Real-time Charging in UMTS Environment

Bálint Dávid Ary
Budapest University of Technology and Economics, Department of Telecommunications
BME Informatika épület
H-1117. Budapest, Magyar tudósok körútja 2.
Budapest, Hungary
E-mail: balint.ary@isolation.hu

Abstract. *The forthcoming UMTS system and the subscriber's new requirements will bring powerful changes in mobile telecommunication. With the available functions and possibilities the role of the infrastructure provider and content / application provider should diverge. In our study, we will show these new functions and possibilities, we summarize the legal and technical difficulties of the advents of 3rd party providers, and we give a functional model which is compliant with the related standards and solves the problems of real-time accounting and QoS measurement.*

Keywords: UMTS, Charging, Accounting, Real-time, Network management.

1. MOTIVATIONS

On the turn of the 20th and 21st century the mobile phones and mobile telecommunication networks developed rapidly. The end-equipment is evolving continuously, bringing more multimedia and more amusement to the users. However, this technical evolution is not always motivated by the subscribers, the newly developed facilities and functions are getting used widely. The mainspring of this evolution is the information society, which spends more and more money on information.

The early GSM system gives place to the UMTS system, with new, packet based services. With the fulfillment of the All-IP concept, all functionality of a home computer would be available on a mobile phone. IP based browsing, and downloading would be common, VoIP and Video Conferencing would appear in communication and several On-Demand services should be available, so as multicast messaging with IPv6. Accounting in such environment is more complex than it is in circuit based systems. Furthermore, pre-paid subscribers need a real-time approach in order to guarantee the right amount of the paid service.

Temporarily, the services accessible with mobile devices are offered by the network provider (who is the content and application provider as well), but with the continuous growth of the number of accessible functions and media it is predictable, that the network operator won't have enough time and / or energy to invent and offer new services, although it could lead to significant superiority in the market competition. With these conditions the supply of infrastructure and content should decouple, so content and applications should be served by 3rd party providers. Because these providers are financially isolated from the infrastructure provider, an accurate and fair accounting mechanism is needed; however such a method in a distributed packet based and real-time system is not so simple, and hasn't been developed yet.

Henceforth in our study, we assume, that services are always offered by a third party.

2. BUSINESS MODELS

The amount of money paid after a service should be shared among the network operator and the content provider. But because the subscriber dislikes paying to more than one supplier for one service (one-stop-shop concept), the two providers should maintain some financial relationship, and settle the liabilities periodically after validation and identification. [1]

These settlements can be solved easily, but for offering proper and fair service the 3rd party provider should be aware of the financial status of the subscriber's account in order to be able to deny the content if the subscriber is unable to pay. Nevertheless, the rendition of the full account due to the user's personal rights is not viable.

As long as a 3rd party provider plays role in a service (to provide the content) the charging and accounting can be performed by the infrastructure or by the content / application provider. In light of these possibilities, we can differentiate three different business models. [6] [1]

2.1 Network operator centric business model

In the network operator centric business model a subscriber maintains financial relationship only with the network provider. While the content and data are going through the mobile network, the operators are capable of determining the costs of services and to perform the accounting procedures. The 3rd party providers are connecting to the mobile network through the public internet, so the network provider can not guarantee full QoS (packet-loss and delay) for the services, because of the Internet's best-effort mechanism.

2.2 Content aggregation centric business model

In case of the content aggregation centric business model, the contents are accessible through a portal. Besides the physical storage, the portals should offer value added services as well. In this model the subscriber maintains a relationship with the network provider and with the content aggregator. The determination of the costs of the services is done by the content aggregator, but the subscriber pays the necessary fees for the network provider in order to gain access to these services. Because of the small number of these portals, it's conceivable, that the infrastructure provider and content aggregator are in business relationship, so the user can pay both fees to only one of them.

2.3 Content provider centric business model.

The content provider centric business model is quite similar to the content aggregation centric business model, but the content provider plays the role of the content aggregator as well. Because of the huge number of 3rd party providers, the realization of a business relationship is much harder, than it was in the content aggregation centric business model. The main disadvantages of this solution are that the content / application providers must solve the accounting of services on their own, which can be more expensive than the service itself [3] and that the subscribers have to maintain an account with every content provider separately. This solution could lead to problems in case there are many providers. This model brings huge freedom to the services offered, but it means enormous administrative overhead as well.

3. TECHNICAL PROBLEMS

In wired communication and in the circuit based GSM systems accounting was much easier. Because of the permanent and reserved bandwidth, the price of the service depends only on the length of the connection. The GPRS and UMTS systems are packet based, so the measurement of value and quality of the service are more complex.

In order to compute the quantity of the service, we should count the bits that have gone through the system. This method would require huge computing capabilities from the network elements (because of the high speed transport) and would impose a big overhead on the system (for N bits the exact size can only be written in $\log_2 N$ bits). Counting of packets isn't the perfect solution either, inasmuch as the packet lengths in IP network vary in considerable range. Because of the defective quality of the

Internet, a correct method should be aware of the lost and doubled packets. The additive cost because of these failures should not be charged to the subscriber.

In non-circuit switched systems the measurement of quality means problems as well, because in best-effort services a fixed reserved bandwidth absent. Without a permanent connection a guarantee for capacity and delay can only be given with heavy signaling. In multimedia services the measurement of QoS is especially hard, because (for example) in case of video-streaming, the actual content affects the minimal requirements for the quality.

We showed that measuring quantity and quality is a quite complex task, which imposes notable overhead on the system. Furthermore, in a pre-paid environment, it must be done in real-time. In the present solutions, most providers solve it by combining data measuring with some easily measurable unit (time based accounting or constant bit rate accounting), or the measurement is done with greater scale (more kilobytes for instance). [5]

Another problem can be derived from the mobility. If we use a fix IP address, and update the router tables in the network, the movement would be transparent for the charging mechanisms, but the router updates would cause overhead and signaling problems in the network. If the IP address changes continuously the network elements (which supply the charging information) must be informed respectively.

In the UMTS system several media are accessible. Such as

- speech,
- voice (real-time / streaming),
- video (real-time / streaming),
- date (download / upload / interactive content),
- messages (SMS / E-mail),
- data-flow (unspecified content),
- accessed web-pages, portals,
- etc.

Standards give possibilities to subscribers to possess separate accounts for all media available in the system. [4]

In case of 3rd party providers, if the accounting is done by the network provider, the operator should be aware of the exact method of charging and the measurable parameters of the service.

4. THE MODEL

Both the GSM and UMTS systems use charging data records (CDRs) in order to charge the services [4]. The CDRs are generated by the network elements, which play a role in the service. If this information originated from a 3rd party, the information and the providers should be validated and identified. The CDR generation is triggered by

- determinate data amount,
- determinate time-interval,
- the change of charging conditions,
- the change of QoS,
- the change of tariff,
- the change of position or cell,
- closure of voice, data or multimedia sessions. [2]

The usability of the CDRs extends beyond charging. With these records it is possible to

- charge the subscribers for the usage of the networks,
- registering the inter-network traffic among network providers,
- analyze service-utilization,
- archiving the service demands for the case of complaints. [2]

For a correct model it is obligatory to suit the related standards. The optimal model can be developed with the proper determination of the free parameters.

Such variable parameters are the amount of data and time that triggers the CDR generation. The fewer amounts we use to trigger, the more accurate the accounting and the larger the network-overhead will be. Other variable parameters are the physical realization of the charging functions, inasmuch as these functions are not attached to hardware entities.

The function responsible to charge

data transfer can be built into the gateway on the border of the mobile network (GGSN – Gateway GPRS Support Node), into the mobile network nodes (SGSN - Serving GPRS Support Node), into the base station or even into the mobile device. [7]
The third variable parameter or method is the measurement of the services. The standards are not dealing with the measuring methods, therefore for data transfer the estimation of bandwidth or the exact bit count are possible solutions. It can be clearly seen, that some trade-offs are necessary.

The reduction of the subscriber's account in pre-paid systems can be done in two different ways. With "immediate event charging", the account is simply reduced by a charging data record from the appropriate service. The "event charging with unit reservation" solution gives the possibility to the network elements to reserve a proper amount of money from the user's account, and to transfer back the remains (if any) after the end of the service. [2]

In our model both the amount of money (or time) that triggers the CDR generation and the place of the charging functions are varied dynamically. In this solution, if the subscriber has more money on his/her account, than a service-specific limit, the accounting is done with the simple immediate event charging method. In this case the problem of real-time charging can be set aside, while with a proper limit and data / time amount the user could not claim more service than he/she had paid for. Whenever the user's account drops below this limit, the billing system delegates the charging privilege to the network element that offered the service (in case of 3rd party provider this privilege is delegated to the gateway element), namely an event charging with unit reservation is done. While the network element that is responsible for charging and termination of the service is aware of the subscribers account, it is capable to terminate the service when the amount of money on the account reaches zero (Fig. 1).

In multi-task systems, it is possible to access more than one service. In such

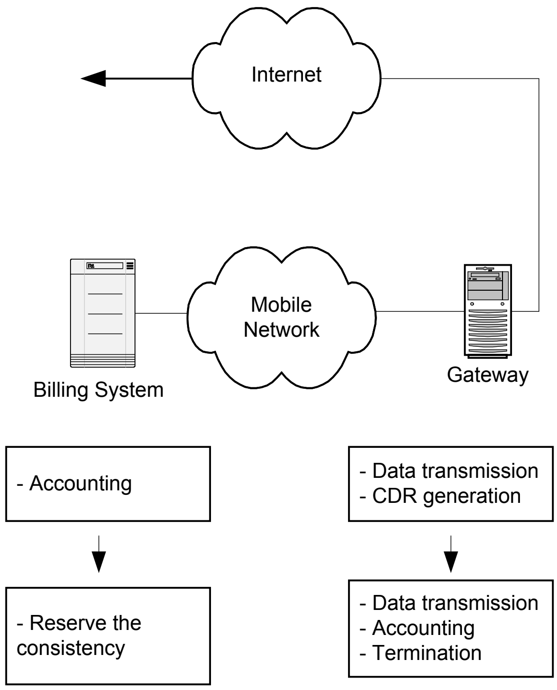

Figure 1. Change of functions of the network elements

cases, when the account drops below the limit, we shall delegate the privilege to multiple network elements. A good solution is to distribute the account among the services with statistical methods, considering the money-consumption and properties of the services, and the behavior of the user.

The UMTS services are based on packet switched network, so we had to count with the packet-loss. The majority of these failures are occurred on the wireless part of the network, but, of course (like on the regular internet), some packet-loss or fault is happening on the backbone as well. Statistical methods can be used, to deal with these failures. Considering the quality of the operator's network, we can send more packets to the user, than it would be necessary with a perfect network, so the user presumably gets the proper amount of packets.

It's practical to include a buffering mechanism between the wired and wireless part of the network, to make the packets resent only from the base station in case of any failure, so the backbone isn't loaded with this traffic. The loss or fault occurring on the backbone is solved with the error correction mechanism of the TCP.

In order to measure the packet-loss and packet based QoS, a presence of a trusted equipment is needed at the end of the connection. This could be the base station, or we can implement the protocol in the low level layer of the mobile phone. The main idea of this solution is that the element has to send some kind of information to the billing system, in order to inform it about the quality. The quality measurement of the data sequence is done by a sliding-window algorithm. After the arrival of a proper amount of packets, the delay (average, maximal delay, jitter), packet-loss, bandwidth and other QoS parameters can be calculated. The retransmission of the lost packets and signaling is done by higher protocols. The measurement of quality can be eliminated with the usage of statistical information about the network, but in this case the results won't mach with the exact situation.

5. ANALYTICAL SUPPLEMENTATION

The user accounts contain "units" instead of real money. The values of services are measured in these units. We can get the correct amount of money from the units with the "rating" method. Our model needs further refinement on order to determine the delegation-limit, the handling of lost packets, and the measurement of QoS.

5.1 Delegation-limit

Let us define a function called unit consumption speed:

$$C(T), \qquad (5.1)$$

with the measure of [*unit/sec*], which represents the consumed units in one second. The consumption speed depends on time to give the possibility to the operators to assign different prices to different time of the day and week. The consumed unit and money can be calculated from the consumption speed with the following equitations:

$$unit = C(T) \cdot t \qquad (5.2)$$

$$money = unit \cdot R(T), \qquad (5.3)$$

where R(T) represents the rating between unit and money. The time-dependence of this function can be used to change the price of the units (in case of inflation or discounts for instance).

Let us use T_c for the time needed to query the user's account. With these notations and definitions the limit for delegation can be calculated. In ideal case it is:

$$L = C(T) \cdot T_c. \qquad (5.4)$$

If we own more units on our account than L, the charging is done by the billing system; otherwise the accounting privilege is delegated to the corresponding network element. If we require more services at a time, the limit can be calculated by the sum of the limits of the services:

$$L = \sum L_i. \qquad (5.5)$$

In case of multiple service demands, the units can be distributed to the serving network elements with the rate of the service's consumption speed. A re-sharing should be done every time a service ends, or an event based service occurs (SMS for example).

When a fix consumption speed could not be assigned to the service (browsing, or interactive content), the average consumption speed should be determined using some statistical models.

5.2 Propagation delay

The events occurring in a distributed, wide network (signaling, queries) have propagation delay, which is not constant in general. If we want to determine the delegation limit exactly, we have to consider the time needed the query (T_c) and to delegate (T_d) the account, together with the variation of these values (T_{cj} and T_{dj}):

$$L = C(T) \cdot (T_c + T_{cj} + T_d + T_{dj}). \qquad (5.6)$$

To ensure accurate charging, we should count with the maximum values of the jitters (T_{ci} and T_{di}). If we want to reduce the values of the delegation limits (in order to reduce the network overhead), we shall count with smaller values (with the expected value for example). In this case the possibility of users gaining more service than they paid for can be calculated from the distributions of the jitters.

In case of re-sharing the control messages should be labeled with proper time-stamps to be able to charge the services gained while the accounting privilege was under delegation.

5.3 Measurement of QoS

The quality can be defined using a sliding-window algorithm; always using the last N packets arrived to the user. With this method, the measured and experienced quality should be close to each other. Let t_j be the transmission starting time and a_j the arrival time of packet j. If the size of the sliding-window is N, the delay (average, minimum, maximum) can be calculated:

$$D_{average} = \sum(a_i - t_i) / N, \quad (5.7)$$
$$D_{min} = \min(a_i - t_i), \quad (5.8)$$
$$D_{max} = \max(a_i - t_i). \quad (5.9)$$

The jitter of the delay is the difference of the maximal and minimal delay:

$$D_{jitter} = D_{max} - D_{min.} \quad (5.10)$$

The packet-loss in case of N arrived, and M sent packets is:

$$Loss = N/M. \quad (5.11)$$

6. CONCLUSION

In our study, we enumerated the motivations for the appearance of 3rd party providers. We have showed the legal and technical issues of this new concept. Most of the technical problems come from the real-time nature and mobility of the packet based network. Finally, we gave a model to solve these problems. The model operates in such a way, that billing is made in the network as usual to a large volume of users (who have more money on their account than the critical amount), with a low CDR transfer. Billing to critical users is more complicated, but they cannot request services in excess of the sum they paid, or the probability that they can is very small.

7. FUTURE PLANS

In the future, in order to develop the complete charging method, it is required to work out the exact method of measuring the data flow and the method to derive the quality of service from the IP based quality. For this, it is crucial to determine the statistic parameters of the services and users. The model is not complete unless the protocols and algorithms are developed.

8. REFERENCES

[1] BUTYKA, ZS. - JURSONOVICS, T. – IMRE, S.: Accounting in Next Generation Networks, *ETIK conference*, Budapest 2004.

[2] Charging and Accounting Mechanisms 3GPP standard: ETSI TR 122 924 v3.1.1

[3] CUSHNIE, J.: Charging and Billing for Future Mobile Internet Services, First Year PhD Research Report, 2000. September.

[4] Charging principles 3GPP standard: ETSI TS 132 200 V5.5.0

[5] SCHWARTZ, S.: "Next-Gen Rating: It Will Be Only as Good as the Network", *Billing World and OSS Today Magazine,* 2003 February, page 16.-22.

[6] Report 21 from the UMTS Forum: Charging, Billing and Payment Views on 3G Business Models, *UMTS Forum*, 2002.

[7] Service Aspects Charging and billing 3GPP standard: ETSI TS 122 115 v5.3.0