3GPP TR 22.801 V12.0.0 (2011-12)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
Study on non-MTC
Mobile Data Applications impacts (Release 12)

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.
The present document has not been subject to any approval process by the 3GPP Organizational Partners and shall not be implemented.
This Specification is provided for future development work within 3GPP only. The Organizational Partners accept no liability for any use of this Specification.
Specifications and reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Keywords
mobile, data, applications

3GPP

Postal address

3GPP support office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© 2011, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TTA, TTC).
All rights reserved.

UMTS™ is a Trade Mark of ETSI registered for the benefit of its members
3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners
LTE™ is a Trade Mark of ETSI currently being registered for the benefit of its Members and of the 3GPP Organizational Partners
GSM® and the GSM logo are registered and owned by the GSM Association

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Some mobile data applications might result in adverse impact to the mobile network, e.g. due to frequent idle-active mode changing, frequent start or stop of services, small data transmission, frequent live update, transmission of data burst. Hence, the network (both RAN and CN) may experience a flood of signalling and data traffic.

This study aims to investigate the service scenario/use cases of different mobile data applications. Their impact to the current system is generalized. Potential service and operational requirements are indentified for possible enhancements to the system.

1 Scope

To make the network better suited for mobile data applications, the aim/scope of this study is to:

- Capture real world data / analysis.
- Identify services scenarios / use cases for mobile data applications.
- Identify potential problems / issues caused by different mobile data applications.
- Identify potential service and operational requirements for possible enhancements to the system.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] "Apple App Store Tops 300,000 Apps," internet source: http://www.pcmag.com/article2/0,2817,2373169,00.asp

[3] "Crossing the 100,000 Applications Mark," internet source: http://blog.androlib.com/

[4] "Smartphones and a 3G Network," Signals Research Group.

[5] "Diversity in Smartphone Usage," H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, MobiSys'10, June 2010.

[6] "A First Look at Traffic on Smartphones," H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, IMC November 2010.

[7] "An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol," S. Baset and H. Schulzrinne, IEEE Infocom 2006.

[8] "Find*Me," internet source: https://market.android.com/details?id=com.gpsshare.ads

[9] http://www.caida.org/data/passive/trace_stats/sanjose-B/2010/equinix-sanjose.2010-0415.dirB.df.xml

[10] http://www.caida.org/data/passive/trace_stats/sanjose-A/2010/equinix-sanjose.2010-09.dirA.df.xml

[11] http://www.caida.org/data/passive/trace_stats/sanjose-B/2010/equinix-sanjose.2010-09.dirB.df.xml

[12] http://www.caida.org/data/passive/trace_stats/chicago-B/2009/equinix-chicago.2009-04.dirB.df.xml

[13] C. Na, J. K. Chen, and T. S. Rappaport, "Measured Traffic Statistics and Throughput of IEEE 802.11b Public WLAN Hotspots with Three Different Applications," IEEE Transactions On Wireless Communications, Nov. 2006.

[14] F. Wamser, R. Pries, D. Staehle, K. Heck, and P. Tran-Gia, “Traffic characterization of a residential wireless Internet access,” Special Issue of the Telecommunication Systems (TS) Journal, 48: 1-2, 2010.

[15] http://www.youtube.com/live

[16] http://www.tuaw.com/2011/01/05/comcast-to-offer-3-000-hours-of-live-streaming-content-to-ipad/

[17] http://www.omaha.com/article/20101205/MONEY/712059914

[18] 3GPP TS 26.346: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[19] http://www.kaspersky.com/news?id=207576158

[20] http://www.appedia.com/news/2148.html

[21] http://www.androidpolice.com/2011/03/01/the-mother-of-all-android-malware-has-arrived-stolen-apps-released-to-the-market-that-root-your-phone-steal-your-data-and-open-backdoor

[22] X. Geng and A.B. Whinston, "Defeating Distributed Denial of Service Attacks," IEEE Journal of IT Professional, Volume 2, Issue 4, Jul. 2000.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

None

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

AoS	Always Online Service
Apps	Applications
CN	Core Network
DDoS	Distributed Denial of Service
EPC	Evolved Packet Core
IM	Instant Messaging
MODAI	Non-MTC Mobile Data Application Impacts
SNS	Social Network Services
SP	Service Provider

4 Background

4.1 Mobile data applications and their adverse system impacts

Today, there exist hundreds of thousands of mobile data applications for mobile devices [2] and [3]. Many of these applications utilize the mobile broadband connections to provide various types of communications to the users. While some of these applications focus on more “traditional” use cases such as web browsing or email reading, other emerging applications such as social networking applications help the users to “stay connected” with their friends on the go. Below is a short list of different categories of mobile data applications:

- Web browsing
- Email
- Weather/News updates
- VoIP (Skype, etc.)
- Social Networking (Facebook)
- Geo services (Google places/location-targeted ads)
- Online games
- Messaging (SMS and instant messaging)
- etc.

A recent study [4] noted that the way social networking applications on mobile devices transmit and receive status update messages can cause significant signalling congestions in 3G networks. There are other operational characteristics of the mobile data applications, such as small data transmission and frequent start or stop of services that can cause adverse impacts to the mobile network.

So as more and more mobile data traffic is handled by the network, the signalling traffic associated with various non-mtc mobile data applications can cause high signalling burden on the network and lead to poor user experience. Also, some of these applications have "keep alive" messages that can generate large amount of signalling traffic.

As shown in Figure 1, it can be observed that, there is a many-to-many mapping between the mobile data apps and their operational characteristics/potential problems. Such mapping implies that:

- One application can potentially cause multiple problems to the network.
- Different applications can contribute to the same problem to the network, thereby aggravating the problems observed in the EPC and RAN.

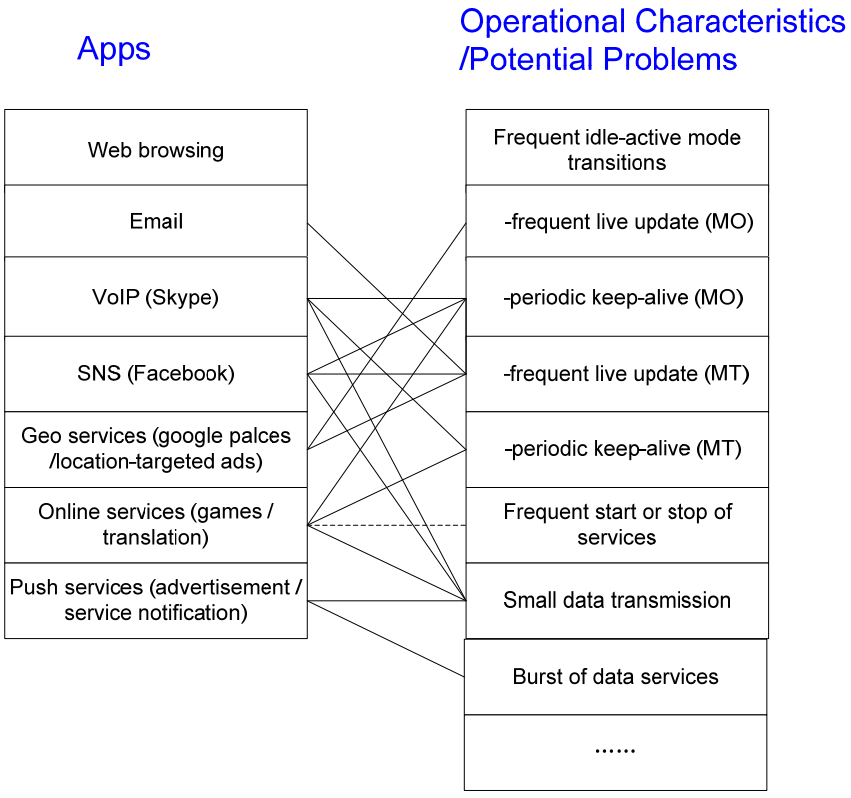

Figure 1: Many-to-many mapping between the mobile data applications and potential adverse system effects

5 Use Cases

5.1 Frequent transmission of small data packet

One characteristic of some mobile data applications is that they generate small data packets, e.g. IM chatting, heartbeat like operation compared to legacy CS/PS services and HTTP/WAP browsing services. Small data packets are exchanged frequently when mobile data application runs on a UE. The time interval between heartbeat messages is often several seconds for some applications, IM presence status update information may change frequently which is compounded by generating large number of small data packets as an update message is pushed to all friends in the buddy list.

Data relating to the presence of small packets over 3GPP access.

When looking at packet traces of major internet POPs [9], [10], [11] and [12], it is clear that roughly 40% of the packets on the Internet today are less than 50 Bytes for IPv4 traffic. Similar observations have been seen in the traffic analysis for wireless access technologies as well [13] and [14]. These are rather large fraction of packets in terms of the total number of packets that flow through the network. These packets could contain a variety of payloads such as TCP ACKs and application related payloads such as VoIP silence suppression and other small payloads.

When these small sized IP packets get transmitted over the 3GPP access, a couple of things are expected to change:

1) ROHC, if enabled will further reduce the size of the packet by compressing the IP/UDP/TCP headers in a very effective manner.

2) The 3GPP access will add PDCP, RLC and MAC related overheads which will bump up the size of the packet.

It is however expected that the net effect of 1) and 2) above will be that the packet size effectively could become even smaller especially in scenarios such as VoIP where the ROHC compression can be quite efficient in bringing down the IP/UDP/RTP headers down to a few bytes in size.

This trend of small packets is expected to be exacerbated as status messages, location messages, instant messages, keep alives etc as generated by the current generation of mobile data apps grow considerably over time.

Hence it is imperative that 3GPP accesses provide enough mechanisms to ensure optimal delivery of such small data packets.

It is to be noted that with IPv6, the small data packet problem will remain the same over 3GPP access even though the IPv6 header is much larger. This is due to the application of ROHC over 3GPP access.

Issues

The frequent exchange of large amount of small data packets may reduce network efficiency and waste network resources, e.g. some IM applications may consume more network resources to transmit same amount of data than HTTP/WAP browsing applications.

5.2 Status updates and keep-alive messages due to always on mobile data applications

Some existing “always on” mobile data applications [5] and [6], such as IM, Social networking apps etc are currently bringing some challenges to operator networks. In general, these mobile data applications involve interactive communications, through operator network, with their application servers in the internet. The server and the application on the UE periodically exchange “heartbeat” messages (also Known as keep-alives) to keep the application session alive and also to avoid the expiry of NAT mapping which causes IP session disconnection.

In addition to periodic keep-alive messages, the applications also generate frequent status update messages to notify the users of status updates relating to the application. Some examples include presence information of buddies in an IM buddy list, update of user location upon user “check in”, update of “Facebook likes” to a user’s friends, etc.

- Frequency of Keep-alive messages:

- VoIP apps such as Skype and Fring generate keep-alive messages from once every 30 seconds to every 8 minutes [4] and [7].

- Frequency of Status update messages:
  - Social networking applications such as FindMe generate status update messages upon geographic position changes. The frequency of such messages ranges from sporadic over a day (e.g. changing from home to work to gym then back to home) to periodic up to every 60 seconds [8].
  - Social networking servers push content and presence update messages of the subscriber's friends to the applications on the UE (e.g. Facebook posts the activities when your friend "likes" a particular article or "becomes a fan" of a particular group). The frequency of such content and presence update messages is estimated in the order of every few minutes [see Annex A].

Two more aspects that can aggravate the impacts of status update and keep-alive messages are:

- These messages can be mobile-originated (MO) or mobile-terminated (MT), e.g. periodic FindMe messages can come from change of location of your friends or can come from the updates of your own location.
- It is not uncommon that a UE will install multiple applications, where each application generates these update/keep-alive messages autonomously.

Issues

When the transmission of keep alive or status update messages are completed, and upon detection of user inactivity, the UE may be moved to a low power state (e.g. from connected to idle) in order to save the UE battery power.

As a result, when the average frequency of status update and/or keep-alive messages is greater than the inactivity timer, the UE will have to cycle among idle, wake up, re-establish the connection, send or receive the update message(s), go back to idle and so on.

Figure 2 further illustrates the timing when the UE experiences such frequent idle-active mode transition problem. From the left-most of the figure, after the phone finishes some data traffic, it stays on active mode for a while and switches to idle state to save power. Soon after the phone enters idle, application #1 generates an update message and the phone wakes up, transmits some signaling messages to establish the connection. We note that, the phone might consume more energy in sending signaling messages than it's in active mode but sending no message. After establishing the connection, it sends the update message and again stays in active state for a while before going to idle state. This cycle repeats as other applications also send/receive update messages (e.g. some MT update messages pushed by the server of app #2 and MO update messages generated by app #3).

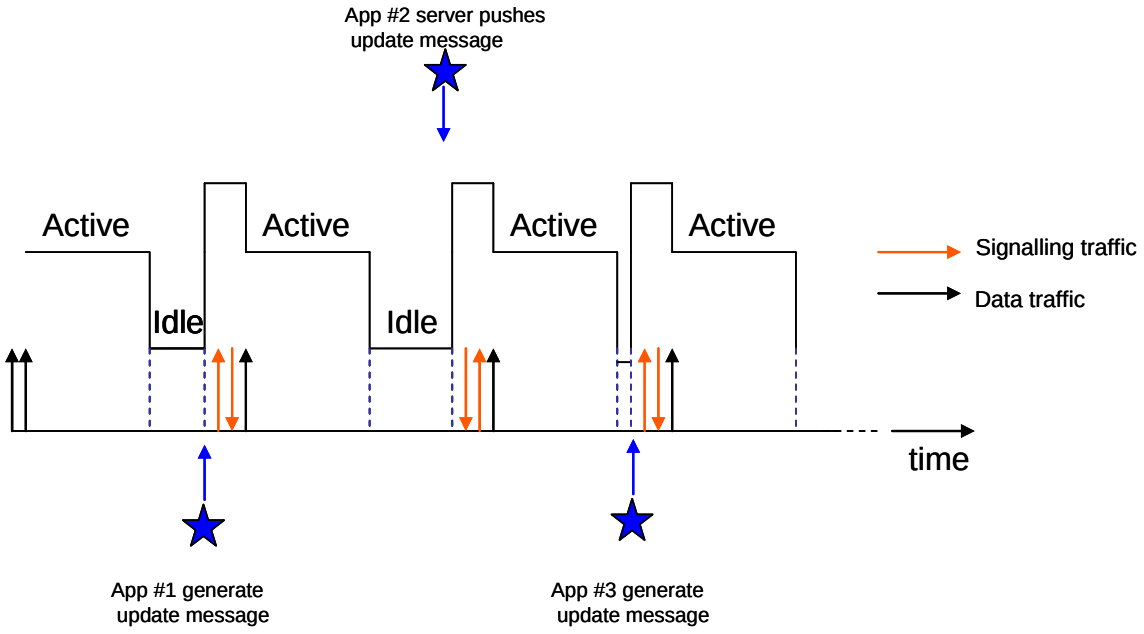

Figure 2: Timing when the UE experiences such frequent idle-active mode changing problem

We can see that, when the UE constantly flips between active and idle state, there are two problems observed.

- Increased control plane signalling:

  There are excessive signaling overhead (both in RAN and in CN) to just send these occasional, very small update messages. To send just one update message, it may take one round of idle-active transition which may incur significant signaling overhead, including multiple RRC messages in RAN (e.g. Service Request, Radio Bear Establishment/Release, and Paging when message is MT) and EPC signaling messages (e.g. Service Request, Connection Setup/Release).

- Reduced battery life of UE:

  In the worst case scenario, when multiple applications generate update messages soon after the phone enters idle state, the energy consumption of the phone increases due to constantly flipping between active and idle state, it may be higher than if the phone just remained in active mode.

Table 1 summarizes the problem scenarios, the sources of problems, and the affected elements of the frequent idle-active state transition scenarios.

Table 1: Signalling inefficiency and reduced battery life caused by mobile data application status updates and keep-alive messages

Problem scenario	Apps that cause the problem	Effect to EPC	Effect to RAN	Effect to UE
MO status update	* Social ntwk: UE owner's status update. * Geo service app: geo-tags, geo-cast etc.	signaling overhead (set-up & tear-down)	RRC signaling overhead	Reduced battery life
MO periodic keep-alive	* VPN * Skype when not in a call			
MT status update	* Social ntwk: friends' content/status update. * Geo service app: location-targeted event/ads.	* signaling overhead (set-up & tear-down)	* RRC signaling overhead * paging signaling overhead to tracking area	Reduced battery life
MT periodic keep-alive	Skype when not in a call			

5.3 Frequent start of service

One of the characteristics of Push-to-talk over Cellular (PoC) service is the frequent start of service. Generally, PoC service is intended for business use as PoC calls are mostly "one to many" calls. Each time users start up the application, the UE goes directly from idle to active, and each user takes an independent wireless resource simultaneously. This means that the resource consumption will be large and will produce a large volume of signalling when many users join the call.

Issues

As the capacity of network is limited, additional signalling produced by users' frequent start of PoC application can lead to shortage of wireless resources and network congestion. In addition, the frequent start of application will reduce battery life of UE.

5.4 Overload caused by live content streaming

Internet live content streaming that offers live TV viewing experience and real-time coverage of popular events [15], [16] has become increasingly popular. In the case of breaking news TV coverage or popular sports events, large numbers of interested viewers flock to the content streaming website and are eager to view the same live program. This phenomenon creates a surge in number of streams trying to flow through the network, consequently creating congestion in the network, especially in the air interface as bandwidth is a scarce resource [17].

Issues

Unlike on-demand based content streaming, Internet live content streaming poses unique challenges to the network operators' bandwidth management strategy because many live events/TV draw a huge crowd of viewers clogging network bandwidth at the same time. It is worth noting that, even though viewers are all interested in the same program, current Internet live content streams are delivered to wireless subscribers in the form of multiple streams of unicast traffic. Since many live content streaming sources come from the Internet rather than the multicast/broadcast services hosted by the network operator, the broadcasting infrastructure within the operators' network, e.g. MBMS [18], can not be utilized to deal with such situations.

5.5 Distributed Denial of Service

With hundreds of thousands of mobile data applications developed by amateur developers on open platforms such as Android and Apple iOS, the emergence of intentionally or accidentally malicious mobile data applications is becoming increasingly high [19], [20].

So far most of the intentionally malicious programs target the UE, such as stealing users' identity, contacts, or making premium SMS/phone calls without users' knowledge [21].

Accidentally malicious programs could be categorized as those applications that are poorly written in turn leads to sub-optimal behaviour of UEs in the operator's network. Instances of such applications are ever increasing due to a huge base of amateur application developers. Furthermore, the problem could be compounded by the presence of a large number of UEs running one or more accidentally malicious programs.

Intentionally malicious or accidentally malicious applications could potentially lead to intentional or accidental Distributed Denial of Service (DDoS) as detailed in this use case.

Distributed Denial of Service (DDoS) attacks usually involve one computer instructing multiple (infected) computers to attack multiple networks or hosts, using the infected computers to mount a powerful, coordinated attack [22]. The effects of DDoS attacks are usually in the form of inundating the victim networks/hosts and blocking legitimate visitors.

Intentional DDoS:

In the context of Mobile Data Application based DDoS attack in wireless broadband networks, the attackers target is to paralyze the wireless networks and take the network and subscribers hostage. The attackers' strategy can be described as follows. The attackers first publishes the malicious programs, often disguised as a legitimate applications, on the open development platform for users to download. The attackers can easily keep track of how many users have downloaded the program, for example by using the download counter provided on the development platforms. When there are enough users that have downloaded and unknowingly infected their devices with the malicious program, the attackers utilizes the back-door left at the malicious program (e.g. Trojan horse) to coordinate simultaneous attack toward the wireless network.

Accidental DDoS:

Accidental DDoS occurs when a lot of UEs end up downloading one or more accidentally malicious programs – which are poorly written applications and cause sub-optimal UE behaviour in the operator network. The sub-optimal behaviour could be in the form of chatty collaboration between UEs, establishing frequent and large number of connections from the UE etc., often leading to excessive data and signalling usage by the UE. This effect can get compounded with higher number of UEs downloading and executing such accidentally malicious applications as well as with more accidentally malicious applications running per UE.

In summary, even though the accidentally malicious application is not intended to attack the network resources, it effectively ends up doing so.

Issues

1) Formats and impacts of the attacks: In one form of the attack, the applications can coordinate the infected devices to initiate continuous and possibly synchronized attach requests to overwhelm the network, causing congestion in both radio links and backhaul links. On the other hand, the infected devices can be co-ordinated to initiate useless but bandwidth-demanding traffic, both uplink and downlink, to inundate the network. The impact of such attack is waste of resources in the network and potentially overage of data charges to subscribers.

In addition, both types of attacks can cause severe damage to the wireless network by blocking legitimate users from using the network and potentially overload and bring down the network making it temporarily out of service.

2) Scale and affected areas of the attacks: the scale and affected areas of the attacks will depend on geographic distribution of infected devices. If large numbers of infected devices in one cell launch attacks at the same time, the radio links become the victim. If the infected device are distributed across an area that is covered by a few cells, the coordinated attack will likely congest the backhaul and elements in access network, e.g. expose processing bottleneck in MME or S-GW, or cause congested links between eNB-MME/S-GW.

It is worth noting that, with the increasing popularity of devices capable of position techniques, in the event of malicious attacks, the attackers might leverage such capability to initiate geo-targeted attacks that aim to take down strategic areas that are of high importance of operators' deployments.

3) Challenges in attack detection and issues of false alarm: it is hard to detect a mobile application oriented DDoS attack, as they are hard to be distinguished from legitimate access requests/congestions from regular UEs.

5.6 Network congestion by push services

Certain push mobile data services do not require real time delivery, e.g. advertisement, service notification, video clips, or some multimedia messaging services. These services may be provided by the network operator, or from third-party providers. Usually these services are distributed to a large number of users in a short period, e.g. news report in rush hours of every morning and every evening. Customers subscribing to these services hope they could receive the content periodically or on time. However the burst of data service will lead to network congestion and impact service experiences.

Issues

Dense simultaneous push services in a local area may cause the network congestion and service break. As such push services do not have strict real-time requirements, sending them to a large number of devices within a short duration is wasteful and unnecessary, especially when parts of the network are already congested. The delivery of such push services could be delayed until there are sufficient network resources available, which not only optimally utilize network resources, but also guarantee acceptable user experience.

In the case where the push services are provided by third-party providers, the lack of knowledge on the current congestion situations of different parts of the network results in the inability of third-party providers to schedule the push services more intelligently.

In the case where the push services are provided by the operator, lack of mechanisms and interfaces to schedule the push services (which are not located in the EPC) based on the current network status at the core and access networks also makes it not possible to schedule the push services efficiently.

5.7 Background traffic of mobile data applications

The background traffic can be large data flow, such as unexpected software version update and large file download. It can also be periodical data transmission, such as regular unwanted collected-data uploading and probing data for connectivity.

Issues

Unlike on-demand or allowed data traffic, background traffic is usually generated automatically and without users' consent. It causes additional traffic which the user may complain about as they will be charged especially in the case of international roaming.

6 Potential Requirements

The network shall be able to provide the capability to reduce the overheads associated with the transport of huge volume of small data packets generated by non-MTC mobile data applications.

The definition of a small amount of data shall be configurable according to network operator policy.

The system shall be able to provide capabilities to minimise signalling surge caused by mobile data application behaviours such as keep-alives, status updates, instant messages etc.

The system should be able to provide capabilities to classify the type of the packet generated by mobile data applications.

The system should be able to use optimized service delivery mechanisms for different types/classes of data application packets.

The system shall be able to provide mechanisms to optimize the traffic due to large number of live streaming sessions for the same content from a given source outside of the 3GPP network (e.g. merger of unicast streams delivering identical content).

Mechanisms shall be provided which allow the network and UE to detect abnormally high data patterns and to provide countermeasures to protect the network and subscribers from data surges that are caused, either intentionally (e.g due to design) or accidentally (e.g. due to bad implementation), by Mobile Data Applications.

The system shall be able to provide mechanism to efficiently manage the delivery of simultaneous push services（e.g. by considering the network status and timing requirements associated with each push service）.

7 Conclusion

This TR has discussed and highlighted that as more and more mobile data traffic is handled by the network, the signalling traffic associated with various non-MTC mobile data applications can cause high signalling impact on the network and lead to poor user experience. Also, the TR has identified the use cases, issues and potential requirements related to Small data packet, Frequent keep-alive and Status update messages, Frequent start of service, Overload caused by live content streaming, network congestion by push services, and Distributed Denial of Service.

It is recommended to start the normative work in SA1.

Annex A:
Example for estimating frequency of Facebook content/status update messages

An example is provided on how to estimate average frequency of receiving Facebook content/status update messages:

a. # of average content items generated per user

a1. 30 billion items of content generated by all users per month[§] / 500 million Facebook users[§] = 60 per user per month

a2. average # clicks of "like" button = 9 per user per month[¥]

a3. average # of comments written = 25 per user per month[¥]

a= a1+a2+a3 = 94 content updates per user per month

b. average # of friends per user = 130[¥]

c. average number of received facebook content/status update messages per user per month = a*b

= 94*130 = 12220 per month => 407.3 per day => ~17 per hour (once every 3.5 minutes)

As a result, we estimate the average frequency of receiving Facebook content/status update messages is in the order of a few minutes.

[§] http://www.facebook.com/press/info.php?statistics

[¥] http://www.digitalbuzzblog.com/facebook-statistics-facts-figures-for-2010/

Annex B:
Example of data traffic model for a mobile IM application

The following data has been gathered from a certain data hotspot of a mobile operator's 2G GPRS network to illustrate the impact of an IM application (App A) behaviour in the mobile network.

Note: In the following figures, "All" includs IM from App A, Web browsing application, and the remaining applications, e.g. SNS applications, news notifications. The values in the following figures are all mean values.

Activity duration is the time between the first and last packets of an closely spaced activity of one usage of such service. Activity inter-arrival time (Activity IAT) is the time between two consecutive activities belonging to the same user and same traffic category. From the following figure B-1, activity duration and activity IAT of IM from App A are shorter than other applications (e.g. web browsing), which means that users tend to use IM application more frequently, but the duration is shorter than other applications.

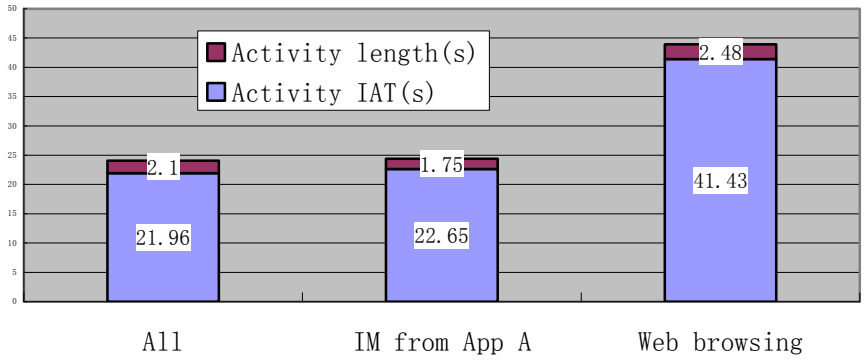

Figure B-1 Activity duration and Activity IAT

Activity Burst Size is the sum of all user plane payloads, including IP headers within a packet burst.

From the following figure B-2, activity size of IM from App A is much smaller than other applications, e.g. web browsing.

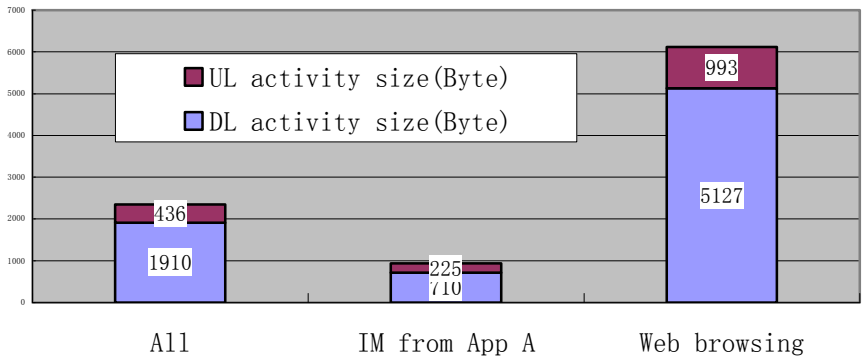

Figure B-2 Activity Burst Size

Activity time is total seconds of active usage of the network for a subscriber in a busy hour. From the following figure, IM from App A has a long activity time, which explains why it contributes to a big part of the total traffic volume despite its low intensity.

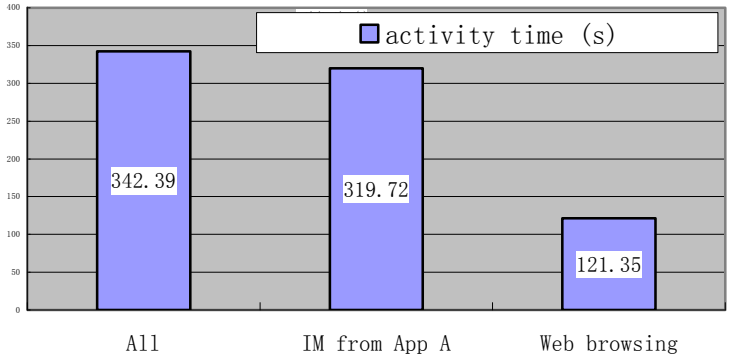

Figure B-3 Activity Time

Some testing data has been accumulated in live cells of 2G GPRS network to monitor signalling failures and it was found that a considerable percentage of failures took place due to signalling channels being congested and a large number of active IM users were frequently activating and deactivating PS sessions.

Some data has also been gathered in live cells to monitor the utilization rate of bearer channels (PDCH, Packet Data Channel which is used to carry payloads). It was found that the utilization rate of PDCH deteriorated to a considerable extent, where there were a large number of active mobile IM users generating a huge number of small packets and reducing radio channel utilization rate.

Annex C:
Data application impact on real UMTS networks

Impact of PS signalling on the network

An analysis was obtained by monitoring network performance counters at urban RNCs of two different operators across multiple days. Network resource utilization, traffic, and signalling counters are analyzed.

Breakdown of DL power at the base station in Release 6

Figure C-1 is a graphic representation of downlink power utilization obtained on an UMTS Release 6 network consisting of about 600 cells. The measurements were conducted over a period of several weekdays (excludes weekends) during the time segments when traffic load is high - from 8 am to 10 pm in each of those days.

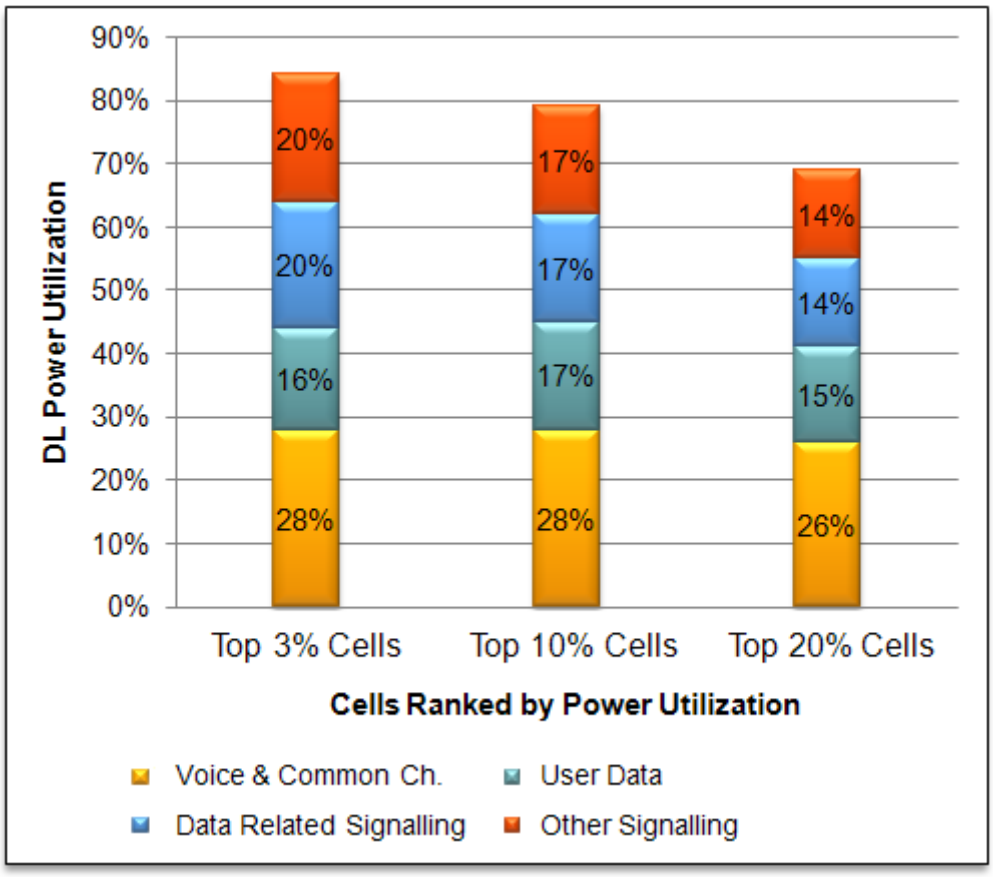

Figure C-1. DL power utilization

The statistics shown were calculated from network counters, which compute total energy used, yielding average power. The three categories of cells shown are based on maximum power utilization (e.g. “Top 10% Cells” refers to the 60 of the 600 cells that were most heavily utilized in terms of downlink transmit power). The percentages shown are relative to maximum DL power rated for the NodeB transmitters.

Transmit power is broken down into 4 categories as shown:

“Voice & Common Ch.” includes power of common/broadcast channels (CPICH, SCH, SCCPCH (PCH, FACH), etc.), in addition to power required for carrying AMR voice frames

“User Data” refers to power consumed by user payload carrying packets

"Data Related Signalling" refers to power for control plane signalling associated with PS Data (non-voice) communication (call setup and in-call signalling), i.e. signalling related to setting up connections, scheduling user payload transmisisons, and connection state transitions direclty relatable to transmission of user payload data

"Other Signalling" refers to power for signalling associated with other services (Voice, SMS, Registrations, etc.)

R99 PS Data traffic is negligible in this network. Hence, power associated with R99 PS Data is not shown in Figure C-1.

Similarly, simultaneous PS and CS sessions constitute a small part of total traffic and they are captured under PS Data power.

Median CS call holding time is 60 – 80 seconds.

Some key data and findings are that, for the 10% of the cells that are the most heavily utilized,

34% of DL power is used for signalling

PS Signalling accounts for 50% signalling volume, 17% of DL power

Other signalling (registration, voice, SMS) accounts for another 17% of DL power

In summary, Data Related Signalling accounts for a very significant part of the signalling volume/impact. That impact is most pronounced in heavily utilized cells, where PS control signalling energy is equal or larger than user payload carrying energy.

Expected performance in Release 8

3GPP Release 8 (HSPA+) features aim to reduce PS Data related signalling .

Performance enhancing features introduced in Releases 7 and 8 define new UE connectivity states and allow for longer UE state transition timer settings, so that observed user activity patterns do not translate to parallel radio access network signalling associated with state transitions. These performance gains vary somewhat with timer settings and cell-to-cell (e.g., cell with predominantly high speed vehicular traffic vs. cell with predominantly indoor coverage).

Disproportionate impact of PS data communication

Another very relevant finding is that PS data communication activities have an impact on the UMTS that is higher than what the penetration of UEs designed for PS data applications in the network would suggest. In other words, as showed for example in Figure C-2, even a modest penetration (16%) of these UEs causes a very large presence (59%) of PS Calls. As the number of applications for data centric devices increases, and MTC device proliferation ramps up, the number of PS calls per UE is likely going to be increasing over time. For reference, the data presented in the Figure C-2 were collected in the spring of 2010.

Figure C-2 represents data collected using OSS performance monitoring tools in urban RNCs of selected networks: Operator A network segment monitored is comprised of 170 cells, and Operator B of 480 cells.

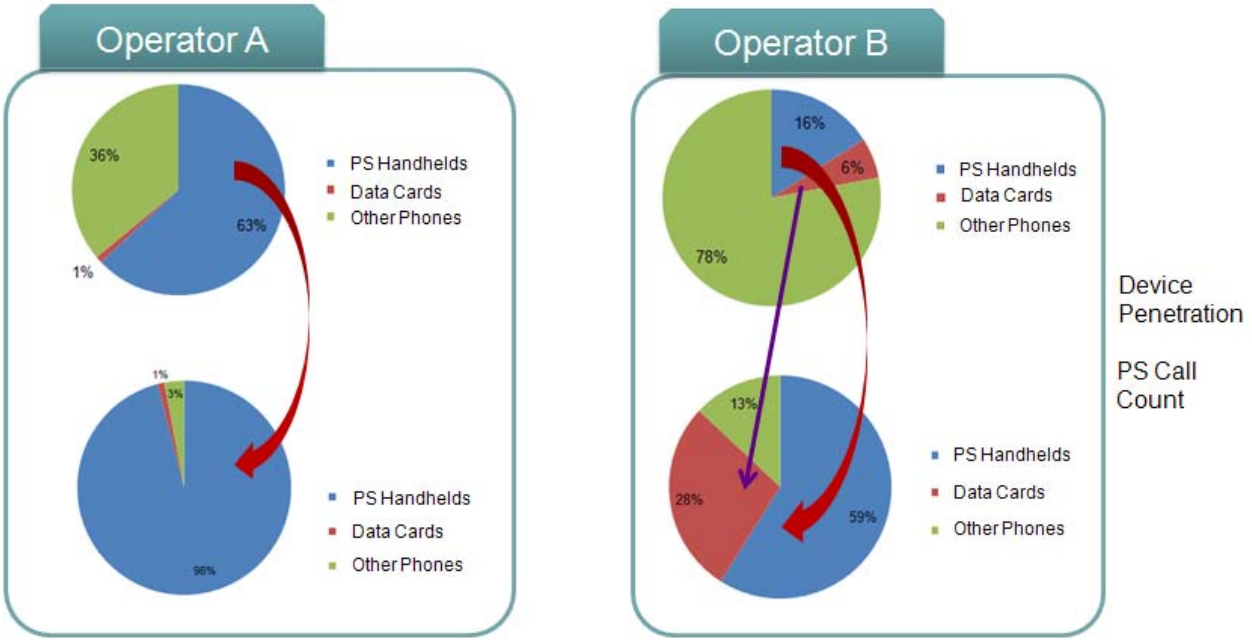

Figure C-2. Device Penetration & PS Call Count

Field data statistics

The following results represent statistics for all devices in the selected RNC's collected over a 24-hour period.

Distribution of PS Call Duration

Figure C-3 shows statistical distribution of duration of PS calls generated by data communication UEs. PS call is defined by signalling messages from RRC/RAB setup and PDP activation, until RRC release.

As expected, many PS calls are "short". Additionally, sharp peak in call duration probablity at around 8 seconds in both networks seems to suggest battery energy saving techniques as a likely cause. Battery energy saving technique refers to the behaviour of some pre-R8 terminals, which, upon completion of a data exchange procedure, would autonomously go into a dormant state, thus releasing the RRC connection. While this UE implementation-dependent (proprietary) action of autonomously entering dormancy is advantageous in terms of conserving UE battery power, it would tend to increase the network impact since it may result in additional signalling due to connection release, only to be followed up shortly thereafter by connection re-establishment, as user activity resumes.

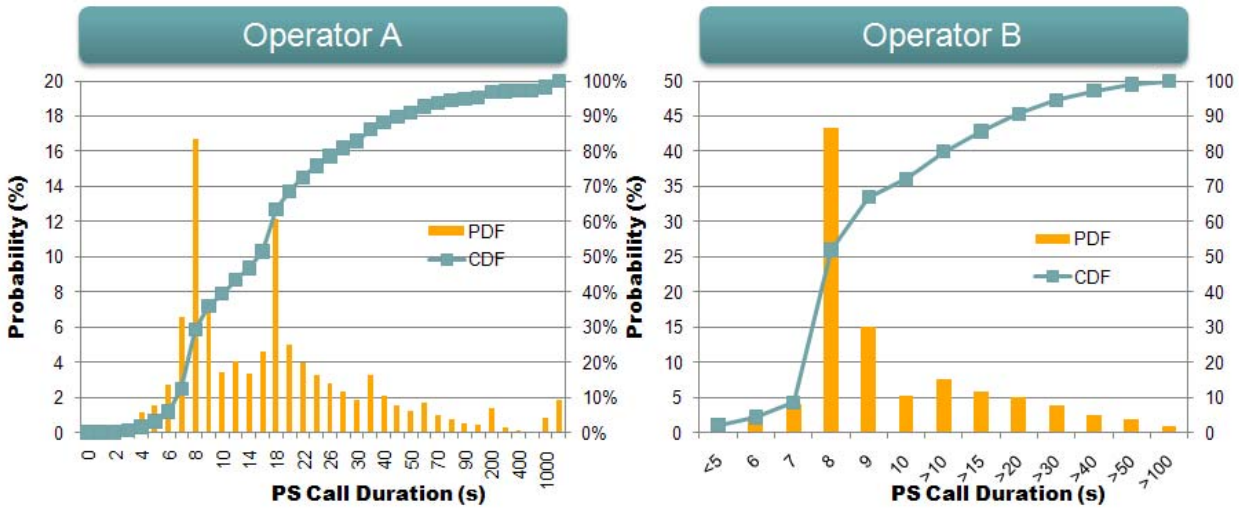

Figure C-3. Distribution of PS Call Duration

CDF: Cumulative Distribution Function

PDF: Probability Density Function

Data Volume of PS Calls

Figure C-4 shows distribution of data volume of PS calls. A large majority of calls is of very small volume: for both examples, a high percentage of calls transfer one kilobyte of data or less. This behaviour is typical of prevailing data applications such as push e-mail, social networking and instant messaging. Activities such as keep-alive messages, status reports, and polling contribute to such behaviour.

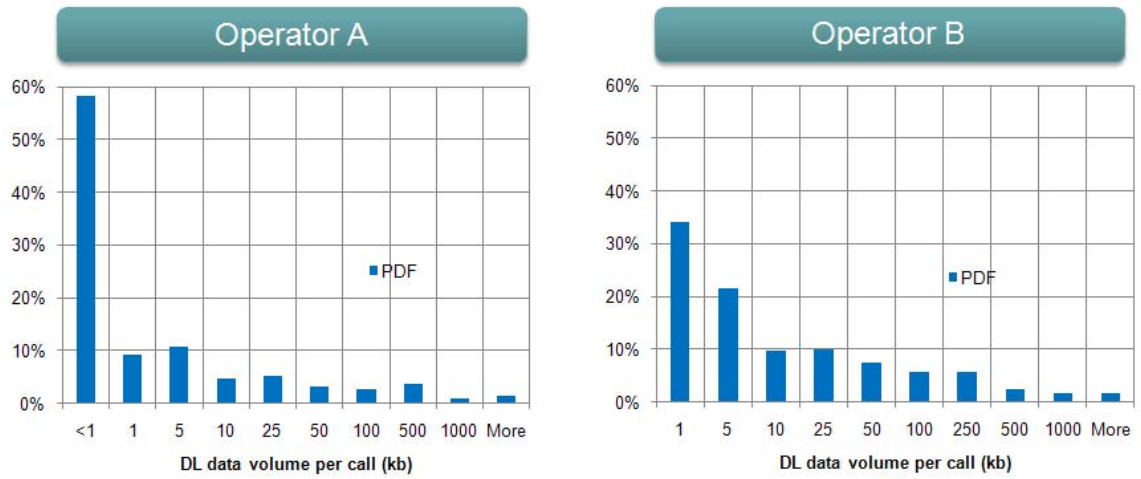

Figure C-4. Distribution of PS Call Volume

Note that the statistics shown are derived from several millions of PS calls taken over the two networks (one in Europe, another in North America) over several weekdays during March-May 2010 timeframe. Thus, the data represent a statistically significant sample for the timeframe in question.

Inter-arrival time of PS Calls

Figure C-5 shows distribution of inter-arrival times of PS Calls for a UE, compiled over many calls.

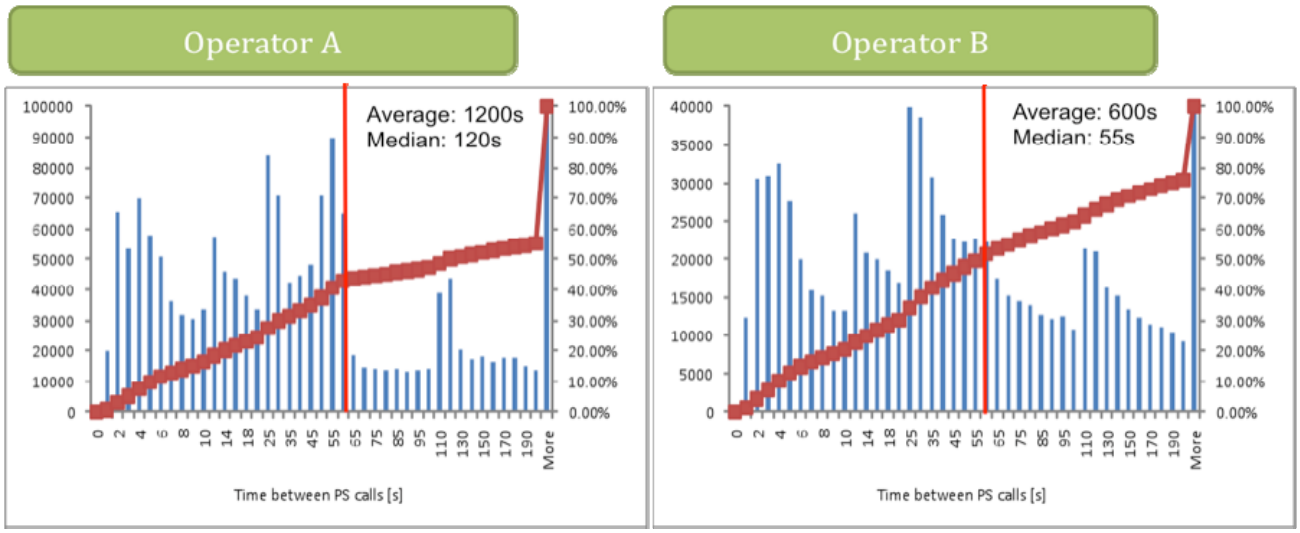

Figure C-5. Distribution of Inter-Arrival Time of PS Calls

The red curve shows CDF of call inter-arrival time (secondary y axis). The purpose of the vertical red line is to show that ~50% of calls occur within 60 seconds of the previous one. This reflects typical user behaviour of closely spaced data communication sessions when using a data communication device actively. The plot additionally shows how "chatty" data applications traffic is.

Annex D:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2011-02	SA1#53				Draft skeleton for review		0.0.0
2011-02	SA1#53				Updated with contributions at SA1#53	0.0.0	0.1.0
2011-05	SA1#54				Updated with contributions at SA1#54	0.1.0	0.2.0
2011-08	SA1#55				Updated with contributions at SA1#55	0.2.0	0.3.0
2011-09	SA#53	SP-110586			Raised by MCC to v.1.0.0 for presentation to SA	0.3.0	1.0.0
2011-11	SA1#56				Updated with contributions at SA1#56	1.0.0	1.1.0
2011-12	SA#54	SP-110817			Raised by MCC to v.2.0.0 for approval by SA	1.1.0	2.0.0
2011-12	SA#54	SP-110817			Raised by MCC to v.12.0.0 after approval by SA	2.0.0	12.0.0