White paper

– 3G – The Third Generation of Mobile Services

AU-SYSTEM

About the authors:

3G Business and Services

Thomas Fransson, Senior Consultant at AU-System in Stockholm. Having worked for both operators, vendors and as a strategy consultant, Thomas has wide experience from various fields of the telecommunications sector since 1984. He has been in the mobile field since 1995 and is a renowned analyst and speaker dealing with the challenge of the changing role for operators and how they are going to make money in the future.

Peter Kaarle, Value Added Propositions Manager at AU-System in Lund. Peter has ten years of experience from the telecom industry. For the last three years he has been focusing on Mobile eCommerce concepts and services. He is currently responsible for development of new Value Added Propositions for Mobile Internet and commerce solutions.

Mats Krekula, Consulting Manager Billing at AU-System in Stockholm. Mats has 14 years of experience in the Telecom Management area, with System Development, Project Management and Sales responsibilities at Ericsson and Hewlett-Packard

Esa Örmä, Senior Consultant at AU-System in Stockholm. Esa has been working professionally in the IT/Telecom business since 1986 and for the last ten years within billing.

3G Standardization

Eric Ljungberg, Vice President of Wireless Development at AU-System in Lund. Eric has experience from the following standardization organizations: ETSI, SMG5, ITU-R, ETSI SMG Plenary, 3GPP Programme Coordination Group and the ETSI Board.

3G Technology

Eskil Åhlin, Senior Consultant at AU-System in Lund. Eskil has experience from working with UMTS in 3GPP and 3G.IP.

Copyright statement

3G – The Third Generation of Mobile Services

Contents

Introduction **4**

3G Business and Services **5**

Business Scope 5
The User Experience 6
Anytime, Anywhere, Anyway - a 3G Service Scenario 7
A Real-Life Example of Situation-Centric 3G Services 7
Implications for the Service Provider 8
Business Models 9
3G Billing 9
Uncertainty in the Marketplace 9
Competition 10
Co-operation 10
Complexity 11
Architecture 11
Content 11
Challenges 12
Conclusions 13

3G Standardization **14**

The Search for Spectrum 14
The Selection of a Radio Interface 14
Core Network Strategies 15
A Global Approach is Resumed 16
3G Standardisation Today 16
3GPP 16
IETF 17

3G Technology **18**

3G Networks 18
The Radio Access Networks 18
The Backbones 19
The IP Multimedia Subsystem 20
3G Terminals 21
Security 22
3G Messaging 23
Service Execution in Terminals 25

Figures

Figure 1	Extended service scope (Ovum)	5
Figure 2	New role categories	5
Figure 3	The connecting quadrant and its roles	6
Figure 4	Service values	6
Figure 5	The user experience	7
Figure 6	Situation-centric services: the market maker and the ecosystem behind	8
Figure 7	The need for new revenue (Microsoft)	9
Figure 8.	Revenue categories	9
Figure 9	Transaction flows (UMTS Forum, Report #11)	10
Figure 10	UMTS billing system	11
Figure 11	Jalda purchase overview (EHPT Safe Trader, product description)	12
Figure 12	Service-centric content billing	12
Figure 13	Worldwide 3G frequency allocation (UMTS Forum)	15
Figure 14	The 3GPP Organization	16
Figure 15	The UMTS networks and domains	18
Figure 16	Close view of the packet switched UMTS backbone	19
Figure 17	The evolution of the circuit switched UMTS backbone	20
Figure 18	The IM subsystem	21
Figure 19	Home control of services	21
Figure 20	UMTS Security	22
Figure 21	The IETF Presence Service model	23
Figure 22	The 3GPP MMS Architecture	24
Figure 23	Java Editions, Configurations and Profiles	25
Figure 24	MExE service execution	26

Introduction

The emergence of the third generation (3G) of mobile services, devices, and networks means that the connected society is here, together with its ubiquitous online services.

The architecture and technology to create this vision are already standardized. Products and networks are underway and the different players in the value network are beginning to see their business opportunities. Operators can now benefit from making the Internet and all its e-services completely mobile, broadband, and even multi-access. Service and content providers can make use of these new market channels with yet unexplored features for customer interaction.

This white paper presents an overview of critical business and technology aspects for operators and service providers with or without 3G frequency spectrum that currently struggle with the question of how to move their business over to a 3G basis.

The first section concentrates on the business and service provisioning side of 3G and discusses issues such as:

- What does the value network look like for 3G service provision?
- What business models are suitable for use with the kind of business that 3G generates?
- What kind of user experience can be obtained and what is the value-added for the user?
- What is the nature of a 3G service?
- How will billing be implemented and how is this related to other forms of payment that will be used over the 3G network?

The second section deals with the story of how a worldwide 3G mobile standard was achieved through years of intriguing techno-politics, while the third section describes the architectures and technology that resulted from that process, covering:

- The 3G network architecture - including radio access networks, the backbone, and IP-multimedia – and how the architecture produces the multi-access and multimedia features of 3G networks.
- 3G messaging architecture including the Instant Messaging and Presence Protocol.
- 3G security issues.
- 3G service execution in terminals.

Note: 3G standards as defined by the ITU under the IMT-2000 umbrella comprise a set of competing standards, UMTS and cdma2000 being the most well known. This white paper is UMTS-centric.

3G Business and Services

The 3G business is driven by a strong vision of a society with omnipresent online services and a sense of both presence and control provided to everyone living in this society. However, several uncertainties remain as to how this vision will be realized, and entering the 3G business to some extent can be characterized as a "leap of faith".

Business Scope

Technology development, "everything is becoming bits and IT", alters the focus in so-called "telecom services" from mainly human-to-human voice communication to content and applications that are human-to-machine, machine-to-human, and - behind the scenes - machine-to-machine.

This change of focus leads to a crossroads for all players currently offering "telecom services". The full scope becomes so enlarged that it is neither natural nor possible for a single integrated telecom player to manage the whole.

Instead, we see a division of activities and players into the following categories:

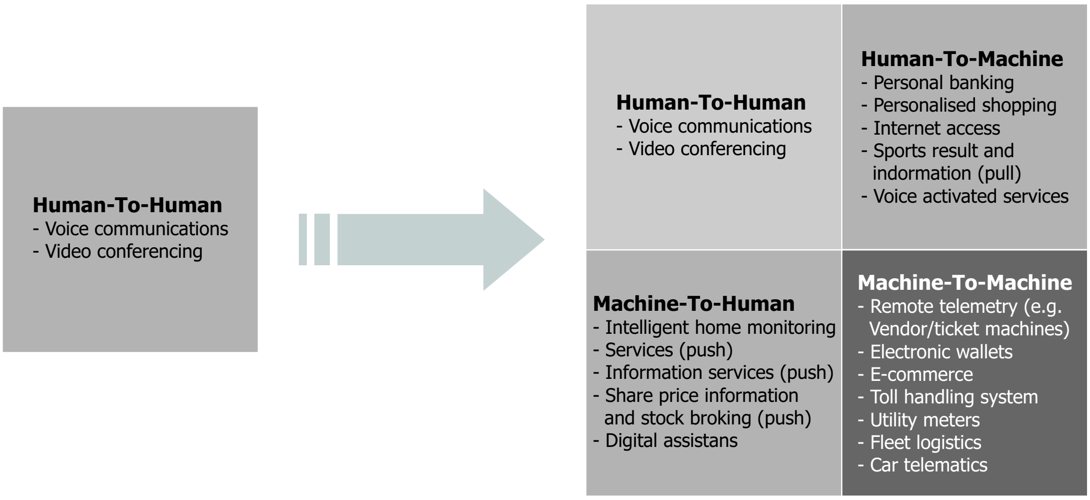

Figure 1. Extended service scope

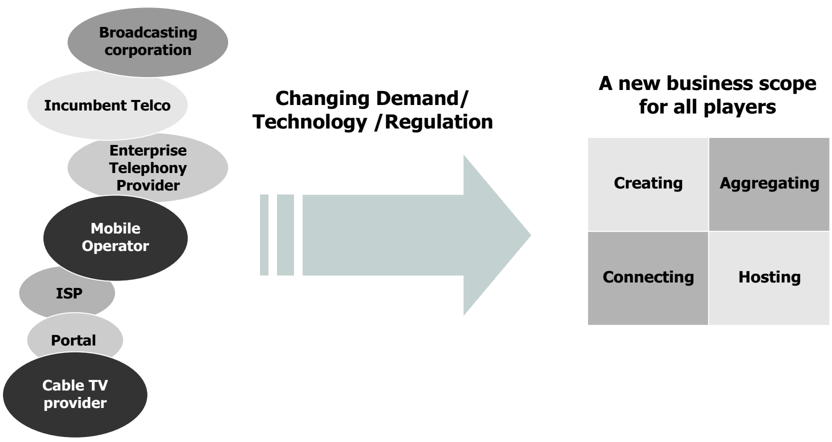

Figure 2. New role categories

- **Creating**
  "Primary production of content and services."
- **Aggregating**
  "Creating context, that is, enhancing the value by tailoring a set of services and content for a specific segment or for a specific situation."
- **Connecting**
  "Enabling people, services, and machines to communicate." This role is the most familiar to operators from the "old world".
- **Hosting**
  "Hosting operations, services, applications, and utilities."

This is the arena in which the 3G players will operate.

The 3G opportunities will begin in the "connecting" category where the 3G networks will enable other players to join the game. However, in a world of declining mobile voice ARPUs, the whole field of players will have to be populated at once in order to start exploiting the entire service scope and to secure a multitude of revenue streams to provide healthy returns on the 3G network owners' investments.

The access independent features of the 3G service architecture will further fuel the current multi-access service trend and put 3G players in a very good position for offering their services across different access and terminal types.

This may lead to a further division of the players within the connecting category, whereby network owners with extensive infrastructure assets operate specific access forms and sell capacity on a wholesale basis to relatively asset-light multi-access telecom service providers. The latter player category may try to stretch into the aggregating quadrant as well by developing intangible assets such as network, brand, and community values.

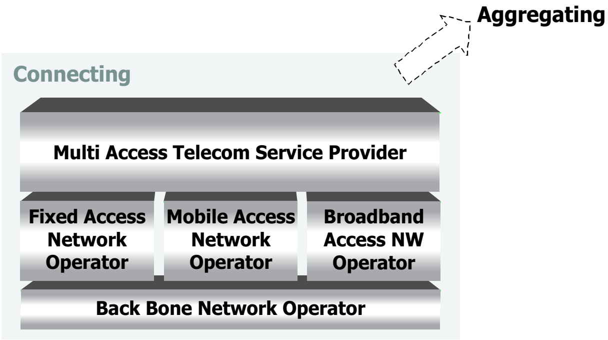

Figure 3. The connecting quadrant and its roles

It would be advisable for players with role aspirations beyond their current positions to act quickly in order to take their new roles before the playing field is defined and changes become difficult or even impossible.

The User Experience

With its multimedia and broadband features, 3G will provide the means to bring a truly differentiated user experience to mobile devices. This opens the door for innovative services both technically and in terms of marketing.

A key feature of all successful services is not only the ability to offer accessibility on the move, but also being situation-centric. That is, to be tailored to the user context in every situation in order to really provide the support most needed at that moment. This can be further developed to transaction fulfillment and instant buying experiences. Climbing this ladder of increased value is of vital importance for all players considering offering services via a 3G network. The aim is to constantly surprise users with enhanced and timely experiences.

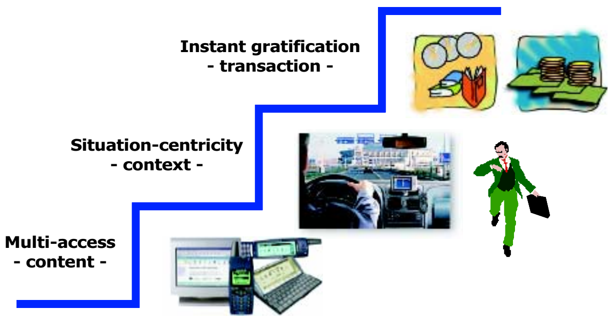

Figure 4. Service values

An essential component in such an offering is the ability for service providers to combine their values in composite services put together on the fly when a situation arises. This requires a B2B ecosystem with dynamic brokering facilities and service providers that have enabled their services with, for instance, standardized discovery and negotiation mechanisms. The emergence of such ecosystems is probably very important for full user acceptance of 3G services. It is also probable that this is the next step in the current mobile portal notion, beyond the current services characterized by static composition.

Apart from being practical on the fly, mobile services based on 3G features have the potential to address even more sophisticated human needs. Experience from i-mode in Japan shows that in an advanced society, services have to address advanced needs to be successful. The most successful services are possibly the ones that address psychological needs such as comfort and recognition, that is, those that manifest the individual as part of a social context. However, offering such services places heavy demands on personalization, selection range, ability to contribute/publish, and lifelike image production. These are all features that 3G will provide the means to support.

The huge service scope potential of 3G means that the 3G players have an excellent opportunity to attract consumer spending from numerous fields other than the former telecommunications field. This, however, requires much more sophisticated and distinct market segmentation than is currently in existence. The 3G players need to offer a rich service portfolio based on a thorough analysis of their customer segmentation to find the qualifying, loyalty building, and differentiating steps for each segment. This analysis has to be iterative and always closely associated with user reactions to remain successful over time.

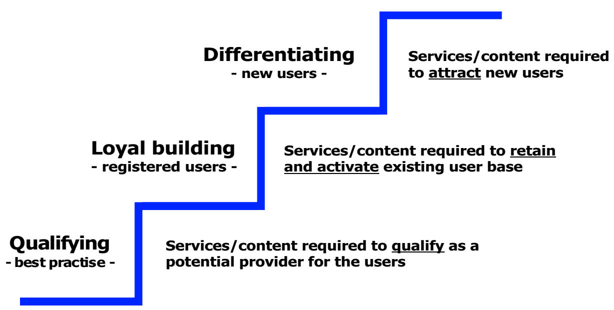

Figure 5. The user experience

The notion of mobility will probably remain so strong that certain 3G players that utilize their mobility asset in an intelligent way will manage to take market maker positions in some ecosystems.

Anytime, Anywhere, Anyway - a 3G Service Scenario

Mobile operators and other players have started to implement mobile portals. They need to take on the role of providers of high-value services with mobile access and accustom their users to more advanced mobile services. The introduction of 3G services will also usher in a new business landscape with more freedom for users to decide which content and services create most value for them.

The main application of the second generation (2G) network is mobile voice communication between people. Whereas with the Internet, e-mail and searching for, providing, and exchanging information comprise the main usage areas. In one respect, we can think of 3G as adding Internet services to the 2G network, with the combination made possible by increased bandwidth and PC-like mobile devices. 3G can also be thought of as adding mobility to limited Internet devices.

A Real-Life Example of Situation-Centric 3G Services

Below are examples of how 3G mobile services can make life easier for a traveling consultant at AU-System. We will examine how features such as multimedia capabilities, positioning, multi-access, and community membership can aid AU-System employee Peter when conducting a four-day business trip in Italy together with a colleague while still keeping close contact with his wife and son back home.

Multimedia Capabilities

By sending sketches from a whiteboard via his 3G terminal to his colleagues in Lund, Sweden and San Diego, USA, Peter is able to solve a complex problem scenario that is discussed with his Italian client. The consultants in San Diego are able to provide Peter with a solution, which are produced overnight before Peter's early morning meeting with his client.

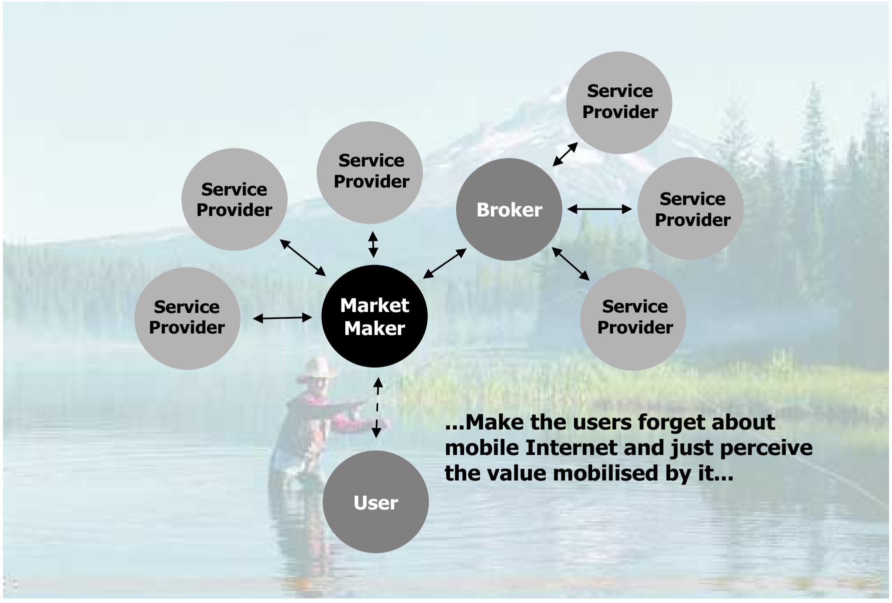

Figure 6. Situation-centric services: the market maker and the ecosystem behind

Positioning

When arriving at the train station in Rome after a visit in Florence, Peter and his colleague access a map of central Rome that shows them the way to their hotel. Using the map, a taxi is able to take them to their hotel.

Multi-Access

Peter has checked in at the hotel in Rome, after a long working day. He wants to check the latest Swedish news from and see how his stock portfolio is developing. Peter logs his 3G terminal into a Swedish media portal and then connects the terminal to the television set in his hotel room. He can now view the Swedish news, which is being streamed to his terminal, on the TV screen.

The solution provided by Peter's colleagues in San Diego is now almost ready to be presented to the Italian client. Peter transfers the slide show to his notebook and then adds the final notes in accordance with his discussions with his colleagues in Lund. Two hours later, Peter is ready to present a comprehensive solution to his client.

Community Membership

Peter is about ready to sleep in his hotel room in Rome. Before falling asleep, he wants to find out what his son and wife have been up to in Sweden. He accesses their private community and retrieves a message from his wife including pictures of their living room with the new sofa that she has bought during his absence. He is also able to watch a recording of his son singing Christmas carols at school earlier the same day.

Peter has been out of office a couple of days and he is curious about the organization plans presented two days ago. By connecting to the department community he gains access to a slide show presenting the outlines, combined with comments from his colleagues.

Implications for the Service Provider

A situation-centric service approach in the 3G context will entail new challenges for service providers. In order to regularly launch new high quality services with high value for the targeted segments, it will be important for all 3G service providers to create a comprehensive platform for service aggregation and intelligent adaptive service architectures.

The 3G service environment will be complex, serving the same users in many different situations and roles with different responsibilities and needs. The same person will be a user at work, at home, with friends, when traveling with family, on business trips, or on the road in a motor vehicle or bus. There will also be simpler service environments based on user personalization and selections. In addition, the next step will comprise community-based services. All these issues have to be considered when deciding an overall service strategy.

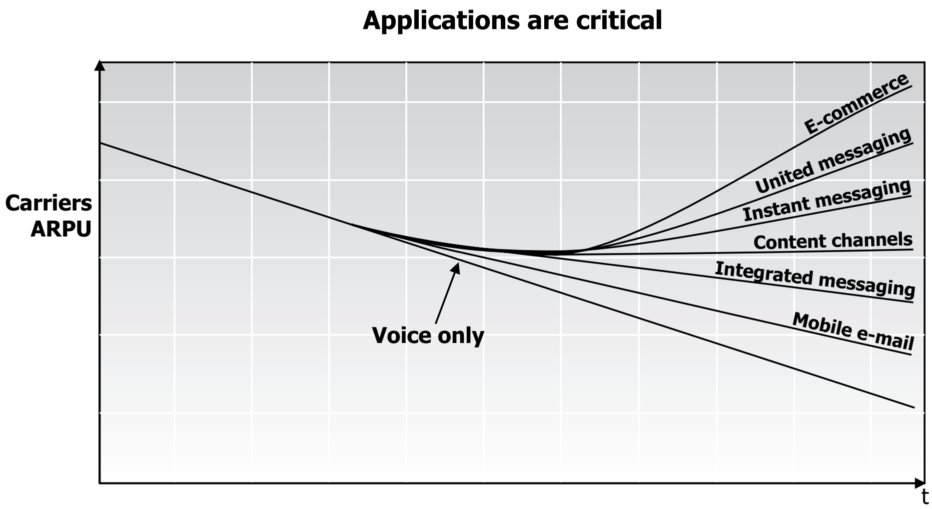

Figure 7. The need for new revenue

Business Models

The advent of the mobile Internet is about to drastically change the whole business logic of delivering mobile services. This change is driven not only by the shift from circuit-switched to IP-based bearers, but also by the Internet business paradigm with its abundance of free content. On top of that, mobile voice ARPUs are also declining and mobile operators are already looking at additional revenue streams from applications that manage traffic and provide alternative revenue types.

In this quite ambiguous situation, 3G players will start to form their business models with a collective responsibility to keep the heavily investing 3G network owners, on which they all rely, in profitable business.

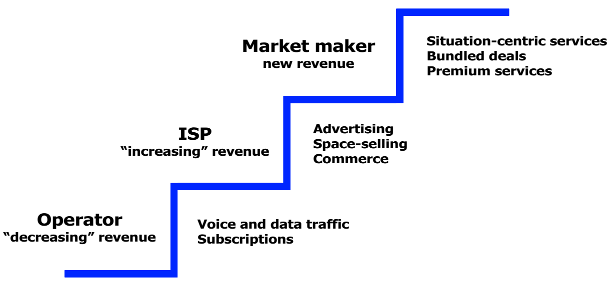

Figure 8. Revenue categories

We believe that it is of the utmost importance for players providing 3G-based communication services to establish a multitude of revenue streams from multiple sources, such as users, service providers, and advertisers. Traffic and subscription revenues will still be there, but declining. The prospect of advertising on mobile devices is a promising one when the location-based opportunities are considered. However, user acceptance is still an open issue and providers must be cautious not to scare people off by pushing commercial messages too excessively. The most promising type of service for new revenue is perhaps the situation-centric services that can increase user activity and willingness to pay a considerable amount. Such services will provide exactly what is needed for every situation, instantly creating savings and providing business advantages for users.

3G Billing

Uncertainty in the Marketplace

The operators that have been granted 3G licenses have in many countries, but not all, paid a great deal for them. Examples include Germany, where six licenses have been granted at a total cost of approximately US$ 50 billion, and the UK, where the licenses have been sold for a total of approximately US$ 35 billion (from an article in Dagens Industri, August 18, 2000). In addition to this, the cost for building, marketing, and administering the 3G networks is very high. In Sweden the costs are estimated to exceed US$ 1,400 and in Germany and the UK US$ 1,100 per user (from an article in Computer Sweden, December 18, 2000). This forces operators to introduce services to ensure rapid ROI. However, the high cost also forces them to obtain revenue from sources other than access charges, as user costs would be too high, which would prevent people from using 3G services.

The operators' roller-coaster ride has just started in terms of how to set different tariffs. These need to be balanced against the contradicting requirements of gaining new customers and earning enough revenue to survive the first years. Another factor is, of course, the other market actors, all of whom also want to gain as much intimacy as possible with the end users, top the value chain, and get the largest piece of the pie.

Competition

In the Internet world, several services are available. 3G can be regarded as just another access method to these services with, however, an important difference. A new range of practical features utilizing user mobility and positioning allow the creation of a new set of services impossible to implement in the fixed Internet. If these new features are not used, the 3G network will just act as an IP-pipe, and all the interesting revenue from the services are at the risk of being collected by other service providers. This is especially true if the role as payment provider is established in the service provider's context whereby the payment provider has an existing payment relation with the end-user, such as through a bank, utility company, or store debit card.

Regulators (like the EU) do not tolerate "locking" the user to the services provided "in-house", so the operator must find means to attract users on its own merits. One way of doing this is, for example, for the incumbent operator to provide branding with bundled services and attractive pricing and discounts, thereby raising customer loyalty. Mobile Internet means that the user is mobile - and not necessarily the terminal used at a particular time - providing the incumbent operator with diversified operations with an advantage over the new entrants. The operator must balance the tariffs cleverly in order to attract customers and to keep them using the value-added services provided in-house.

Co-operation

In the world of new services, an array of different roles can be identified. One model from the UMTS Forum, shown in *Figure 9*, describes the transaction flows. To maximize revenue and retain control, the operator would like to enter several of these roles, but also needs to coexist with other parties filling the same roles.

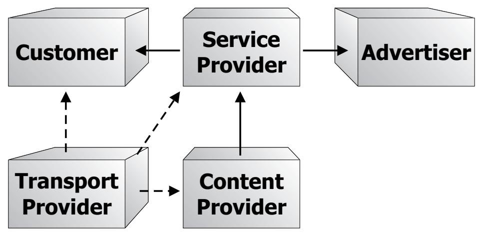

Figure 9. Transaction flows

To be able to share the revenue, inter-operator billing between the different parties is an area of vital importance, as no player will act alone. Service providers will emerge "everywhere", and the operator will need to find the means to cooperate with them. Information available in the operator's network, such as the position, direction, and speed of mobile users, the large customer base, user identities, and geographic patterns, is of great value to content providers. Some of the information is in direct conflict with integrity requirements and will perhaps not be available without written consent from users.

From an independent service provider's perspective, the transport provider is doing the same amount of work whether the value of that particular item of bit stream is high or low. The key issue is to ensure that all parties can trust each other by making the billing information available at both ends, ensuring a fair split of the revenue. Models where information is only available to one party will not gain market acceptance, so this must be considered very carefully.

Some emerging revenue sharing models include:

- Commission charges - the operator receives, for example, 10% of the revenue generated by the content provider. Simple business model, easy to understand.
- Airtime revenue share - the content provider receives a share of the additional traffic revenue generated in the operator's network.
- Payment provider model - the operator can take on the role of a trusted payment provider, aggregating payments from different sources and presenting one bill to the end-users. The content provider does not know the end-users' identities and has no direct relation with them.

Combinations of these models are also feasible, and will most probably be adopted.

Complexity

The billing interfaces in 3G offer many options that must be carefully considered by the operators. It will be possible to base billing on, for example:

- Time
- Destination
- Location
- Volume
- Bandwidth
- Quality
- End-user value
- Service elements in the Virtual Home Environment (subscriptions, service transfers, service upgrading, service usage, and roaming)

However tempting to design advanced billing tariffs, they must be balanced against the risk of confusing customers – which end-user is prepared to pay for Mbytes? The tariffs must be kept simple and the internal support systems must be able to cope with the massive amount of information generated within the networks. The billing systems must also be able to cope with different access networks (GSM, GPRS, UTRAN, HiperLAN2, etc.) and roaming between these networks.

The difference in quality of service is often singled out as a differentiator when comparing 3G with existing access networks. This is true and can be used if the service is accessed, end-to-end, in a 3G environment controlled by the operator. However, at present, the user accesses a service via the Internet, so no guarantees concerning quality of service can be made. The existing paradigm on the Internet is "best effort", which prevents billing based on quality of service.

Architecture

The UMTS billing systems will be divided (as are most telephony billing systems) into Mediation Systems, End-customer Rating and Billing Systems, and Interconnect Billing Systems. See *Figure 10*.

Mediation Systems are placed close to the network elements and obtain network session and naming information, which is then used to bill for the service. These systems can also provide content information.

The "traditional" airtime information is accessed via mediators.

End-customer Rating and Billing and Interconnect Billing retrieve information from different mediator platforms, so-called mediation systems. This information is then combined with the rating and billing rules, which are still to be defined and will therefore be inconstant in the beginning.

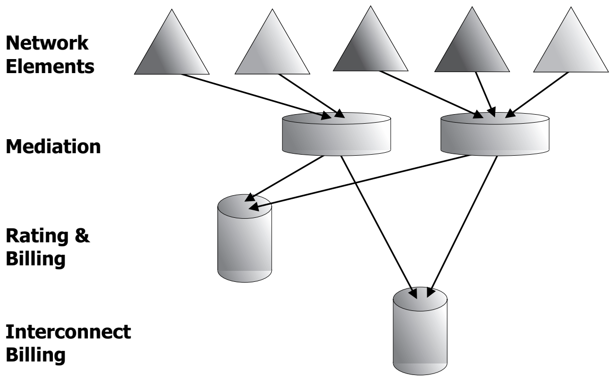

Figure 10. UMTS billing system

Content

One new aspect arising with 3G billing is the ability to bill for content, which adds another dimension to the billing challenge for operators, which have previously handled billing based on time and destination. Content billing has not been implemented before and existing systems were not designed with content billing in mind. However, the operators are forced to accept this challenge. If other actors in the value chain can better implement this, they will also gain customer loyalty, making operators mere access providers for others.

Content billing can be handled in the traditional way, according to the architecture in *figure 10*, where the content servers are regarded as network elements. The content servers produce simple events, such as the start of a streaming video clip, and the event is mediated into the rating and billing system. The operators either run the content servers themselves or enter a back-to-back agreement with a content provider on a static basis. This approach is network centric and requires modifications in the mediation and rating systems each time a new service is introduced. This architecture will be used in the first 3G installations, but will not be optimal when content offerings are expanded.

Alternatively, content information can be obtained via the emerging Internet payment solutions, creating a service-centric billing perspective. In this case, mediation is taken care of by a payment solution, which also provides payment and payment information. This enables a dynamic relation between the operators and content providers, where one content provider can sell its services to anonymous end-users, allowing payment to be handled by different payment providers. The emerging Internet payment systems also allow micro payments and innovative pre-paid solutions where, for example, the accounts can be topped up directly from mobile terminals. The Internet payment solutions provide a secure environment where the consumer and seller do not have a direct payment relation, as the payments are handled via a third party, the payment provider. See *Figure 12*.

If a service-centric billing approached is applied, services can be billed independent of the access method used at the time, assuming that the operator has taken on the payment provider role.

If the operator wants to price the services differently according to access method (for instance, a video clip costs more/less from a mobile terminal than a fixed terminal), then a combination of both network- and service-centric views needs to be applied, thereby integrating the two systems.

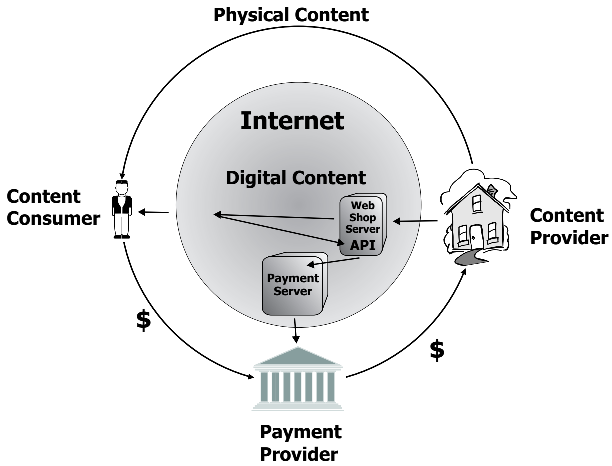

Figure 11. Jalda purchase overview

Challenges

For some time now, the selling points for new billing systems have been rating and billing plan flexibility, convergence, discount bundling, and real-time capability. The new solutions have often been hindered by limitations in the legacy systems, and the results have been disappointing in terms of achieving design goals. With 3G, the operator has to start from scratch and cannot compromise on any of the goals, as the billing system's architecture and functionality are critical for the operator's mission.

Mediation and/or Internet payment systems must be able to handle the large volumes of billing related

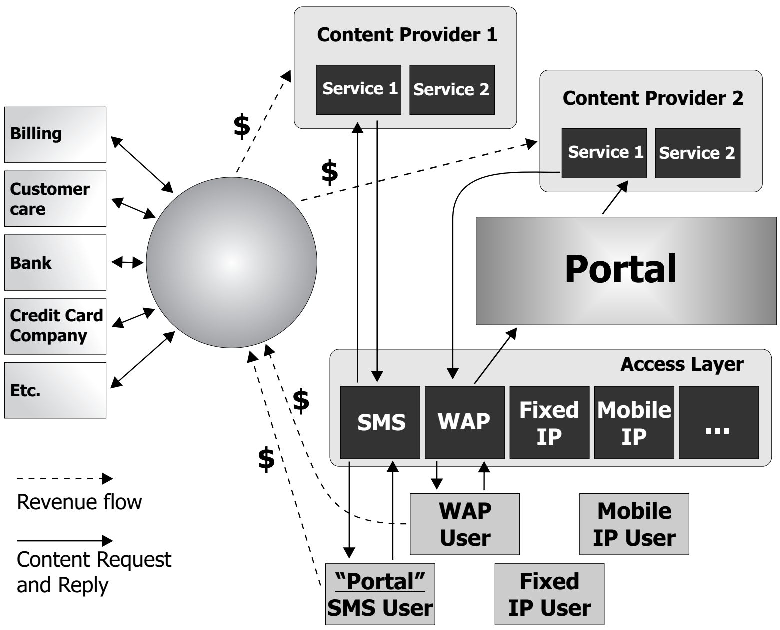

Figure 12. Service-centric content billing

information created by 3G networks, estimated to be 10-100 times the number of events in existing 2G networks. Efficient event filtering is crucial for success in this area. If Internet payment systems are used, they must suit the operator's business model, fulfill the reliability and performance criteria, and be trusted and accepted by all parties, including the content consumers.

The necessity to combine information from different sources in order to create billable records is a new requirement for the mediation systems.

Conclusions

Telecom billing is a multi-billion industry and the operators' billing processes have been developed over a long period of time and with considerable effort. The processes are devised to secure revenue for the company and streamline the work process in order to create cost-effective operations. The billing processes are part of the core business of any operator - and operators want to keep control in their own hands.

To change the processes and systems, and to include new elements and new thinking, is not a simple task. However, the legacy processes and systems are in practice impossible to modify to handle 3G billing, as nobody has the time to make an attempt. They have to be replaced.

The operators have to cooperate with billing vendors, system integrators, and consultants in order to achieve this within the time frame allowed and to attain ROI in the 3G networks.

3G Standardization

The 3G story began back in 1985 and is full of interesting techno-politics. The first major milestone was obtaining a global frequency band, without which there would be no global mobile system. The selection of the radio interface and the three "eras" in the core network arena comprise the other major milestones.

The Search for Spectrum

ITU-T (at that time CCITT) began work in 1985, the needs of the developing world were a clear objective and probably the major cause for commencing work at all at that time. The focus on the developing world then quickly declined. All assumptions at that time were built around the ISDN paradigm: a multi-service network with separated signaling that would reside in a new frequency band with all the capacity and services you could possibly need. This would replace older mobile systems.

Until 1990, work was concentrated on gaining consensus for a common set of objectives and motivating global spectrum. At the global spectrum conference WARC-92, a set of global frequency bands (the FPLMTS bands) was identified and work on system specifications could begin. In 1995, WRC-95 defined this band in greater detail and WRC-2000 in Istanbul managed to find some additional frequencies.

Early frequency spectrum calculations already showed that the bandwidth issue would prove problematic. In fact, there was no more bandwidth available other than that for GSM1800, but several of the targeted services would require ten or one hundred times the bandwidth of a normal voice call.

Around 1995, a movement started around "migration" and "development" with three objectives:

- Vendors should develop their 3G products as a further development of the 2G product platforms.
- Operators should re-use the existing installed equipment base.
- Operators should be able to use the 2G and 3G bandwidths together in order to meet the overall capacity requirements.

Consequently, the UMTS standardization in Europe became GSM-centric, resulting in multi-mode operation GSM-UMTS.

Would existing GSM operators be able to acquire UMTS frequencies, or would regulators prevent this in order to enhance competition? The tendency in the UK, where British Telecom was initially held back from a GSM license, served as a warning. The development of GPRS and EDGE provided fallbacks for GSM operators (and a development path for TDMA and GSM networks in the US).

An auction in UK was devised in order to provide the government with the maximum possible payment for the designated frequency spectrum. The UK government thus potentially set mobile Internet development years back by demanding investments in fees rather than networks. A similar development took place in Germany, but at the end of 2000, the demand for overpriced UMTS bandwidth in other countries had dropped.

Figure 13 shows the 3G (IMT-2000) bandwidth including the new frequencies identified by WRC-2000: The 1710-1885 MHz, 2500-2690 MHz, and 806-960 MHz bands have not yet been allocated by the authorities in any country for UMTS. These bands provide the longer term possibilities for expanding UMTS services.

The Selection of a Radio Interface

The European Framework project Codit came to play an important role in the selection of a radio interface. Ericsson and Nokia, among other companies, participated in this project, which led to the early foundation of expertise concerning W-CDMA and, not least, a set of key patents.

In Japan, NTT DoCoMo planned to specify a W-CDMA system. Very often, historically, NTT have been able specify their own systems and solutions and are a large enough corporation to ask vendors to produce what they need. This concept has been very successful in the local Japanese market. NTT DoCoMo's decision and their pressure on global vendors at an early stage had a crucial impact on the outcome.

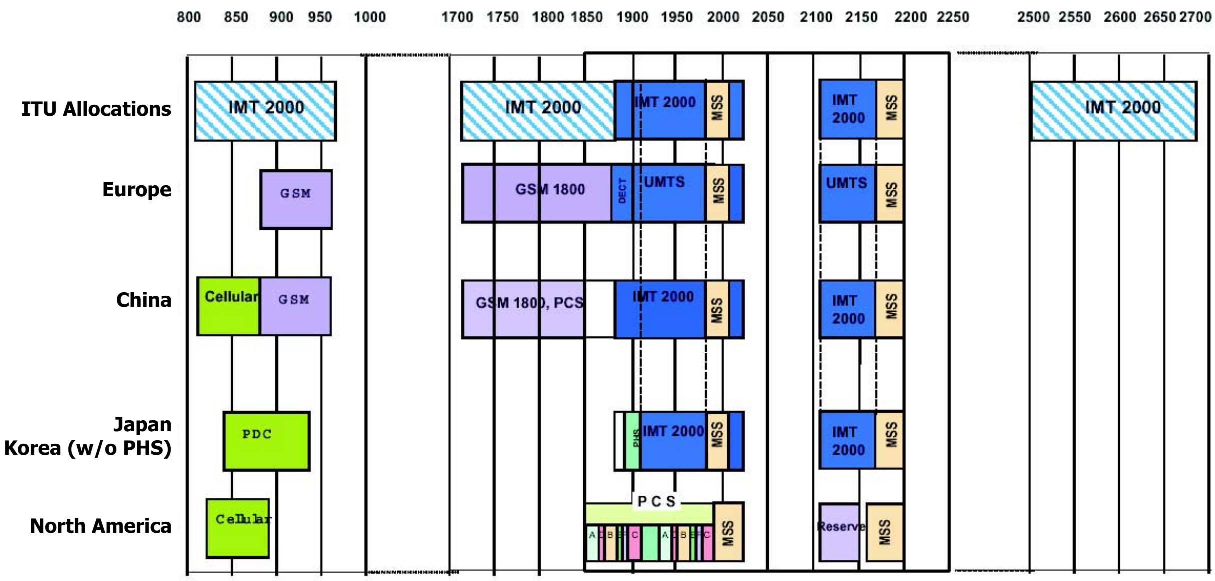

Figure 13. Worldwide 3G frequency allocation

For a while, the 3G arena was anything but global. Japan had selected W-CDMA and parts of the US chose to develop IS-95 toward multi-carrier cdma2000. Europe was yet to decide. All players seemed to override the ITU.

In Europe, several alternatives emerged. The major contestants were W-CDMA (lead by Ericsson & Nokia) and TD-CDMA (lead by Siemens & Alcatel). Ballots were held in ETSI TC SMG at the end of 1998/ beginning of 1999, but no single formal winner could be announced. W-CDMA received a majority, but not the absolute majority required to win. A compromise directing W-CDMA toward the FDD bands and TD-CDMA toward the TDD bands was made. Harmonization between the two modes followed.

The Chinese were pushing for another TDD solution, called TD-SCDMA, to be brought into the 3G family of systems.

With the choice of W-CDMA in Europe, Asia and among selected US players, together with adoption of specific TDD modes, a global solution was found. UMTS will be the first global mobile system in the world.

Core Network Strategies

As work on 3G commenced, the B-ISDN and IN paradigms were fashionable. Much effort was put into this architecture during the years 1990-1995. Some terms from this period, such as “service capabilities”, are still in use.

The targeted architecture was, however, not compatible with events in the GSM world. GSM became a huge global success during these years and the GSM-centric way of thinking gained in popularity and importance. Following the ITU-R and ETSI SMG migration reports, the situation changed rapidly. In fact, a large set of specifications was abandoned and a new set based on the GSM/GPRS core network architecture was adopted. The Japanese also accepted this core network solution as a compromise once ETSI had selected the W-CDMA radio access method.

Another major player, AT&T Wireless, made a significant footprint on the UMTS scene in the spring of 1999. From a TDMA network perspective, the company was looking to secure a migration plan for their network. A group of operators invited some of the leading vendors to form a strategy called “All-IP”. This name was adopted by the market very rapidly, but the issue at hand was the implications for the architecture. Marketed as a revolution, the GPRS architecture was enhanced with multimedia capabilities (based on the SIP protocol) and the traditional GSM MSC structure (optionally) was detached from the network. This meant that all services, even the traditional circuit-switched services such as voice calls, could be produced via the GPRS platform in Release 4 and 5 of the UMTS system. This would provide greenfield operators as well as operators evolving from a TDMA architecture with the possibility to move toward EDGE or W-CDMA by investing only in the GPRS part of the GSM/GPRS core network. Moreover, exiting operators would have the possibility to migrate their traffic from MSCs to IP-based xGSNs whenever business required it.

A Global Approach is Resumed

The original intention of the ITU was to review a number of proposals and devise a set of radio interfaces characterized by a maximum of common denominators. Following US PCS auctions, Japanese W-CDMA decisions, and GSM migration plans in the middle of the 1990s, it really did not seem probable that a global system would be the outcome of the process.

Following negotiations between Japanese and European players (W-CDMA and GSM/GPRS core networks), a large part of the world was again united.

Globally, the Operators Harmonization Group (OHG) contributed significantly to a worldwide harmonization solution, allowing both cdma2000 and W-CDMA radio access networks to be connected to both IS-41 and GSM/GPRS core networks, subject to operator preferences. Global roaming between all types of networks followed from the same decision. The world's largest operator, Vodafone-Airtouch, using GSM in European networks and IS-95 in the US, was one of the beneficiaries of this solution.

IMT-2000 could again be called a global system.

3G Standardization Today

There are several standardization bodies and special interest groups involved in the standardization of 3G. Even if they have different cultures, organizations, and working procedures, they have the same overall goal: to realize the wireless mobile Internet vision.

The two most important standardization groups, 3GPP and IETF, stem from different industrial cultures. IETF comes from the Internet world and 3GPP from the telecom world. Both their working procedures and fundamental philosophies differ. With the merger of Internet and mobile wireless, 3GPP and IETF are obliged to cooperate.

3GPP

The 3GPP is not a legal entity but a joint project between regional standardization bodies from Europe, North America, Korea, Japan, and China. Since the 3GPP has no formal power, all standards have to receive official approval from the founding members. This is, however, merely a rubber stamp process.

The goal of the project is to finalize and further develop the UMTS specifications. This encompasses the W-CDMA and EDGE radio access networks, the backbones, the terminals, the service enabling nodes, and the management functionality. To do this, 3GPP has a project coordination group (PCG) that controls the project and five technical specification groups (TSG) that are each responsible for a particular technical area. The TSGs are divided into smaller working groups that perform the actual standardization work. *Figure 14* outlines the 3GPP organization with the PCG on top and the TSGs and their subgroups below.

The work procedure of the 3GPP begins with the formulation of specifications. These are first approved within the working group and then sent for approval on higher levels. If errors and/or enhancements are identified for an approved document, formal alteration requests have to be drafted and approved.

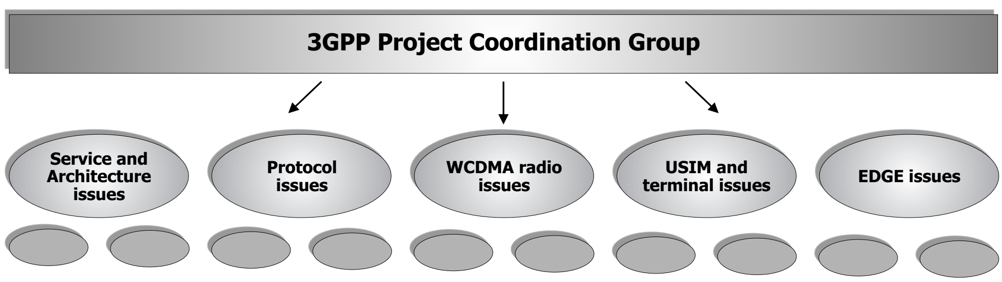

Figure 14. The 3GPP Organization

IETF

The IETF defines themselves not as a standardization body but as a large open community of individuals with a common interest in the development of the Internet.

The IETF is responsible for the protocols, both mobile and fixed, used on the Internet today. The group is organized in a similar way to 3GPP, with a controlling board at the top and working groups clustered into technology areas such as “Security” and “Routing”. Currently, there are approximately 130 different working groups within eight technology areas.

Within IETF, the work procedure is different from that within 3GPP. Basically, there are two document types, the Internet Draft and the RFC. Anyone with what they consider a good idea may submit an Internet Draft and present it at an IETF meeting, as long as it is in the correct format. The Request For Comments (RFC) is the finished standard, although there are several different kinds of RFCs with different statuses. To make an Internet Draft into an RFC, independent implementations of the standard in question must show interoperability at testing gatherings called bake-offs. This hands-on work method contrasts with the 3GPP work procedures. Of course, the RFCs also have to be approved by the IETF. Moreover, note that IETF does not alter approved RFCs; if an RFC needs updating, a new RFC making the old RFC obsolete is created instead.

3G Technology

3G technology isn't just a new radio interface. Even if the fixed backbones are based on an evolution of second-generation technologies such as GSM/GPRS, the overall picture is quite a bit more complex than that which 2G presents. 3G is not just a matter of building a totally new infrastructure, but also entails having migration plans ready, both for the short term and the long term. Existing 2G operators will also have to juggle different radio access networks.

The 3G terminals will also have new capabilities that 2G terminals do not. Multimedia messaging and other multimedia services based on standard Internet technology will bring the worlds of cellular telephones and the Internet closer together. Users will be able to download Java-based applications to their terminals, just as they do today on the World Wide Web. Of course, with new possibilities come new security threats and even if security is also enhanced, the risk of security lapses will be unacceptably great unless extra precautions are taken.

3G Networks

UMTS is both packet and circuit switched and a further development of earlier public mobile systems such as GSM and GPRS. In addition, by the end of 2001, IP multimedia will be an integral part of the UMTS standards. *Figure 15* shows the various domains of UMTS and some external networks.

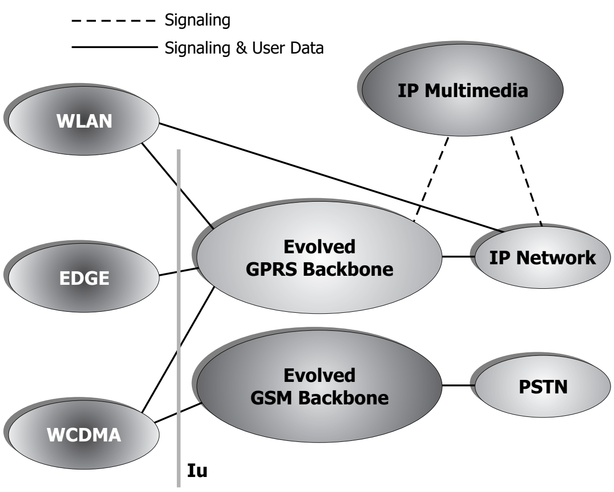

Figure 15. The UMTS networks and domains

The circles to the left show three different radio access networks, which are attached to the backbone networks in the middle via the Iu interface. A Wireless LAN (WLAN) need not be connected to the GPRS backbone but could also be linked directly to an IP Network. The upper circle to the right depicts the SIP based and access independent IP Multimedia subsystem. To the far right are legacy PSTNs and an external IP network (e.g. the Internet).

The Radio Access Networks

One of the main design criteria of UMTS is that different radio access networks be attachable to the backbone via a common interface, the Iu interface. In *Figure 15*, three radio access networks are shown - all of which can be used to access UMTS services.

- Wideband Code Division Multiple Access (W-CDMA). When most people refer to 3G or UMTS radio access they mean W-CDMA. W-CDMA uses the 2 GHz frequency band, the bandwidth for which operators e.g. in the UK and Germany, paid large fees. In theory, W-CDMA can reach data transfer speeds up to 2 Mbit/s, but in reality transfer speed is closer to a couple of hundred kbit/s, the actual bit rate depending on factors such as cell load and mobility of user (see *Figure 13*).
- HiperLAN/2 is a Wireless LAN (WLAN) technology that offers very high bandwidth (~25 Mbit/s to the user), new advanced radio functions, mobility support, a strong security infrastructure, and guaranteed quality of service. It uses the 5 GHz frequency band and the service range for one access point (or base station) is typically 30 m indoor and 150 m outdoor, thus making wide area coverage a practical impossibility. There are no HiperLAN/2 products on the market yet but several major vendors have stated that they will have products out before the end of 2001 and some have officially demonstrated prototypes.
- EDGE is an enhancements of the "2.5G" radio access GPRS, which in its turn is a development of the 2G circuit switched GSM. EDGE and GSM/GPRS are so closely coupled that the cell structure can be left intact. EDGE can, in theory, reach data transfer

speeds of up to 384 kbit/s, but the real speed is most likely lower.

Figure 15 illustrates how the W-CDMA radio access network is attached both to the evolved (circuit switched) GSM backbone and the evolved (packet switched) GPRS backbone. UMTS has a packet switched and a circuit switched domain. This white paper deals primarily with the packet switched domain as it is predicted that the majority of new and innovative services will be developed for that domain.

The circuit switched capabilities of UMTS are of use to an operator that doesn't own a circuit switched 2G PLMN (such as GSM) or needs more bandwidth for voice services. An existing GSM operator that deploys a W-CDMA network will probably use the GSM network for standard voice service and the W-CDMA network for new data and multimedia services. GSM operators need to have a plan for how to gradually move away from 2G voice services, either to 3G circuit switched or packet switched IP multimedia.

It will be difficult to build a radio access network with sufficient capacity using only W-CDMA technology. If the vision of the mobile Internet revolution is fulfilled, the bandwidth in certain areas, hotspots, will most likely exceed the supply substantially. It is a good strategy to not look at W-CDMA as the only 3G access form. An operator will benefit from considering WLAN (e.g. HiperLAN/2) service in hotspots like airport lounges, train stations and conference hotels. One unsolved problem with this strategy is how to support roaming and hand-over when users move from WLAN service to W-CDMA service or vice versa. There are two approaches to solving this issue, either by using the Iu interface or by dealing with mobility over IP. Using IP has the advantage of being very general, since IP and IP functionality would be the "least common denominator" for every access form. On the other hand, performance problems would be an issue with today's techniques, hand-over, for example, (with Mobile IP, RFC 2002) could not be achieved.

The Backbones

The packet switched backbone has evolved from the GPRS backbone. Many of the basic functions are identical, but improvements have been made. At the bottom of the protocol stack, IP is still to be found. Also still in the stack is a special tunneling protocol used to encapsulate user data, perform signaling, and handle mobility tasks. See *Figure 16* for a closer view of the packet switched UMTS backbone.

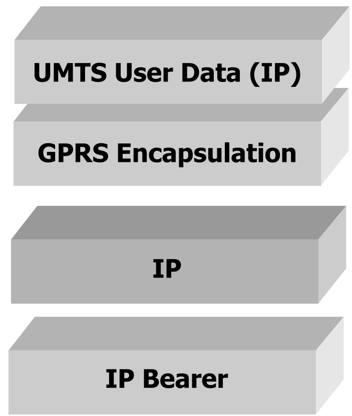

Figure 16. Close view of the packet switched UMTS backbone

At the top of the protocol stack in *Figure 16* is the UMTS user data, i.e. IP data. This user data is encapsulated in the GPRS specific tunneling protocol mentioned earlier. At the bottom is the IP bearer layer. This IP bearer could be a Gigabit Ethernet or ATM, for example. It is up to the UMTS operator to select and use the IP bearer that is most suitable for its specific requirements and existing architecture. The separation of the layers allows one layer to evolve independently from another. In this way, the IP bearer can be upgraded without having to adjust the upper layers.

The IP layer in the middle of *Figure 16* need not only be used for UMTS data transport. Great benefits can be experienced by reusing the same IP backbone for several different traffic types. The same IP backbone could, for example, be used for an operator's UMTS transport and ISP/broadband services.

The evolved GSM backbone is still circuit switched, but in later releases of UMTS this fundamental architectural principle will change. *Figure 17* shows the evolution of the circuit switched backbone where the MSC and GMSC are broken down into servers and media gateways.

At the top of *Figure 17* is the first version of the circuit switched UMTS backbone. From a high architectural point of view it is identical to GSM.

In the evolved backbone, at the bottom of *Figure 17*, the MSC has been divided into an MSC server and

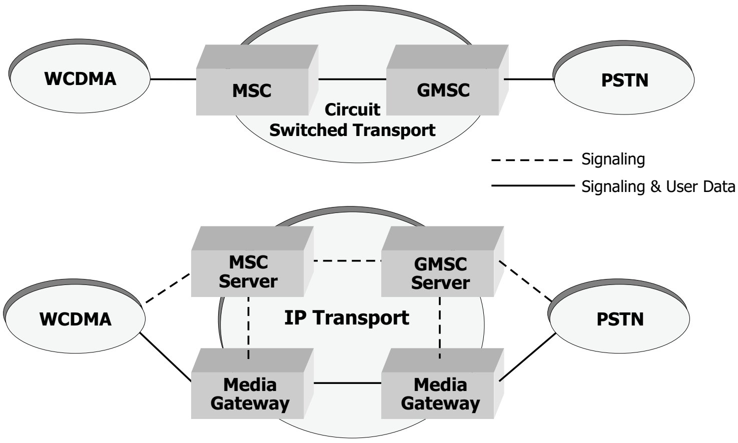

Figure 17. The evolution of the circuit switched UMTS backbone

a media gateway. The MSC server handles all the signaling and controls the media gateway. This allows it to know what circuit switched trunk is mapped onto what IP address/port number and vice versa. The media gateway moves the media (e.g. voice) from a circuit switched bearer to an IP bearer, and may also have additional functions such as echo canceling and transcoding. Both user data and bearer control signals are transported between the two media gateways. The bearer control is used for setting up/dismantling media channels and to negotiate bandwidth usage, for example. The GMSC is broken down in the same manner as the MSC, with a GMSC server and a media gateway.

By separating the user plane and the control plane, two things are basically gained:

- The media translation makes it possible to use a common IP network for both packet switched and circuit switched data. This IP network could also be used for other traffic types at the same time.
- The unit-based architecture enhances flexibility and scalability, as it allows one MSC- or GMSC-server to control several media gateways. In addition, signaling and user data capacity can be scaled independently.

This disassembly of circuit switched architecture is one aspect of the popular “all-IP” vision that is currently being widely discussed and implemented in the wireless mobile industry. Another part of that vision is the IP multimedia subsystem, something that will result in significant added value for the end users.

The IP Multimedia Subsystem

The IP multimedia subsystem (or IM subsystem) is a service network that only handles signaling, i.e. actual user traffic takes a different route and do not traverse the IM subsystem. The idea behind this is to enable multimedia and new innovative services, not only to 3G subscribers, but also to users of different IP access networks. To follow the idea of the mobile Internet and facilitate service development and inter-operability it is important to merge with standard Internet protocols. However, inter-operability must be balanced with scalability requirements and the need to make the system “future proof”. Bearing this in mind, the IM subsystem was designed with the following primary characteristics:

- Based on standard Internet specifications (e.g. SIP, see below).
- Uses the next generation of the IP protocol (IPv6).
- Access independent, i.e. fixed and/or mobile access.

The Session Initiation Protocol (SIP) was developed by the IETF and is used for handling interactive sessions on the IM subsystem, the Internet, or another IP network. An interactive session could be, for example, a video conference or an on-line game.

SIP is not concerned with the content or type of a session, it merely finds the endpoints and carries information concerning the kind of session to be set up, ended, or added to. It is up to the particular application that uses SIP to handle the media stream, supply a GUI to the user, maintain an address book, and so on. Since it was developed by the IETF, SIP uses the Internet procedures and resources, such as DNS and MIME coding. One user advantage arising from this alignment with Internet architecture is that a user's SIP address could be identical to his email address.

The basic design of SIP is constructed so that most of the intelligence in the endpoints, which is in keeping with the Internet paradigm. However, there are two types of SIP servers that can be used to add SIP functionality inside an IP network, the redirect server and the proxy server. The IM subsystem uses proxy servers to make it possible for an IM domain operator to offer multimedia services and roaming capabilities to its subscribers.

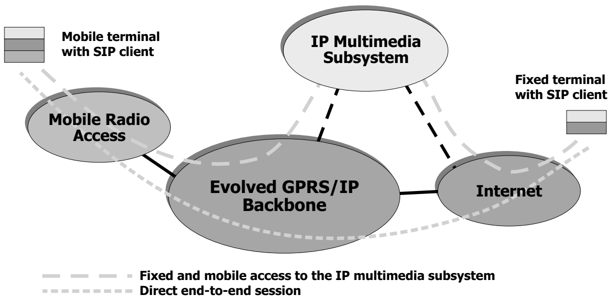

Figure 18. The IM subsystem

Since the IP multimedia subsystem is access independent, it provides an excellent opportunity to achieve service mobility, see *Figure 18*. By implementing new, and some old, services in the IP multimedia subsystem, these services would in principle be accessible wherever there is IP connectivity.

A future UMTS subscriber will always be able to do anything end-to-end and thus bypass the IP multimedia subsystem completely (see *Figure 18*). For an IM subsystem operator this is crucial. To be successful it is imperative that an operator provide an environment that is attractive to 3rd party service/content creators. The aggregate of 3rd party services/contents is what the operator sells to the subscribers. An attractive environment in this context means access to customer stock, joint marketing efforts, access to charging & billing functionality, and guaranteed quality of service.

In the IP multimedia subsystem, the next generation of the IP protocol (IPv6) is used exclusively. One of the major advantages of this is the abundance of IP addresses available. This abundance is necessary in order for a wider range of terminals to be "always on". A less apparent advantage is that IPv6 facilitates the use of SIP in the long term, since publicly routable IPv6 addresses can be used end-to-end. It is important for operators to have a reviewed IPv6 strategy, i.e. address allocation and interworking functions.

The standardization of the IP multimedia subsystem allows operators to optionally perform all service for roaming subscribers, both control and execution, in the home network. *Figure 19* shows how home control of services works.

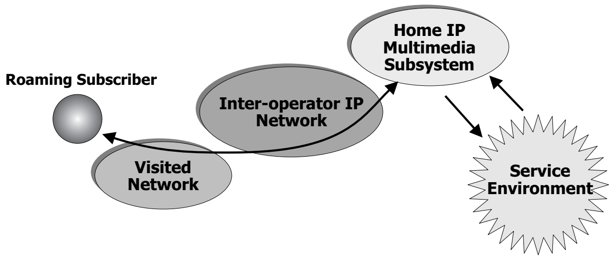

Figure 19. Home control of services

The roaming subscriber in *Figure 19* has attached to a visited network, with which his home operator has a roaming agreement. To access services, signaling (i.e. SIP signaling) is transported to the home domain of the subscriber and service is executed in the home domain. From a service point of view, the subscriber is still in his home network with the exception that the visited network always has the final say when it comes to resource allocation. The advantage to the subscriber is that he can use the same services he is used to, with the same interface. The home operator can rapidly develop and deploy new services without having to wait for standardization since the service interfaces to the service environment and subscribers are internal.

3G Terminals

3G terminals will be multi-mode, i.e. be able to support more than one radio system. Multi-mode terminals differ noticeably from multi-band terminals. A multi-band terminal (like GSM dual-band) can handle different frequency bands but not different radio systems. The first 3G terminals for the European market will most likely be dual-mode and dual-band. They will have a GSM call control stack, session management, and radio interface implemented together with the corresponding UMTS terminal functionality (i.e. an

GSM/GPRS/900/1800/UMTS terminal). 3G terminals will also have to be multi-mode to support global roaming, i.e. be able to handle the W-CDMA and cdma2000 modes.

W-CDMA itself has two different modes, the Frequency Divisions Duplex mode (FDD) and the Time Divisions Duplex (TDD) mode. The distinction lies in how the uplink and downlink is handled. In the FDD, the downlink uses a different frequency than the uplink, and the TDD mode uses the same frequency but multiplexes on time instead. The first W-CDMA terminals on the market will be FDD terminals without support for TDD. This is simply due to the fact that FDD development is at a more mature state than TDD. It is more difficult to build wide area coverage with TDD, but easier to cater to asymmetric traffic than it is with the FDD mode. One possible future scenario is that operators choose to use FDD for outdoor service and TDD for indoor service.

Security

GSM security was only slightly improved with the introduction of GPRS. Essentially, GPRS uses the same security mechanisms but with new algorithms and slightly longer cryptographic keys. In UMTS on the other hand, the security architecture has new and improved security features and cryptographic algorithms. The cryptographic algorithms are also public, and the publication of the algorithms is believed to increase the credibility of the UMTS security. In UMTS, keys of twice the key length as in GPRS are used (128 bit), making crypto attacks considerably more difficult to carry out.

UMTS security can be broken down into four parts, see *Figure 20*. The controlled access to the mobile terminal is called Terminal Security. This concept comprises User/USIM authentication (e.g. entering of PIN code) and the ability to lock out undesired USIMs from a certain terminal. Both these features work the same way they do in GSM/GPRS:

Network Security is about securing signaling and data through the wireline part of the UMTS network, something that is particularly important to roaming subscribers. Very little about Network Security is standardized, but if IP were used as transport, then the IPsec standard from IETF would be a strong candidate for a solution to this problem.

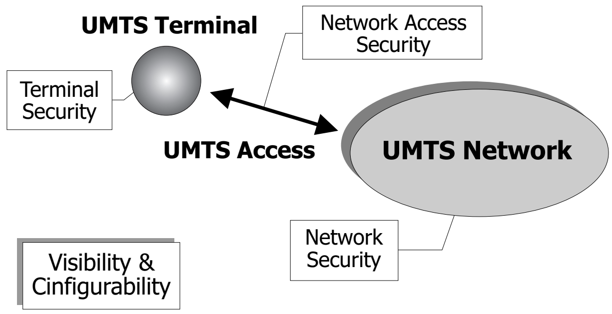

Figure 20. UMTS Security

The visibility property of the UMTS security comprises a set of features that make it possible for a user to find out whether a certain security feature is in operation or not; i.e. if encryption is used. Configurability is what enables the user to decide whether the use and provisioning of services should depend on a certain security feature or not, rejecting incoming non-ciphered calls or rejecting the use of a certain encryption algorithm, for example. Visibility and configurability are important UMTS concepts that do no exist in GSM/GPRS.

Most of the security work in 3GPP has been done on the Network Access Security, the largest security area to cover, and has several features:

- User identity confidentiality. This hides the user's permanent identity and current location. Achieved by the use of temporary identities like in GSM/GPRS.
- User untraceability; protects a user's integrity by making it hard for an attacker to deduce what services are utilized. This feature is achieved by changing temporary identities often.
- User authentication; used to allow network access to authorized users (subscribers) only. Works as in GSM/GPRS with shared secrets and challenge-response, but with improved algorithms and key lengths.
- Network Authentication. A means for the user to authenticate the network. This is a completely new feature that eliminates the threat of a radio channel being hijacked by an intruder. An encrypted sequence number is included in the authentication request from the network, and if this sequence number is not in keeping with what the terminal has stored, the authentication request is rejected.

- Data and signaling confidentiality. The encryption of signaling and user data. Highly sensitive signaling and all data are encrypted in GSM/GPRS as well, but the algorithms in UMTS are improved and the keys are longer.
- Data and signaling integrity. Signaling and user data are not only encrypted, but also protected against such things as reply attacks and tampering. This integrity is achieved by using the same sequence numbers for network authentication and by encrypted authentication codes appended to the data and signaling frames. GSM/GPRS has no integrity protection and relies solely on encryption.

A UMTS user must determine the level of security required to meet his specific needs. For the private consumer that only wishes to use the Internet casually, the security provided by the normal UMTS features is enough. For banking services, m-commerce, and professional use, end-to-end security is required.

If a company wishes to use UMTS to access form their corporate intranet, it is crucial to keep virus protection up to date and preferably to also install personal firewalls in the UMTS terminals. Having a clear security policy and making sure that all co-workers realize its importance is even more crucial.

UMTS (and GPRS) terminals will exist in an Internet environment with all its possibilities but also with all of its dangers. In principle, all that can currently happen to a fixed host attached to the Internet could happen to a UMTS terminal in the future, but the consequences would probably be more severe. For example, the radio interface is expensive and sensitive to denial of service attacks, so email viruses like the infamous "Melissa" or "ILOVEYOU" viruses would do major damage in an UMTS environment.

3G Messaging

According to the Swedish regulatory authorities of the telecom industry (PTS), approximately 160 million SMS messages were sent in Sweden during the first half of 2000. This is more than what was sent during all of 1999. The turnover for the Swedish SMS market during the first six months of 2000 was 32.5 M euro, comprising a 4 % stake of the entire mobile services market.

This surge of simple SMS message traffic in Sweden, and across Europe, hints that one of the Killer Apps (a fundamental and successful application) in a future 3G environment will likely be some kind of messaging, an evolved form of SMS.

IETF Instant Messaging and Presence Protocol

IETF is currently conducting work on an "Instant Messaging and Presence Protocol" in the IMPP work group. One example of a current presence service is the immensely popular ICQ, which allows subscribers to see when their friends are on-line. An example of a current instant message service is of course the SMS service or sending small text messages to other players when playing an interactive on-line game. It is important to realize that even if they work well together, presence and instant messaging are two separate issues. One does not need to have a presence service to do instant messaging and vice versa.

The model of the Presence Service part of the IMPP is shown in *Figure 21*.

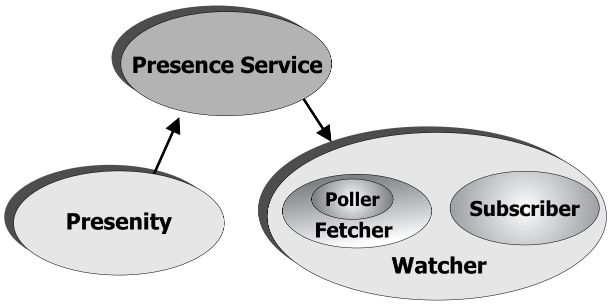

Figure 21. The IETF Presence Service model

The Presence Service has two kinds of clients, the Presentity and the Watcher. Even though these are logically separated they may very well be implemented together. There are also two different kinds of Watchers, the Fetcher and the Subscriber. The Subscriber requests notification of future changes to one or several items of the Presentities' presence information. The Fetcher does not receive any notification automatically, but must ask for updates on presence information. The Poller is a special type of Fetcher that makes periodic requests for presence information.

When a Presentity changes its presence information, it notifies the Presence Service, which in turn notifies all

Subscribers that have requested notification of changes to that specific Presentity's presence information. The Fetcher has to explicitly ask the Presence Service to receive an update.

The IETF also has a simple model for the Instant Message service, which consists of a Sender and an Inbox.

One way to implement the IETF IMPP could be with SIP, extensions to SIP and XML. There is currently work being done in the IETF IMPP work group along these lines.

XML could be used to represent presence information. This new XML format would describe a range of contact means for a certain user, such as email or an Instant Inbox address, and an indication as to whether the user is reachable via any of these means. For example, you can send an email to a user at any time (almost), but the Instant Inbox would only be open when the user is on-line.

To send presence information between the Presence Service and clients, there is a proposal to use SIP and existing SIP architecture, but with new extensions defined for it. The reason for using SIP to handle presence information (which is outside its original scope) is that it already uses similar information to find the endpoints needed to set up an interactive session. There are well-defined methods of sending this information from an endpoint to a SIP server. Enhancing SIP to distribute presence information from the Presence Service (SIP server) out to the endpoints is a relatively simple matter.

SIP has also been proposed as the core of the Instant Message Service. Again SIP is being used for something it wasn't originally designed to do. The general motivation for this usage is that sending an instant message and sending an initial set up request for an interactive session are very similar in nature. In both cases one needs to find the other endpoint and/or receive an error message if it isn't reachable. Another advantage of using SIP is that it can carry any data that is MIME codeable, such as text, images, sound, and video.

Placing the IETF IMPP in perspective against the IM subsystem paints a particularly attractive picture. The same architecture can be rather easily used to implement a wide range of communication services, regardless of IP access form. The user can use the same address (e-mail address) for all his communication needs and the different means of communication can be seamlessly forged together, as long as there is IP access available.

Multimedia Messaging Service

3GPP has a framework for a function called Multimedia Messaging Service (MMS), and even though it's a 3GPP standard, it can also be used in a GPRS environment. Unlike the IETF IMPP, MMS also includes interworking between networks and translation between different message formats.

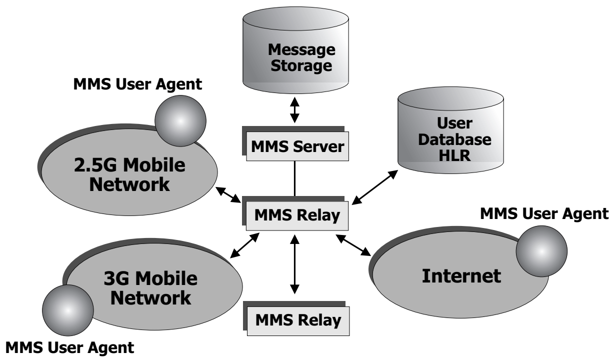

Figure 22. The 3GPP MMS Architecture

The MMS user agent carries out the notification, presentation, and composition of incoming multimedia messages to the user, and may also have optional features such as encryption and the handling of external devices like cameras or microphones. At a minimum, the User Agent must be able to handle plain text, but it is recommended that other media formats widely used on the Internet like MP3, JPEG and QuickTime also be supported. The user agent has an interface with the MMS Relay trough which to send and receive messages than can be based on WAP.

The MMS Relay is the heart of the MMS environment. All multimedia messages go via this node. Thus, one of its most important functions is to generate charging data. The MMS Relay keeps track of subscription profiles via the HLR and controls access to the MMS environment. It also does media conversion and address translation and has filtering and personalization features. In other words, the user should be able to decide what kind of messages he wants to receive, when to receive them, and from whom to receive them.

Messages that are waiting to be delivered, or for some other reason have to be stored, are kept in the MMS Server and Message Storage. The MMS Server can e.g. be an e-mail server or an SMS-center. Depending on the business model, the MMS Relay and MMS Server may be combined or separate.

When a multimedia message is sent across different administrative domains, or for some other reason sent between two MMS Relays, the Simple Mail Transfer Protocol (SMTP) is used. Between the MMS Relay and the MMS Server, HTTP (which is used on the WWW), SMTP, or some other transfer protocol can be used. The interface between the MMS Relay and the MMS User Agent is based on WAP, as specified by the WAP Forum.

Service Execution in Terminals

Handheld 3G terminals, and 2.5 G terminals as well, will have several tools available with which to enable and execute a wide range of new services. The flexibility that comes with these new tools is immense, but also poses new security threats. WAP still exists in 3G, but will be complemented by new Java technologies and Internet protocols such as SIP (see section *The IP multimedia subsystem* on page 20).

Java Technology in 3G

To diversify the Java technology in a controlled manner, Sun categorized it in three editions that all have different capabilities and are aimed at different kinds of platforms. They are the Java 2 Platform Standard Edition (J2SE), Java 2 Platform Enterprise Edition (J2EE), and the Java 2 Platform Micro Edition (J2ME). It is the J2ME and its refinements that are interesting with regards to handheld 3G terminals.

The J2ME has two separate configurations, the Connected Device Configuration (CDC) and the Connected Limited Device Configuration (CLDC). The CLDC is designed for smaller connected devices with limited processing power, very constrained memory, and battery operation. The CDC must run on a device with a complete implementation of the Java Virtual Machine, it is made for higher end connected devices like set-top boxes and web-enabled phones. To quantify, the CDC requires 256K of RAM and 512K of ROM while the CLDC needs 32K and 128K respectively.

The configurations of J2ME (CDC and CLDC) consist of a small sized virtual machine and some core APIs suitable for the type of devices the configurations are designed for. On top of the CDC and CLDC are different profiles to complement the Java functionality required for a specific set of devices. One of the more interesting profiles for handheld 3G terminals is the Mobile Information Device Profile (MIDP). *Figure 23* shows where the MIDP fits in the Java world, as well as illustrating the Personal and PDA profiles.

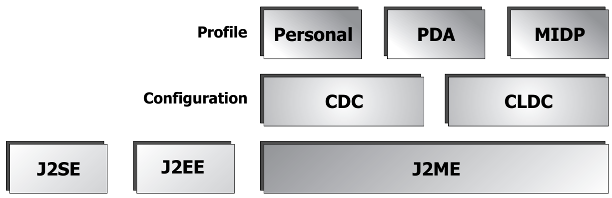

Figure 23. Java Editions, Configurations and Profiles

MIDP, together with CLDC, provides a set of APIs and a small sized virtual machine that is suitable for implementation in handheld mobile devices such as the 3G terminal. The functionality provided by MIDP and CLDC to enable the creation of MIDP applications (MIDlets, the applet equivalence in a MIDP environment) includes the ability to create powerful user interfaces and a simple database to store persistent data.

When it comes to networking, MIDP can handle HTTP 1.1 client connections, but has no support for datagrams. This means that some underlying transport must be in place. Native TCP/IP can be used, but WAP transport can be used as well.

The Personal profile on top of the CDC is the next generation of Sun's PersonalJava application environment. PersonalJava represents the first effort to address the software needs of the kind of devices that is relevant to CDC. PersonalJava has gained wide industry acceptance and is, like MIDP, a part of the MExE specification.

MExE

Mobile Station Application Execution Environment (MExE) is a framework for an environment where applications can be executed in a mobile terminal to create new services quickly and easily. MExE itself does not specify any new technologies but defines how capability and content negotiations are performed, how user profiles are handled, and service provisioning and management. The MExE framework also provides a solid security infrastructure to counter the potential security risks that come with the flexibility of MExE.

Currently MExE supports three types, or classmarks, of execution environments on MExE enabled terminals.

- MExE classmark 1: WAP environment
- MExE classmark 2: PersonalJava environment
- MExE classmark 3: J2ME CLDC MIDP environment

A terminal may support more than one MExE classmark. For example, a MExE classmark 1 and 3 terminal could use its WAP browser to reach the contents of the WWW and enable new services in the form of MIDlets.

The counterpart to the MExE enabled terminal is the MExE service environment where service nodes provide MExE services that are transferable to the terminal via, for example, UMTS access, Bluetooth, or infrared. Capability and content negation is required so that the MExE service environment can adapt the services according to such factors as terminal screen size, audio/video input, and MExE classmark.

MExE services can be categorized in four categories, based on how a service is executed. *Figure 24* shows the four possible service execution styles.

At the upper left in *Figure 24*, the service is executed remotely on the service node. Such cases are typically classmark 1 MExE terminals accessing content on a web server. In a certain sense, no service is actually executed on the terminal; content is merely requested and then transported out to the MExE terminal.

To the upper right, a service is executed in stand-alone mode on a MExE terminal. Here the service node is a file repository that MIDlets, for example, can be downloaded from. An example of a possible service falling into this category would be some sort of non-interactive game.

Simultaneous service execution both on the service node and on the MExE terminal is shown to the lower left of *Figure 24*. There is interaction between the MExE terminal and the service node. The transport mechanism used to handle this could be WAP or regular TCP/IP, for example. An example service that falls under this category would be e-mail with a client on the MExE terminal and a server on the service node.

Finally, to the lower right, a case is illustrated where two (or more) MExE terminals interact directly and the service node is a storage site for the files needed. An instant messaging service would fit this scenario.

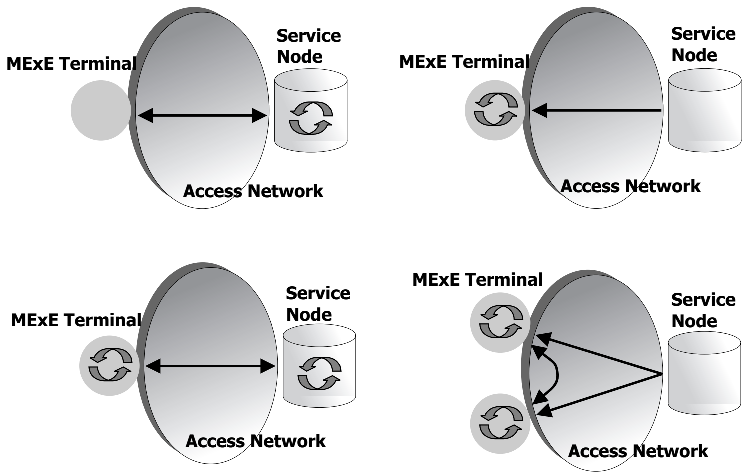

Figure 24. MExE service execution

Acronyms

2G	Second Generation of Mobile Services
3G	Third Generation of Mobile Services
3GPP	3G Partnership Project
API	Application Programming Interface
ARPU	Average Revenue per User
B-ISDN	Broadband Integrated Services Digital Network
CCITT	Comité Consultatif International Téléphonique et Télégraphique
CDC	Connected Device Configuration
cdma2000	Code Division Multiple Access 2000
CLDC	Connected Limited Device Configuration
EDGE	Enhanced Data for GSM Evolution
ETSI TC SMG	European Telecommunications Standards Institute/Technical Committee/Special Mobile Group
FDD	Frequency Divisions Duplex
GMSC	Gateway Mobile Switching Center
GPRS	General Packet Radio System
GSM	Global System for Mobile Communications
GUI	Graphical User Interface
HiperLAN2	Wireless LAN with very high bandwidth
HTTP	Hypertext Transfer Protocol
ICQ	I Seek You, an Internet service
IETF	Internet Engineering Task Force
IMPP	Instant Messaging and Presence Protocol
IMT-2000	International Mobile Telecommunications 2000 MHz
IN	Intelligent Network
ITU-R	International Telecommunication Union - Radio Sector
ITU-T	International Telecommunication Union - Telecom sector
J2EE	Java 2 Platform Enterprise Edition
J2ME	Java 2 Platform Micro Edition
J2SE	Java 2 Platform Standard Edition
JPG/JPEG	Joint Photographic Experts Group, a compressed image format
MexE	Mobile Station Application Execution Environment
MIDP	Mobile Information Device Profile
MMS	Multimedia Messaging Service
Mobile IP	Mobile Internet Protocol
MP3	Mpeg 1 Audio Layer 3, a compressed audio format
MSC	Mobile Switching Center
PCS	Personal Communications System
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
RFC 2002	The Internet Request For Comments document regarding IP Mobility Support
ROI	Return On Investment
SIP	Session Initiation Protocol
SMTP	Simple Mail Transfer Protocol
TD-CDMA	Time Division-Code Division Multiple Access
TDD	Time Divisions Duplex
TDMA	Time Division Multiple Access
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
VPN	Virtual Private Network
WARC-92	World Administrative Radio Conference 1992
W-CDMA	Wideband-Code Division Multiple Access
WLAN	Wireless Local Area Network
WRC-95	World Radio Conference 1995
XML	Extensible Markup Language

www.ausystem.com
3g@ausystem.com