AGENT-BASED RESOURCE MANAGEMENT FOR 3G NETWORKS

J. Bigham¹, L. Tokarchuk¹, D.J. Ryan¹, L.G. Cuthbert¹, J. Lisalina², Manuel Dinis³& S. Robles⁴

¹Queen Mary, University of London, United Kingdom, ²Swisscom AG, Switzerland, ³Portugal Telecom Inovação SA, Portugal, ⁴Universitat Autònoma de Barcelona, Spain

Abstract: This paper discusses an architecture for efficient, scalable and robust real time control of 3G mobile systems in the context of realistic business models of Network Providers, Service Providers and Customers and the relationships between these actors. The aim is to bring financial benefits to mobile users through improved efficiency; extensions to the same approach would allow management of a variety of contingencies with social benefits.

INTRODUCTION

Resource management is becoming increasingly important in mobile networks as the resources at the air interface are necessarily limited. Third Generation (3G) mobile systems are seen as the new technology to bring the new broadband services being developed for the Internet to the mobile user. However, providing flexible higher bandwidth services in a mobile environment leads to increased complexity in resource control and management¹ due to the variable bandwidth requirements of the applications, the new radio architecture (e.g. CDMA “cell breathing”) and the varying demands on the fixed part of the infrastructure. Such complexity requires the use of sophisticated control and management techniques. This situation is further complicated by deregulation of telecommunications networks, as users buy services from Service Providers (SPs), who in turn buy capacity from Network Providers (NPs). Resource management is, therefore inextricably linked to the business strategies of these providers.

Intelligent agents have been used in many domains, from workflow management to buying and selling on the Internet. Agent technology has been used in the management of telecommunications systems [1,2,3,4] to address the problems of distributed and decentralised resource control and management, as well as incorporating the “business dimension” with different SPs. Much of that work has been specifically targeted at the management of connection set-up in fixed ATM networks but the concepts can be extended not only to the fixed part of 3G mobile networks, but also to the set-up of connections over a radio interface.

¹ Resource control is seen as a short-term response to events (as in the Control Plane) whereas resource management is seen as being a longer-term mechanism, part of the Management Plane.

The additional problem in a radio network is the allocation of bandwidth to radio cells in order to avoid local congestion or degradation of the Quality of Service (QoS). Much work has been developed in the USA for AMPS mobile networks. Fixed Channel Assignment (FCA) [5] is the simplest scheme, but strongly constrains resource utilisation and optimisation. If the traffic demand in a cell is greater than the available capacity the excess demand will be blocked, regardless of the traffic conditions in surrounding or overlaying cells. This constraint is too limiting and several strategies have been proposed to optimise resource allocation and minimise the probability of blocking new calls or dropping existing calls.

The radio access techniques using WCDMA for 3G systems bring much more complexity in areas such as congestion control than in earlier generation FDMA / TDMA technologies [6,7]. Intelligence and speed are inevitably key factors in improving performance. With CDMA receivers are more complex, particularly with the need to adapt to different code sequences; more accurate power control is necessary as signal sequences must be carefully controlled at the receiver to avoid suppression of weaker signals by stronger ones. The number of code channels on the same frequency channel is limited by the SIR which cannot exceed the value required for reception by the correlator. The different techniques for providing variable transmission rates (e.g. adapting the spreading factor, or the allocation of more code channels to a Customer) and the means of allocating the spectrum inside the cellular hierarchical structure strongly affect the choice and features of the congestion control strategies. It is vital for an efficient deployment of 3G to assess the constraints and the applicable strategies for congestion control inside these chosen technologies.

These factors are being addressed by the IST project “SHUFFLE” which is using agents to control and manage both business interactions and radio resources.

FUNCTIONAL ARCHITECTURE

Figure 1 depicts an outline functional architecture demonstrating relationships between SPs, NPs and Customers. In the past, mobile networks in particular

have generally restricted the Customer access to only one SP. [Notice that the SP and NP could be the same organisation (NSP), but it is assumed that the business roles are separated].

With the influence of the Internet it is likely that a more widespread choice of SPs is to be accessible to 3G users. In fact, a burgeoning of SPs is the driving force that would make the introduction of agent technology highly desirable in such a scenario, if only as a means of enabling the Customer to find his or her chosen information or service. Use of agents would allow SP selection by price, QoS or value-added service type.

However, the introduction of agents goes beyond the Customer-SP interaction. As shown by the European Union ACTS Projects, IMPACT [8] and FACTS [9], agents have a significant role to play in managing resources between SPs and NPs. Moreover, as Bodanese has shown, agents are valuable in managing radio resources within a mobile network. The quote driven approach of Bourne [10] may be particularly useful in this aspect of resource control as the signalling load is likely to be significantly reduced compared with traditional multi-auctions as a mechanism for negotiation or coordination. This may be more scalable and possibly more secure, by virtue of a reduction in the necessity for a centralised trusted authority.

The Facilities Plane houses the entities that ensure secure interactions in the negotiation plane. Issues such as SP registration and key distribution feature here. Confidence, as determined by the facilities provided, is also an important concept. Facilities are provided to allow the agents in the negotiation plane to participate in indirect trading relationships such as blind auctions, as well as infrastructure services. We will reserve trust to refer to the whole model built up from confidence, reputation, loyalty, fairness, reliability and survivability.

The Negotiation Plane is where all the interaction between Customers, SPs and NPs occurs. Here the SP winning business sets up the connection using the NP of its choosing. This plane offers an interface for the agents to communicate with each other as well as with the services provided by adjacent planes. Business models and Service Level Agreements (SLAs) feature in this plane. Reputation and the maintenance of statistics for its measurement are important concepts here to provide rapid decision making.

The Resource Plane is where the NP manages its resources both across and within calls. It represents the "goods" layer of the marketplace. Reliability and survivability are important operational considerations here. This layer lodges all low level network services including handover and cell selection, cell size management and offloads to other networks, and provides an interface to other layers.

Figure 1 Illustration of functional architecture adopted by SHUFFLE

Splitting the system into these three planes supports the definition of well-defined interfaces. Different services are then more independent, providing flexibility (as behaviour of services may change) and adaptability (new services can be added without modifying the design).

Notice that there is no direct Customer-SP negotiation. The assumption is that the Customer? (and hence the User) is "owned" by the SP rather than a NP (in opposition to the current situation). A Customer may subscribe to many SPs but the choice of which to use for a specific request is made at the terminal, either by manual selection or by intelligence in the terminal (this could be a form of gatekeeping).

This model describes the interaction in terms of setting up a connection1 and activities associated with handling sets of Customers. However, there is a separate business interaction between SPs and NPs in longer-term negotiation on capacity, but it is important to stress that because of the scarcity of the radio resource (unlike the fixed resource in IMPACT [8], for example) it is not possible for SPs to reserve capacity from NPs4. Rather, the SPs lease a right to access capacity, governed by some form of SP-NP SLA using probabilistic performance as the measure of compliance. The SLAs are written to be legally

1 Using the distinction between the Customer, who pays for the service, and the User, who makes use of it.
2 Connection is used here in the loose sense of either a circuit-switched virtual connection, or a connection-oriented service such as TCP.
3 A simple calculation shows that if there are 4 SPs each reserving the same capacity from a NP the most any SP could have in a 2Mbit/s pico-cell is 0.5Mbit/s: this fixed reservation is unsatisfactory and dynamic allocation is needed.
binding and hence capable of interpretation by a lawyer, yet they can also be specified in electronic form (perhaps XML) and interpreted by the agents acting on behalf of these actors. The definition and implementation of such SLAs is part of the work of SHUFFLE but outside the scope of this paper.

Facilities Plane

The facilities plane houses all facilities used by agents to achieve indirect agent-agent trading relationships. We call a facility any entity in the marketplace that provides a set of similar services. A facility also provides a well-defined interface and a set of protocols. Some facilities control the infrastructure of the marketplace, forming the institution of the marketplace. Examples are the registration office and the trusted authority. Other facilities provide the system with basic trading services, such as the auction house. Finally, there are facilities to help agents to perceive the environment better, thus allowing complex architectures to have better decision-making capabilities. Examples are the internal and external information offices. Some examples of facilities on this plane are:

- Auction House
- SLA Repository
- Trusted Authority
- Registration Office / Directory
- Legal Service
- Internal and External Information Service
- Financial Office

All these facilities need to be regulated by a set of norms. The norms of the institution are set before any marketplace activity and depend on the scenario. The marketplace, as described above, is in itself a multi-agent system and a secure e-trading and e-commerce centre oriented to business to business activities.

Negotiation Plane

The basic operation to initiate a new service at the negotiation plane is described below. The negotiation plane model can be tailored to suit different business models. Here a general model is described, but many of the ideas apply to more restricted models. It is assumed that all NPs provide connectivity to all SPs, for their own benefit. However, if they do not, then the gatekeeping software in the terminal could choose a network that did. This means that the initial request dialogue can be carried over any network, but that a specific network could be used for the transport of the service itself. A typical sequence might be:

1. The Customer, having chosen a SP, requests a service.
2. The terminal sends a request to the SP for the connection; the SP adds location information to this set-up request (if not already present in the set-up message).
3. The SP selects from (or, more generally negotiates with) all, or a sub-set of the NPs for a transport-layer connection. As geographical location is known, the response can include a "cost" based on the current network load in that region as well as other factors. This negotiation can include constraints based on the SP-NP SLA.
4. Having successfully negotiated a transport-layer connection, the SP can instruct the terminal which NP to choose and what set-up message to send.
5. At this point the SP drops out of the loop; billing information in terms of network usage, QoS, and NP costs is sent by the NP to the SP.

Note that this interaction is completely separate from the resource plane. Moreover, at this functional level it is not necessary to define a mechanism for the negotiation. Several candidates are being considered but only one approach based on reputation is outlined here.

Reputation. The ability to negotiate in several multiple timescales is essential in many environments, and central to management of resources in the multi-SP, multi-NP environment proposed here. Some of the important interactions required to support the management of resources include:

1. bilateral SP-NP negotiations to agree acceptable relaxations of QoS (which can be provided in different ways via the NP's resource agent) when the SP's call request cannot be granted;
2. Customer selection of SP;
3. SP-SP & NP-NP interaction for exchange of resources. Though mechanisms for this are of interest they are not described further here.

Some negotiations (e.g. SP-NP) will have very tight real time constraints, while others (e.g. SP-SP) are more relaxed. To provide mechanisms that support fast real time decisions we concentrate on the use of the concept of reputation allied to negotiation, but this is beyond the scope of this paper.

Agents. It is beyond the scope of this paper to describe all the agents in depth, but a summary will be given of all of them, plus a more detailed description of the reactive layer of the NPRA to illustrate the functionality that can be introduced:

- User Agent (UA) resides in the user's terminal and is responsible for requesting the service. It could also choose the SP but, as that has functionality
implemented in an earlier EU project (albeit for fixed networks) it is not being repeated in SHUFFLE.

- **Service Provider Agent (SPA) / Network Provider Agent (NPA)** are responsible for managing the business of the two actors. They will use information from their resource agents and instruct them on business strategy. It is here that different competition strategies can be implemented.

- **Service Provider Resource Agent (SPRA)** acts on behalf of the SP to select an NP to handle a particular connection and to negotiate:
  - service portfolios with UAs
  - SLAs with NPNAs
  - (bulk) bandwidth with NPNAs

The SPRA’s reactive layer is thin, as it needs the ability to react quickly: the important issue here is the planning layer. A fast approach to the network selection function of the reactive layer is to use the concept of reputation as well as price.

- **Network Provider Negotiation Agent (NPNA)** has its main role in the planning layers, although there is a possible reactive layer function to change the requested QoS of a connection if the NP decides this is necessary in order to meet its SLA commitments.

SP-NP negotiation only comes into play when the NP recognises that it cannot offer the required QoS. This may be immediate, or may be a result of feedback from the reactive layer, and is in the interests of both SP and NP as previously discussed.

QoS and status of the SLA is multidimensional for both parties, and negotiation involves making trade-offs along the dimensions of QoS in the context of the current SLA compliance. Proposals and counter proposals need to be made. Typically, one-to-one negotiations are conducted by making concessions, adopting different kinds of tactics:

1. **Time-dependent**: these model the times during a negotiation when concessions should be made. For timeliness, it is likely that a time limit, or more practically a limited number of negotiation cycles would be allowed. A very busy NP may prefer to make concessions quickly rather than try to compute and iterate through trade-offs.

2. **Resource-dependent**: for example, simply concede more if resources are tight.

3. **Behaviour-dependent**: make concessions dependent on the behaviour of the other party, e.g. tit-for-tat.

**Resource Plane**

The resource plane is the province of the NP and it is here that the network is managed at a detailed level to meet the SP-NP SLA. If there is no congestion, the network operates as provisioned. If load is increasing, however, or congestion has started to appear, the NP can take remedial actions, all with QoS implications, such as:

- At call select, choose a different type of cell, providing a lower QoS to the Customer, e.g. assign a macro-cell instead of a pico-cell thus providing a lower bit rate;
- Forcibly handover a connection to a different type of cell using a different frequency band;
- Offload the call to an overlapping GSM network owned by the same NP (GSM offers lower QoS);
- Degrade the connection’s QoS without handover.

These are **cell-scale** operations and would be the NP’s first choice. Next choice would be to manipulate radio resources:

- Change a cell’s effective radius to sacrifice coverage for capacity gain;
- Force termination of those calls consuming more resources than average, to provide overall service to more Customers;
- Reduce bit rate of non real-time services;
- Reduce bit rate of real-time services
- Deny power-up commands;
- Adjust radio parameters;
- Deny access to adjacent cells to reduce interference to a congested cell
- Turn off soft handover capability
- Offload the call to another NP.

This latter option may seem unusual, but NP-NP cooperation may provide mutual benefit with better loading for all. Conversely a NP may take a position in the market whereby it guarantees to carry all connections over its own network, at least where international or other off-net connections are not required.

Many of the techniques for resource control are independent of the specific mechanisms used in the negotiation plane and do not depend on the described general business model.

**NPRA ARCHITECTURE**

Both the planning (deliberative) and reactive layers of the NPRA observe the state of the network and its resources. The local planning layer is in charge of setting the policies for how:

- calls are assigned to NodeBs;
- calls are admitted to NodeBs;
- QoS relaxations can be made with and without contacting the NPNA;
to manipulate cell and radio parameters on a group basis to satisfy QoS;

- to achieve scenario driven tactical considerations.

According to these planning layer policies the reactive layer is responsible for:

- NodeB assignment;
- Connection Admission Control (CAC);
- suggesting reasons for CAC failure and triggering calculation of a revised acceptable QoS request if appropriate for the type of call;
- manipulation of cell and radio parameters on an individual basis.

The temporal ordering of the above functions is not prescribed by the architecture. Normally these functions should finish by returning a solution for their part of the processing, within prescribed bounds. Each function may fail, and in these cases the reactive layer has fallback actions, particularly relevant in the case of CAC, that may perhaps allow the function to succeed. The critical point is that the reactive layer does not need to consult the planning layer to fall back. These fallback, or exception management, strategies are not guaranteed to succeed, but are valuable because:

- the planner is slower than the reactive layer and so must be based on predictions which will inevitably be wrong some of the time;
- imbalances created by exception management are temporary and will be compensated for in the next planning cycle.

The planning layer can monitor the reactive layer’s performance and trigger re-planning when too many exceptions are executing. Alternatively, re-planning could be performed continually over a suitable time interval.

CONCLUSIONS

Significant progress has been made in the design of an agent system to support a variety of business models for the management of 3G radio networks. Preliminary implementation is concentrating on the reactive layers of the individual agents to ensure that adequate real-time response is maintained. Simulations have shown that the technique of handing down policy to a simple reactive engine at runtime can be used by a rudimentary planning layer. Future work will elaborate the higher layers.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge financial support from the Commission of the European Union under the IST Programme, as well as the contributions from other partners in the SHUFFLE Project. The full list of partners in SHUFFLE is: Martel GmbH (CH), Queen Mary, University of London (UK), Nortel Networks PLC (UK), National Technical University of Athens (GR), Portugal Telecom Inovação SA (P) and Swisscom AG (CH).

REFERENCES


