Meeting 46, May 2005.
[40] Ishwar Ramani and Stefan Savage. Syncscan: Practical fast handoff for 802.11 infrastructure networks. *IEEE InfoCom 2005*, March 2005.
[41] Ramachandran Ramjee, Thomas F. La Porta, S. Thuel, Kannan Varadhan, and S. Y. Wang. HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks. In *ICNP*, pages 283–292, 1999.