RTT TECHNOLOGY TOPIC
November 2002

SIP and Session management

Session Initiation Protocol is a protocol originally developed to manage IP voice calls. Like many IETF protocols, it now has wider ambitions and is presently promoted as 'a signalling protocol for establishing, routing, modifying and terminating calls or communication over Internet Protocol networks.'

It has been adopted by 3GPP, the body responsible for 3G cellular standards and is included in Microsoft's Windows XP.

It promises to replicate existing PSTN circuit switch functionality, to add additional session management capability **and** reduce signalling overhead.

Ambitious claims.

In this month's HOT TOPIC, we explore what's needed from a session management protocol in a (3G) wireless network and provide a context in which SIP's future capabilities can be assessed on an objective rather than subjective basis.

How session complexity increases over time

Let's start by revisiting the topic of session complexity that we touched on in our June HOT TOPIC (The Circuit Switch to Packet Switch Transition).

Session complexity is determined by the number of users participating in a session and the complexity of the content exchanged.

As session complexity increases, session value increases. Session complexity is therefore beneficial provided we can bill for it and have transparency to cost information (see below). If session complexity can be realised as session margin then it is advantageous to have session management software and session management protocols that help to increase session complexity (encourage participants to participate) as a session progresses.

For example, an initial one to one peer to peer exchange will ideally (from a session value perspective) be developed into a one to many, many to one or many to many exchange.

Similarly a simple one to one text message exchange will ideally (from a session value perspective) be developed into a simultaneous text, voice, image and video exchange.

Note however that as session complexity increases, bandwidth requirements and

signalling overheads increase.

In a 3G cellular network, radio bandwidth (physical layer resource) is allocated using a combination of static rate matching (additional code streams to support voice, image or video) and dynamic rate matching (increasing or decreasing allocated bandwidth at 10 millisecond intervals).

This implies an intimate relationship between the radio layer and higher layer bandwidth allocation protocols.

As session complexity increases, session persistency (generally) also increases - both contribute to session value - they increase (or should increase) billability.

Session complexity value can however only be realised if session quality is maintained. Session quality is determined by factors such as text integrity, speech and audio quality, image and video fidelity (colour depth, resolution, frame rate) and application integrity. Session quality is also dependent on session consistency.

Session consistency becomes harder to maintain as session complexity increases. We may find for example that we need to manage and control time inter-dependency between multiple per user channel streams which means we need to measure and manage delay and delay variability - i.e. the need for session immediacy increases with session complexity.

In addition, as session complexity increases, we will tend to see a parallel increase in policy complexity.

Policy complexity is determined by the presumption that we bill users, devices or applications differentially depending on their right of access to delivery or storage bandwidth.

In an IP network this requires us to prioritise packets and packet streams (delivery bandwidth) and provide differentiated access to storage and server bandwidth The policing of access rights in turn depends on our ability to authenticate devices and/or users as they join or re-join a session.

In a wireless network, users may leave a session involuntarily (they walk into a tunnel and the session disconnects). When they rejoin the group (walk out of the tunnel) they will need to be re-authenticated and go through a re-admittance procedure.

In addition to introducing discontinuity, the radio layer will or may introduce variable quality into the media exchange.

Session properties therefore change as a session progresses. Some of these changes are intentional and can be controlled; some are unintentional and cannot be controlled.

Session properties can be described in terms of their dynamic range - how many users in the session, the types of media being exchanged, the bandwidth of the media being exchanged (influenced by the quality requirements of the media) and the

rate of change.

Note that the session management protocol needs to have sufficient dynamic range and a fast enough rate of response to preserve session value.

SIP and session value

This brings us back to SIP. We have implied that the job of the session management protocol is to maximise and preserve session value, i.e. it should be capable of responding to changes in the properties of the session as the session progresses.

Figure 1 shows a horizontal view of SIP. Session establishment requires a negotiation to be undertaken between users participating in the session, and the devices and applications being used.

Session modification requires the session to be continuously modified as session requirements change. Session modification may be problematic if the rate of change is high.

Session clear down requires the capture of session specific and user specific billing data (session data records/service data records) for delivery to the billing system.

We can begin to see that session management is really not so simple.

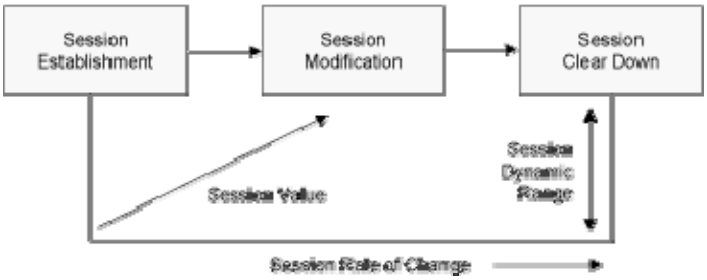

Figure 1: A horizontal view of SIP

Session properties and session management

Figure 2 shows a vertical view of SIP referenced against the traditional OSI model.

The user has application software that is capable of defining, requesting and negotiating its particular session specific requirements. (A declarative application is an application capable of declaring what it needs and wants). The application will be running on a target device that itself will have certain capabilities and constraints that will need to be accommodated by other session participants (a declarative device). The user is probably declarative as well if given the opportunity.

In the figure we have used Windows XP as an example of declarative application software. To be effective in a mobile environment, XP (or Symbian or competitor application software) will need to take into account device profiles in addition to using SIP or competitive or complementary protocols such as RSVP to communicate with

all other layers of the OSI protocol stack,

The session protocol has to arbitrate these session requirements and match them to available transport, network and radio resource. Note that data link layer and physical layer resources in a 3GPP1 radio network are allocated using an ATM ABR protocol (though it is not described as such in the standards). Network bandwidth allocation decisions are made on the basis of network congestion measurements; radio bandwidth allocation decisions are made on the basis of radio interference measurements. These measurements are presently captured by the SS7 signalling layer.

Given the complexity of the network and radio admission procedures, it is inconceivable that SS7 should be dropped as a signalling standard. For SIP to survive it will therefore need to work with these existing signalling protocols.

The Traditional OSI Model	The User	Software and protocols
	↓	
Application Layer	Declarative Applications	Windows XP
↓	↓	↓
Preservation Layer	Declarative Devices	HTML/XML
↓	↓	↓
Session Layer	Session Management	SIP/RSVP, etc
↓	↓	↓
Transport Layer	Adaptive Transport Resource Allocation	TCP
↓	↓	↓
Network Layer	Adaptive Network Resource Allocation	IP
↓	↓	↓
Data Link Layer	Adaptive Radio Resource Allocation	ATM
↓	↓	↓
Physical Layer	Radio Bandwidth Allocation	WCDMA or CDMA2000

Figure 2 A vertical view of SIP

It comes down to the question of who is in control of the session - the user, the application, the user's device or the management protocol. Will it work? Well, possibly, but the specific claims are that SIP can or will do all the above while at the same time saving on signalling overhead. This is hard to believe.

The additional need for session cost transparency

As session complexity increases, signalling overhead increases which in turn increases the cost of delivery. This doesn't matter as long as we achieve a session tariff premium, which is greater than the additional costs incurred. But how can we tell? At present there is no consistent way of describing the cost of delivery of a complex two user or multi user packet driven session. Without knowing session cost,

we cannot calculate session margin. We cannot know whether our bandwidth is making or losing us money.

So perhaps it's not a session initiation protocol that we need, perhaps it's not a session management protocol that we need, perhaps what we need is a cost management and margin management protocol (CMMMP).

Session value is discussed in detail in Geoff Varrall and Roger Belcher's new book '**3G Design**' published by John Wiley New York December 2002.

About RTT Technology Topics

RTT Technology Topics reflect areas of research that we are presently working on.

We aim to introduce new terminology and new ideas to clarify present and future technology and business issues.

Do pass these Technology Topics on to your colleagues, encourage them to join our Push List and respond with comments.

Contact RTT

RTT, the Shosteck Group and The Mobile World are presently working on a number of research and forecasting projects in the cellular, two way radio, satellite and broadcasting industry.

If you would like more information on this work then please contact

geoff@rttonline.com

00 44 208 744 3163