

Research Article

Challenges for Coexistence of Machine to Machine and Human to Human Applications in Mobile Network: Concept of Smart Mobility Management

Rajarshi Sanyal,¹ Ernestina Cianca,² and Ramjee Prasad³

¹ Engineering Expert (Roaming and Signalling), Belgacom International Carrier Services, Rue Lebeau 4, 1000 Brussels, Belgium

² Department of Electronic Engineering, University of Rome Tor Vergata, via del Politecnico 1, 00133 Rome, Italy

³ CTIF, Aalborg University, Niels Jernes Vej 12, 9220 Aalborg, Denmark

Correspondence should be addressed to Rajarshi Sanyal, rajarshi.sanyal@bics.com

Received 23 April 2012; Revised 9 August 2012; Accepted 13 August 2012

Academic Editor: Jianhua He

Copyright © 2012 Rajarshi Sanyal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A key factor for the evolution of the mobile networks towards 4G is to bring to fruition high bandwidth per mobile node. Eventually, due to the advent of a new class of applications, namely, Machine-to-Machine, we foresee new challenges where bandwidth per user is no more the primal driver. As an immediate impact of the high penetration of M2M devices, we envisage a surge in the signaling messages for mobility and location management. The cell size will shrivel due to high tele-density resulting in even more signaling messages related to handoff and location updates. The mobile network should be evolved to address various nuances of the mobile devices used by man and machines. The bigger question is as follows. Is the state-of-the-art mobile network designed optimally to cater both the Human-to-Human and Machine-to-Machine applications? This paper presents the primary challenges for the coexistence of M2M and H2H devices in a mobile network and draws emphasis for revisiting the mobility management aspects and congestion control in the-state-of-the-art network. Further, we set out a mobile network architecture with smart mobility management which aims to reduce the signaling interaction between the device and the network to optimise the power and bandwidth.

1. Introduction

The concept of Machine to Machine (M2M) communications fits into the new trend of devices that we see around us disseminating information through the mobile network to the remote peer or to the cloud. These objects have their own Internet protocol addresses. They are embedded in complex systems and interfaced with sensors to obtain information from their environment (e.g., food products that record the temperature along the supply chain).

Some guidelines are given for particular aspects of this M2M technology, such as ETSI's TS 102.689 [1]. But a unified approach contemplating the coexistence and seamless interoperability between the different device types is still "work in progress." This poses a risk, as the state-of-the-art mobile networks were designed keeping in mind Human to Human (H2H) communication. Our contention

is that the state-of-the-art mobile network may not be resilient to the data surge likely to be caused by M2M applications. It is probable that a group of M2M devices for a particular industry application may initiate services and access the mobile network at the same time, causing network congestion and degrading the quality of other services. For example, a fleet management company may have moving carriers in different geographical locations trying to access the network and upload data at the same instance. Considering the high proliferation of these devices, an avalanche effect on signaling and data traffic can be generated, thus impacting the gross QoS (Quality of Service) offered by the network. It is imperative that the mobility management processes need to be revisited in order to cope up with the surge in the mobility management messages in the mobile network and to maintain the same QoS without being heavy on the bandwidth. Attempts should be made to

reduce the interaction with the network and to simplify the mobility and location management-related processes.

This paper addresses the technical architecture of a Smart Mobile Network Access Topology (SMNAT). We call it smart because it is

- (i) lite and does not engage complex application layer processes for mobility management and thereby reduces signaling overhead and wards off congestion,
- (ii) agnostic to the device type: no intricate network adaptations, scheduling, topology control for M2M applications to thwart congestion,
- (iii) identifying and addressing the mobile nodes directly at the physical layer.

SMNAT covers broad implementation scenarios, mainly in the cellular network domain and is not confined to explicit setups that we witness in ad hoc or sensor networks comprising of nodes with self-organisation capabilities. Hence there are no tailor-made amendments in the framework or operational procedures for specific scenarios (like swarm networks as an example). Topology control (physical or graphical) is not implemented as we assume in this paper that the M2M devices will acquire the mobile network like any other H2H device and will not establish direct multipath/multihop communication between the wireless nodes. The purpose of SMNAT is to ensure that the coexistence of H2H and M2M devices in the same network does not impose any collateral impairment. It is designed to abridge the intricate mobility and location management processes as well as the addressing principle to render more operational efficiency and cater the needs of the M2M devices.

The rest of the paper is organised as follows. In Section 2 we dwell on the issues in addressing and location management due to the coexistence of the M2M and H2H devices in the “state-of-the-art” network. In Section 3 we describe the smart mobile network essentials. The unique addressing scheme of SMNAT implementing the physical layer, the mobility and location management processes are subsequently discussed. In Section 4, comparative analysis of the signaling processes and the overhead due to mobility management processes between SMNAT and state-of-the-art is made; Section 5 addresses the key issues with the available mobile network topologies and how they are addressed/resolved by the smart approach. Conclusions are drawn in Section 6.

2. Challenges in Mobility Management in the Homogeneous State-of-the-Art Mobile Networks for H2H and M2M Applications

Challenges faced by the present mobile network in the light of the M2M scenarios are as follows.

- (1) Number of M2M devices can be fairly large compared to the H2H devices. So the length of the addressing parameters that are used today (MSISDN, IMSI, or even the IPV6 addresses) seems to be insufficient to identify the high volume of devices in the network.

- (2) Power management will play an important role. For example, an M2M device meant for location-based application in a freight container will need to be sustained for weeks, if not months, may be without an external constant power source. So the battery power should be carefully conserved, and one of the effective ways to do so is to reduce the signaling interaction between the handset and the network to maintain location and presence information.
- (3) Variation of QoS requirements: some real-time applications have high demands on latency and reliability (as e.g., accident sensors). While the other applications like reporting services may be contended with a lesser priority scheme but with a higher bandwidth requirement.
- (4) Different traffic patterns: some M2M devices may not transmit any data for months, for example, a crash sensor. While other sensors may transmit data continuously or periodically.

The foremost challenge for 4G is to support seamless interoperability, handover and roaming for the H2H and M2M devices without compromising the “grade of service,” “round trip delay,” and latency. It has to meet the desired reliability for the M2M application, while supporting the high throughput demand of a wide variety of H2H multimedia services such as multimedia web browsing, video, online gaming. Various methods [2–6] for optimising bandwidth have been worked out for mobile and 802.1X networks like dynamic bandwidth allocation according to location/QoS. Some methods [7] adhere to the principle of fixed slot allocation for all cells which is not optimal in terms of bandwidth utilisation given the fact that different device types coexist in the network with different nature of bandwidth requirement. Also, the traffic pattern varies over time and location area [7]. Due to the high proliferation of the M2M devices, there may be significant impacts like network congestions, which will in turn affect the H2H services. The M2M scenarios are characterized by colossal amount of devices that interact frequently or infrequently by small amounts of data. These devices may be clustered in a small zones leading to competition amongst the network nodes. A majority of the devices can initiate connection leading to peaks in signaling and data. This penalizes the non M2M devices.

The architecture of the IMS/LTE networks [8–10] is expected to keep large use of IP and use a large spectrum of core and access technologies, stemming from the necessity to cater various types of mobile applications depending on the device type/capability. Wireless 4G networks create various challenges due to their architectural heterogeneity in terms of differences between access schemes, resource allocation techniques, and QoS requirements. To address these challenges, the existing proposals require a significant modification or new Medium Access Control (MAC).

The primary issues that may impact the state-of-the-art mobile network due to M2M communication are as follows.

Radio Network. The eNodeB in the LTE network needs to connect a large number of M2M devices. The devices will contest to use the same channel leading to collisions.

Core Network. The MME needs to attach a large number of devices. The network also needs to provide the IP and NSAPI address to those devices. A lot of devices imply sporadic and transient use of the bearer, leading to an overhead. The HSS needs to have the subscription and profile definition for the M2M devices. There can be signaling congestion towards HSS when a large number of devices try to register in the same HSS.

In M2M networks, we seldom experience congestion in the data plane. This is because devices send and receive small amounts of data. But a lot of devices transceive data simultaneously leading to congestion mainly in the Enhanced Packet Core (EPC) part between the Serving Gateway (S-GW) [10] and the Packet Gateway (P-GW) [10]. Capacity augmentations need to be done for adjuncts like MME, HSS, AAA servers, CDR processing engine.

The goal of 4G is to replace the current proliferation of core mobile networks with a single worldwide core network standard, based on IP for control, video, packet data, and voice. The IP protocol used is IPv6, which uses 128 bit addresses vis-a-vis the 32 bit addresses of IPv4 protocol. Hence the IP V6 protocol which is directly used for mobility management in 4G can offer a wider address pool for the mobile users. This is essential to accommodate all the M2M devices in the mobile network. To enhance mobility in IPv6, “micromobility” protocols (such as Hawaii, Cellular IP, and Hierarchical Mobile IPv6) have been developed for seamless handovers that is, handovers that result in minimal handover delay, minimal packet loss, and minimal loss of communication state. However, the core issues of IPv6 to be involved in mobility management still remain.

Three major issues are highlighted and the improvisations of the solutions are as below.

Issue 1: Paging Support. The base IPv6 specification does not provide any form of paging support. Hence to maintain connectivity with the backbone infrastructure, the mobile node needs to generate location updates every time it changes its point of attachment, even if it is currently in dormant or standby mode. Excessive signaling caused by mobility leads to a significant wastage of the mobile node’s battery power, especially in environments with smaller cell [11]. Therefore, it is impractical to rely completely on location updates, and it is essential to define a flexible paging support in the intradomain mobility management scheme.

In order to save battery power consumption of devices, IP paging is proposed as an extension for Mobile IP [12]. Under Mobile IP paging, a device is allowed to enter a power saving idle mode when it is inactive for a period of time. During idle mode, the system approximately knows the location of the mobile node within the paging area comprised of multiple subnets. A device in idle mode does not need to register its location when moving within a paging area. It performs location update only when it changes paging areas. However, with the high penetration of the M2M devices, the size of the

paging area will reduce. This will culminate in high signaling network load when the M2M device (say, implemented for a fleet management company) alters the location frequently.

Issue 2: Round Trip Delays. According to the IPv6 implementation principle for mobility management [12], the mobile node needs to send binding updates to the home agent in a given periodicity. This action implies that an authentication procedure needs to be actuated for each bind updates resulting in an increase in round trip delays.

For these reasons a new Mobile IPv6 node, called the Mobility Anchor Point (MAP), has been suggested in RFC 4140. MAP can be located at any level in a hierarchical network of routers. The MAP will limit the amount of Mobile IPv6 signaling outside the local domain. As an evolution from IPv6, Hierarchical Mobile IPv6 (or HMIPv6) proposes the implementation of the MAPs. Networks are divided into domains and subnets, with each administrative domain having a Mobility Anchor Point (MAP) at the highest level. Intradomain mobility of a mobile host is handled separately from interdomain mobility.

When the Mobile Host (MH) changes points of attachment within the same domain, the MAP of that domain is informed of the change in care of address of the MH through binding updates. Binding updates are also sent to correspondent hosts within the same domain.

This process reduces signaling traffic due to lesser binding updates. Handoff latency also decreases as “far-off home agents” and “correspondent hosts” need not be updated every time a mobile host changes point of attachment. IPv6-based mobility management ensures minimal handoff latency to achieve better QoS for real-time data conveyance.

However HMIPv6 has its drawbacks as the hierarchical, addressing model may impact the real-time applications. Due to the hierarchical nature of the protocol, each anchor point is only aware of the next anchor point down in the hierarchy. Each node stores a mapping of source destination addresses of the previous and next nodes in the hierarchical structure. These addresses are called VCoAs (Virtual Care of Address) [13]. Only the lowest anchor point in the hierarchy stores a mapping of VCoA to PCoA which is the physical care of address of the MH in the foreign environment. As the source is aware of only the VCoA of its nearest anchor point, it sets that address in the destination field header of the IP packet. At each hop, the packet is processed, depending on the source address, the new destination address (VCoA) is decided. The packet is then forwarded using the new destination address and the node’s own address as the source address. This processing occurs at each hop and can effect real-time applications running at the mobile device as it creates significant delay especially if the CH (Correspondent Host) and the MH are many hops away with the MH part of a big hierarchical structure.

Issue 3: High Signaling Exchange for Mobility Management. With the evolution of the mobile networks as we find from 2G to 4G [8] we attempt to inject more symbols (data) per unit time. This has been made possible by the evolution of the modulation schemes from BPSK in case of 2G, to

64 QAM for LTE-Advanced. In the current Mobile Radio Access Network [14–16] there is a host of network processes, tightly synchronized and orchestrated by intelligent network elements, such as:

- (i) handover management,
- (ii) location management,
- (iii) call drop off management,
- (iv) interoperability and downward compatibility management,
- (v) service control (like roaming control), and
- (vi) feature management.

These processes involve complex signaling operations across the radio and the core network. The attempt to simplify the network in IMS [9] and LTE [10] is focused to make the core and access networks all IP. Following the evolution trail from 2G to 4G, we do not discern a significant philosophical drift towards simplification in terms of the service logic related to mobility and location management. We see a similar process for handover, frequency reuse, location updates, and cancellations in 4G [10] as compared with 3G/2G [15]. Hence the network elements of the state-of-the-art still need to be equipped with the intelligence and processing power to handle all these complex signaling operations.

With the recent initiatives from the European Commission [1], LTE-A network will be extended to support the M2M devices. The IPV6 addressing mechanism enables a wider device reach. But there is still no evidence of significant emphasis on reducing the mobility management processes for optimised operations.

2.1. Related Work. A lot of proposals exist for optimising the access network and the nonaccess stratum (NAS) to adapt to the machine-type communications (MTC) scenario [17, 18]. Some proposals on the access layer end dwell on realisation of an adaptive RACH [19] to reduce RACH collision probability, to control network overload, and to enhance system performance. Some other solutions [17] eyeing for optimisation at access layer are time backoff classes, slotted access, and group coordination. 3GPP has a focus on adapting the network architecture and the associated parameters to be more compatible with the M2M devices. The primary goal is to provide a long-term solution for addressing mechanism and the congestion. Network improvements for MTC (NIMTC, Rel. 10) [17] and system improvements for MTC (SIMTC, Rel. 11) [18] are two mentionable developments.

In Rel 10 for, we have witnessed some improvisations namely,

- (i) RAN overload control by Extended Access Barring (EAB),
- (ii) APN-based prioritisation and congestion control, and
- (iii) throttle control.

In Rel 11, the adaptations are

- (i) avoidance of E.164 number allocation for the M2M devices,
- (ii) dedicated interface for the MTC Server to ward off link congestion in the AAA and for MSISDN (E.164) less device triggering.

There has been significant work done by the research community on congestion control. These proposals [20–22] deal with scheduling algorithms in the access part and channel reservation.

In [23], a group-based mechanism was proposed. Within the group there will be a cluster head in a device area. All the signaling information passes through the cluster head towards the eNodeB (BTS). Due to aggregation mechanism, the traffic distribution is more optimised resulting in reduction of the signaling overhead.

Most of the solutions worked out so far are poised to reshape the existing mobile network in light of the M2M scenario by implementing scheduling logic, new congestion control mechanism, defining domains and cluster-heads, proposing new device addresses, and so forth. Many of these solutions attempt to reduce network congestion caused by the proliferation of the M2M devices. We attempt to shift from this paradigm by proposing SMNAT which is agnostic to the device type. We try to nip the problem in the bud by designing an access methodology which will eliminate the application layer messages itself for mobility and location management, rather than devising relentless network adaptations by standard bodies for M2M.

3. Introduction to SMNAT

With the SMNAT, we are able to perform specific layer 5 and 7 processes related to addressing and identification of a node in the RAN, right at the physical layer [24–26]. This relieves the application layer from the arduous processing in the context of the mobility management and addressing. It wards off the continuous interexchange of signaling data between the device and the network for reducing congestion.

A mobile node is provisioned with 3 network parameters, namely, the symbol coordinate, time slot, and the primary AFCRN, which are essential for channel acquisition and for processing the call or other network operations [26]. Figure 1 elaborates these parameters. The available bandwidth is divided in multiple subfrequency bands called AFCRN (Absolute Radio Frequency Channel Number). The AFCRNs are numbered in numerical order. Each AFCRN is linked to one Time Frame. Following completion of the addressing procedure, the data channel or a time slot for the traffic conveyance is identified. The access layer is involved with limited layer 7 functionalities [25]. A new entity called Coordination Processor (CP) [25] sits in between the core and access network domains. It facilitates addressing and essential signaling functionalities for mobile user node at the physical layer level. It also takes care of some layer 7 functionalities that are still required for AAA (authentication, authorization, and accounting), supplementary services, and

value-added services. The RAN comprises ad hoc repeaters, referred to as Access Points in the network area implementing TDMA/FDD multiple access.

In [26] we presented the modulation scheme, time frame structure, and the addressing principle of SMNAT. The modulation scheme is a blend of M1 PSK (for outer ring) and M2 PSK (for inner ring), where $M1 < M2$. Figure 1 shows the constellation diagram in the case: $M1 = 8$ and $M2 = 4$. The outer ring comprises of M1 symbols for user traffic, and the inner ring comprises of M2 symbols that will be used for addressing the mobile users. In the proposed multiple access scheme, a user is identified in the network with respect to the symbol coordinate of the outer ring in the complex plane. The constellation diagram in Figure 1 pertains to a specific time slot, $T1$, of the frame. The frame corresponds to a specific AFCRN. Symbols of the outer ring are conveyed at a rate of fr and the symbols of the inner ring are exchanged with a much higher rate denoted by fd . In particular, we could set $fd = 8 * fr$. When the Mobile Station (MS) is provisioned in the network, it is allocated a specific AFCRN, a specific time slot in the TDMA frame, and a specific symbol coordinate pertaining to the M1-PSK. After the addressing part is completed by actuating a layer 1 handshake, the time slot is used for data transfer. There is no separate time slot for conveying the control signaling and the data part. A single-time slot can carry M1 symbols, used for addressing M1 users. Later when the time slot is seized, it will be used to convey the symbols for data traffic.

The network based on SMNAT consists of Access Points (APs) which transceive the data to the local multiplexer. Multiple local multiplexers converge the data towards an aggregate multiplexer, which in turn connects to a central processor called Coordination Processor (CP) [25, 26]. The CP can be compared with the time/space switching matrix 2G/3G network, MME of LTE [10]. It bridges the access and the core network and is responsible for formulating the time frames actuating physical layer addressing in the forward and reverse channels. Figure 2, demonstrates the end-to-end handshake process for call initiation by A party and termination to B party. All the associated network elements in the call trajectory and the actions taken by each of them in the process are also furnished in Figure 2.

When a mobile node A calls mobile node B, the mobile node A initially scans for the time slot in the time frame pertaining to primary AFCRN to check whether it is free to carry any symbol for a new operation, or is already in use. If it finds that the time slot is free, then it injects a symbol for itself and sends it in the uplink channel. The mobile station includes the specific symbol value in the given time slot and forms a frame. It synchronises with the access point and synchronises with the frame before populating the symbol in the time slot. The frames pertaining to all the AFCRNs are aggregated by the Central Mux and finally fed to the Coordination Processor. So the layer 1 message with the specific symbol for node A reaches the Coordination Processor. The Coordination Processor injects the same symbol in the same time slot and AFCRN in the reverse direction (downlink), which acts as a layer 1 page response. If the AFCRN is busy and not available, the response message is

not sent back. The mobile node has a timeout and retries the same process (i.e., invoke a page message), but with another AFCRN.

After the Coordination Processor generates the response in the downlink and the Node A receives the page response, subsequently the time slot between the Node A and the Coordination Processor is seized for layer 7 signaling data and traffic data transport. Subsequently, a layer 7 message is transmitted by the mobile node 1 via this “seized” time slot containing the A and B number plus the information elements with the details of supplementary services of node A and the authentication parameters. Once the CP obtains the B number, it queries the HSS/HLR over DIAMETER (S6a)/GSM MAP interface mainly for 3 objectives:

- (i) authentication and authorisation of mobile node A,
- (ii) to resolve the essential network parameters provisioned for B, namely, the symbol coordinated, time slot, and primary AFCRN in order to invoke the downlink message,
- (iii) To check the available supplementary services/tele-services for mobile node B, only in case mobile node B is provisioned in the same network. This can be analysed by the Coordination Processor by examining the dialing plan.

After the signaling exchange with the HSS/HLR, the Coordination Processor generates a layer 1 message to B node. If the primary AFCRN provisioned for Node B has been seized before by another user, the Coordination Processor tries with another AFCRN according to the AFCRN scanning procedure. Node B listens to the time slots in all the relevant AFCRNs, whether a symbol (for itself) has arrived. When it finds one, it generates a layer 1 message with the same symbol and in the time slot towards the CP. After the CP receives the information A, it comprehends that it is a response message from the B party. It distinguishes between a response message and a new call request, as it maintains the call context immediately after generating a page signal in the downlink and starts a timer to await exactly the mirror message as generated in a different time frame. After this process is completed, it is acknowledged both by Node B and the CP the time slot is available for data transfer between the two entities and is now sampled at a greater rate by sub dividing in more time slots. At this stage, a layer 7 message is invoked by the CP towards the Node B to let it know all the details of the call, like A party number, B party number, supplementary services, and so forth.

Note that the procedure explained above is actuated only when the mobile node or the network commences a network operation, like mobile originating/terminating calls, data service, SMS, or other value-added services. However, the mobile node does not get attached to the network and hence there is no periodic location updates, location update during network acquisition, or location update due to change in cell. The network does not keep a track of the location or presence of the mobile node in the network area. This helps in substantial reduction of signaling messages due

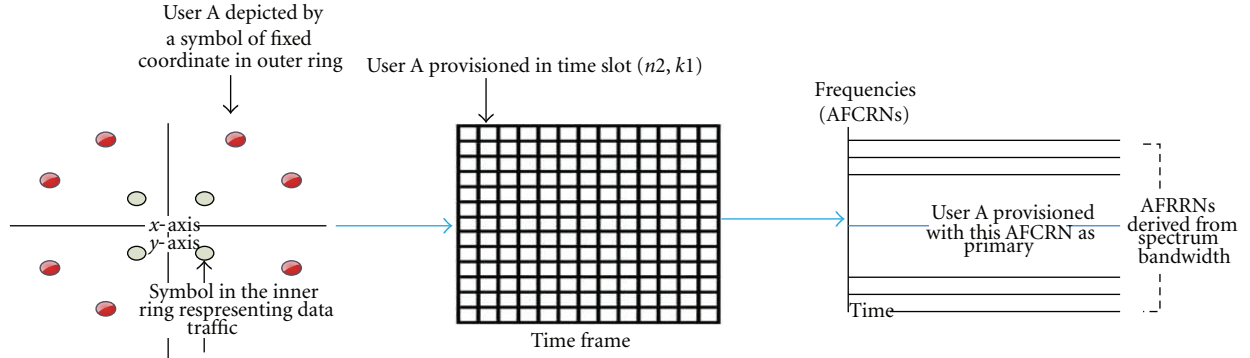


FIGURE 1: Parameters at the physical plane allocated for the mobile users.

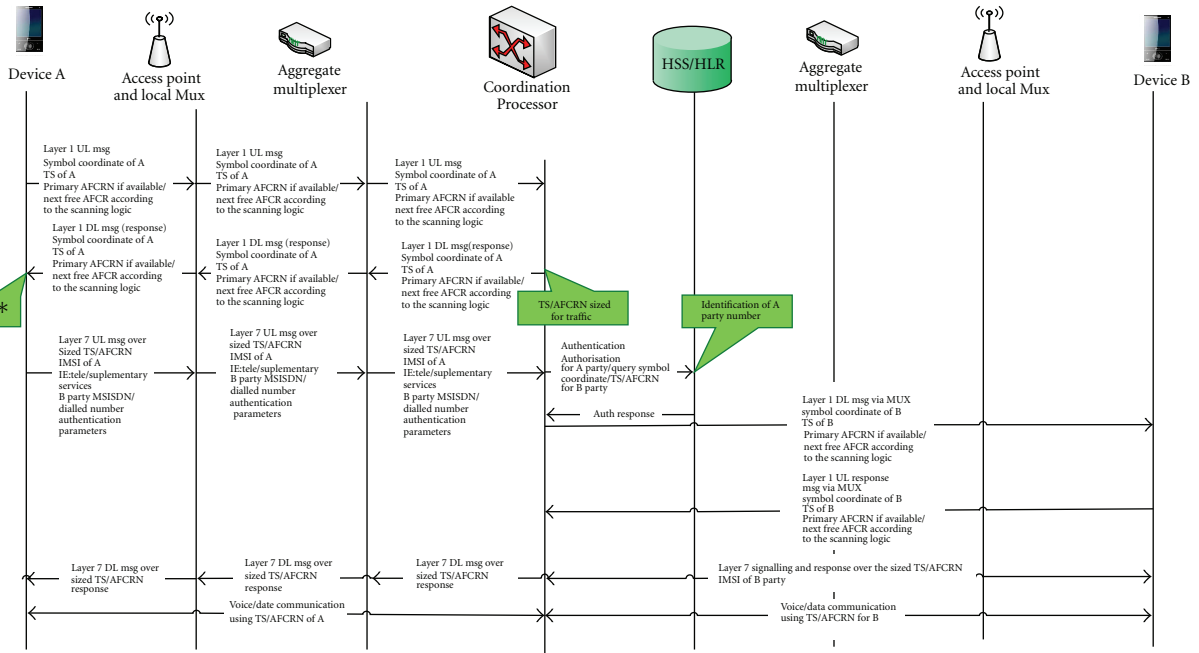


FIGURE 2: End-to-end call flow between A and B parties.

to mobility management. Section 3.1 covers this topic in details.

3.1. Mobility and Location Management. The cellular network has a checker board design as shown in Figure 3. Each block represents a square-shaped cell which is not identified by a Cell Global ID. Instead they are assigned Geographical Information System (GIS) ID. This parameter is not parsed in any signaling messages related to mobility and location management. So the network is agnostic to the exact location of the mobile station. The white blocks are allocated AFCRN from 1 to n , while the black ones are allocated AFCRN from $n + 1$ to k .

Say we consider $n = 64$, $k = 128$ in this paper to keep parity with allocation methodology in traditional GSM. In the mobile station, the values of n and k are known, All the AFCRN in white shade are mapped sequentially to the

AFCRN in the black shade in the mobile station. When a Mobile user is provisioned in the network, a primary AFCRN is allocated. The primary AFCRN pertains to a given shade of the cell, either white or black. If the mobile station attempts to invoke a message or respond to a page in a cell with opposite shade, the AFCRN mapping process (between white and the black shades) as above is invoked.

Each mobile station is fitted with 2 transceivers [25, 26]. There is an active section which takes part proactively in the network operations within the cell where the mobile user is currently located. The dormant part keeps on scanning the adjacent cells and is mainly responsible for actuating the handover procedure as in Figure 3.

A central processor in the mobile station monitors and compares the signal strength as perceived by the 2 transceivers, the active one which is already in a call and the passive one which compares the pilot. The MS also has an inbuilt GPS unit, and if in open air (with LOS with

the satellite) can also determine the direction and speed of motion. It also has an electronic compass to analyse the direction and interpret the direction of motion if the line of sight (LOS) with the satellite is not available. Analysing the signal strength of the 2 received signals (and also gathering intelligence from the GPS/compass unit if available), the handset decides to initiate the handover.

In such a case, the passive transceiver as in Figure 3, conveys the time slot and the frequency which is in use by the active transceiver. The base station serving the cell establishes a channel (TS), the same value which was intimated by the passive transceiver. This happens exactly when the active transceiver releases the channel so that the base stations of the adjacent white and the black cells can establish channel on the same TS which is currently in use between the base station of the white cell and the MS.

Hence the physical trajectory of the call is:

passive transceiver of the MS to → base station/access point of black cell → base station/access point of the white cell (which was already in conversation phase directly with the MS before the handover) → central Coordination Processor.

After the handover, the passive transceiver in the MS becomes the primary one. The transceiver which was acting as primary before the handover process now becomes dormant which commences scanning the AFCRN of the adjacent cells for evaluating the signal strength.

The handover never happens between the cells with the same AFCRN band that is with the same colour. The lateral handover is actuated directly between the white cell to black cell or vice versa. However, the diagonal handover happens through a process of 2 lateral handovers [25, 26].

4. Comparative Analysis of the Signaling Processes and Overhead for Mobility Management in the-State-of-the-Art and the Proposed Smart Network

The unique mobility management procedures in SMNAT are not designed specifically for one specific application scenario but for supporting coexistence of M2M, M2H, H2M, and H2H applications. The motivation is to reduce the signaling interaction between the network and the device, actuate a better congestion control, and thereby alleviate the impact due to existence of heterogeneous devices in the mobile network. The following section scrutinises the major differences between the state-of-art and the SMNAT in terms of signaling and mobility management is applicable for all types of User Equipments including the M2M terminals.

4.1. Synopsis of Mobility and Location Management in LTE. The signaling overhead due to mobility management is related to mainly 2 subprocesses, location update and paging. During TA (Tracking area) update procedure [27], the Mobility Management Entity (MME) [10] records the TA in which the User Equipment (UE) is located. When a UE, that is, the Mobile device moves to a new TA, tracking area update is initiated by the device. Paging comes in play when the UE is being called for voice, SMS, network-initiated

location messages (for location services), and network-initiated USSD. In order to place the call to the UE, MME broadcasts paging message in all cells of the UE's registered TA. In LTE, the Mobility Management Entity (MME) is responsible for the mobility management function. The MME is connected to a large number of evolved Node Bs (cells) that are grouped into the Tracking Areas (TAs). The TAs are further grouped into TA Lists (TALs). When a UE moves out of the current TAL, it reports its new location to the MME. If the LTE network attempts to connect to the UE, the MME asks the cells in the TAL to page the UE. In LTE paging, the MME may sequentially page a cell, the TA of the cell, and/or TAL of the cell.

In the following sections, we investigate the 3 main processes of a mobile network related to mobility and location management which imparts signaling overhead in the network. A comparative analysis is made between the state-of-the-art and the SMNAT.

4.1.1. Location Update State-of-the-Art (LTE). Location update in LTE can be categorised as following.

Static Scenarios.

- (i) Always-update: in this scheme, the user updates its location whenever it moves into a new cell.
- (ii) Never-update: in this scheme, the user never updates its location, which means that the location update overhead is zero, but may lead to excessive paging.
- (iii) Reporting cells: in this scheme, the user updates its location only when visiting one of the predefined reporting cells.
- (iv) Forming LA: in this scheme, the user updates its location whenever it changes an LA.

Dynamic Update Schemes.

- (i) Selective LA update: in this scheme, the LAU is not performed every time the user crosses an LA border.
- (ii) Time-based: in this scheme, the user updates its location at constant time intervals.
- (iii) Profile-based: in this scheme, the network maintains a list of "Location Areas" where the user was located in specific time periods.
- (iv) Movement-based: in this scheme, the user updates its location after a given number of boundary crossings to other cells in the network.
- (v) Distance-based: in this scheme, the user updates its location when it has moved away a certain distance from the cell where it has last updated its location. This process is primarily based on reference signal received power (RSRP) calculations by the network.
- (vi) Predictive distance-based: in this scheme, the network determines the probability density function of the user's location based on location and speed reports.

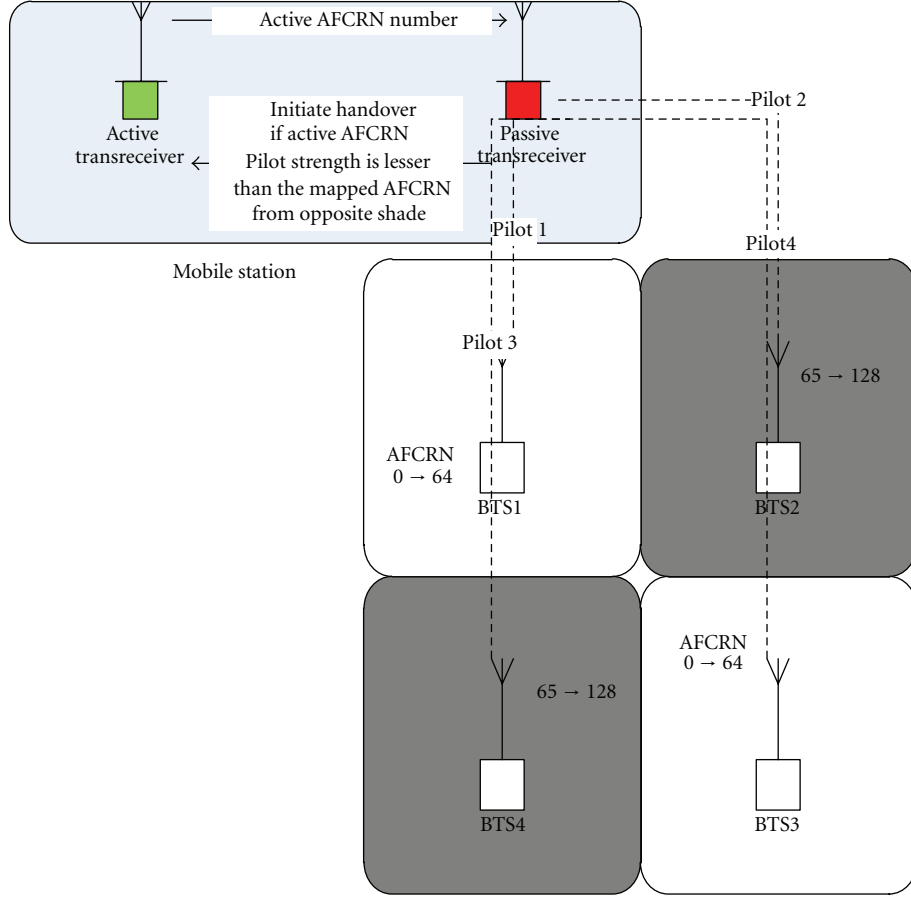


FIGURE 3: Active and passive transceiver and its network interaction.

The MME houses the location update algorithm to cater to any of the above-mentioned scenarios. The signaling overhead for a location update message as seen in the core network is around 378 Bytes. Depending upon the scheme chosen, the total overhead for exchanging the location update message can be determined. Figure 4 shows a location update message showing layer 3 to layer 7.

Before the application layer message for location update is sent, the LTE handset initiates the synchronisation and random access process to acquire a channel. Figure 5 demonstrates the synchronisation and the random access process for the LTE access layer.

The processing of MIB/SIBs reception by the handset (40 ms in case of MIB and 80 ms in case of SIB 1) consumes battery power in the handset.

The bandwidth required for location update can be calculated as follows.

The set of cells in a network is denoted by $N = \{1, \dots, N\}$, and the set of TAs currently in use is denoted by $T = \{1, \dots, T\}$. The vector $\mathbf{t} = [t_1, \dots, t_N]$ is used as a general notation of cell-to-TA assignment, where t_i is the TA of cell i . TA design \mathbf{t} can be alternatively represented by an $N \times N$ symmetric and binary matrix $\mathbf{S}(\mathbf{t})$; in which element $s_{ij}(\mathbf{t})$ represents whether or not two cells are in the same TA,

that is,

$$s_{ij}(\mathbf{t}) = \begin{cases} 1 & \text{if } t_i = t_j, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

h_{ij} is the number of UEs moving from cell i to cell j .

The overhead (B_{up}) of one location update

$$B_{\text{up}} = \sum_{i \in N} \sum_{j \in N: j+1} (c^u h_{ij} (1 - s_{ij}(\mathbf{t}))). \quad (2)$$

4.1.2. Location Update in SMNAT. There is no location update process as there is no network acquisition required in SMNAT. More importantly, there is no location update required prior to the execution of the other network operations/processes, for example while receiving calls/SMS or during handovers. For these operations the necessary precondition is successful paging of the UE at physical layer but the UE does not need to attach to the given network. Thus by avoiding location updates, we save substantial battery power of the UE as well as spectrum bandwidth.

4.1.3. Paging in LTE. In most cases, this paging process happens while UE is in idle mode. This means that UE has

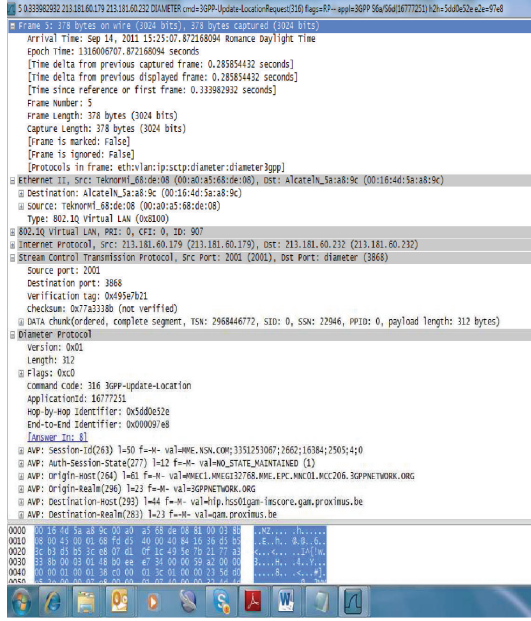


FIGURE 4: Update Location Request on Diameter S6 Interface.

to monitor whether the networking is invoking any paging message to it and it has to spend some energy (battery) to run this “Monitorin” process.

The paging procedure is used by the network to request the establishment of a NAS signaling connection to the UE. The network shall initiate the paging procedure for EPS services using S-TMSI with CN domain indicator set to “PS” when NAS signaling messages are sent to the UE.

Paging in LTE happens via PDCCH channel. The UE listens to the PDDCH channel in given intervals, but not continuously. This is done by DRX (Discontinuous Reception) in order to save power.

The DRX LTE has two timing units as many of other technologies. Timing Unit in Frame scale (SFN: System Frame Number) is one unit and the timing unit in subframe level (Subframe Number). It means that the network needs to know both SFN and Subframe Number to locate exact position in LTE time domain. Regarding the paging cycle, PF (Paging Frame) + PO (Paging Occasion) let you know the exact timing when UE has to wake up to catch the paging message being sent to it:

$$PF = SFN \bmod T = (T \div N) \times (UE.ID \bmod N). \quad (3)$$

According to 3GPP 34.104, T is defined as follows.

T is DRX cycle of the UE. T is determined by the shortest of the UE specific DRX value, if allocated by upper layers, and a default DRX value broadcast in system information. If UE-specific DRX is not configured by upper layers, the default value is applied.

It means UE can get the T from two different sources, one from the system information (SIB2, IE default paging cycle) [28] and the other from upper layer. If upper layer sends the value, it use the T value from the upper layer, otherwise UE has to use the value from SIB2. If T is 5 milliseconds, then

the handset will not be able to receive a page message unless the DRX sleep cycle is completed.

Paging in LTE is done at PDCCH using the S-IMSI (SAE-Temporary IMSI) for addressing which is a 32 bit value.

4.1.4. Paging in SMNAT. The address for the page message is the symbol coordinate at a specific time slot in the radio frame (in the Predefined AFCRN) assigned to the UE. The UE is always in the sleep mode, except the fact that it has to wake up to listen to the specific time slot assigned to itself in the available AFCRNs available within the spectrum. The UE does not need to listen to the entire radio frame. This is not possible in the-state-of-the-art as the paging information can arrive dynamically from any time slot in the frame.

The time slot carries only a single symbol with a repeated transmission of 4 times. If each symbol measures a time span of 1 microsecond, the total time of the radio frame is 10 ms (say), then the UE has to remain live for 4 microseconds. Lower time to be live implies better battery power conservation.

Secondly, no higher layer addressing parameters are used for paging which also conserves the bandwidth. The layer 7 parameters are exchanged only when the paging process is completed and the time slot is seized between the UE and the CP (Coordination Processor) for the conveyance of the signaling and data information. This is the primary advantage of the SMNAT over the-state-of-the-art where it is not possible to confine the paging process to a layer 1 activity.

The time period during which the UE needs to wake up can be calculated as follows (refer Figure 6) Let each frame pertaining to a single AFCRN contain n to $n + k$ time slots.

Say the coordinate symbol (of the complex plane) assigned to the specific UE arrives in K th time slot.

Let there be AFCRNs ranging from i to j .

Hence the time period for which a UE needs to be live within the radio frame T live

$$T \text{ live} = 4 \sum_{i=1}^j t(n), \quad (4)$$

where n can range from 0 to K .

As we have a repetitive code (4 repetitions) for error correction, so the total time is multiplied by 4.

5. A Summary of the Key Issues with the Available Network Topologies for M2M Devices to Interoperate, and How They Are Addressed/Resolved by the SMNAT

5.1. The Radio Network

Issue with Present Generation Networks. it has a complex process of mobility management and radio resource management. Basic philosophy remains the same across all the generations (2G/3G/4G). Complexity exists in radio network design. Each cell in the network area is allocated a range of frequencies. Each cell has its own cell id. Network efficiency depends much on the frequency allocation and the frequency

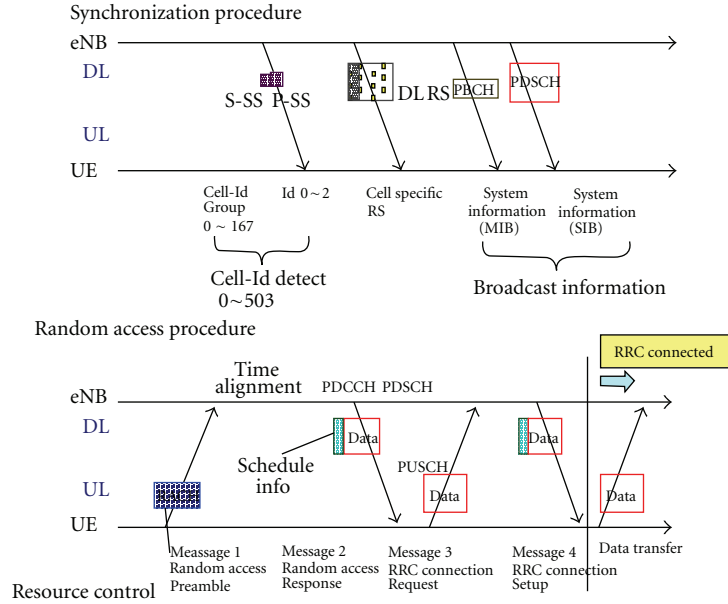


FIGURE 5: Synchronisation and the random access process for the LTE access layer.

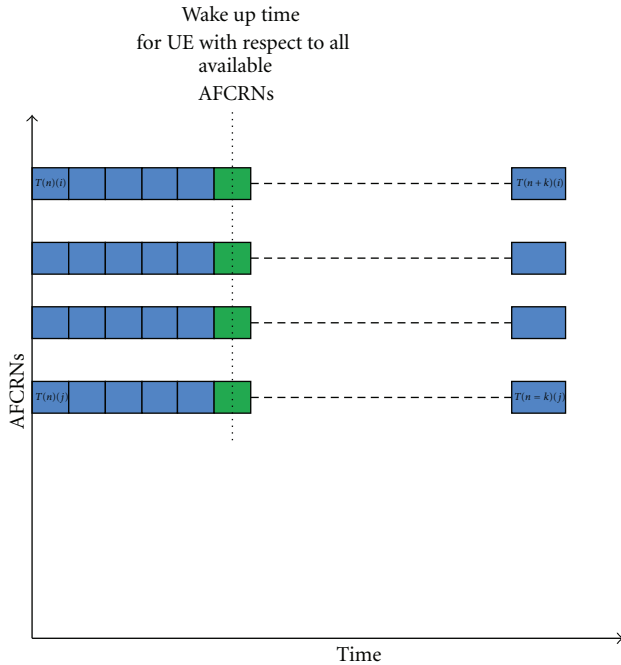


FIGURE 6: Wake up time for the UE.

reuse pattern defined by the operator. Handover of a call when the subscriber is moving from one cell to the other is a process that requires lot of signaling interaction between the handset and the network. As the penetration of the M2M devices will be substantial, hence we will see a sharp increase in the signaling messages in the RAN and CN primarily due to handover, periodic location updates, location updates triggered by change in location. This poses a challenge for the M2M device itself because of the impending needs to power

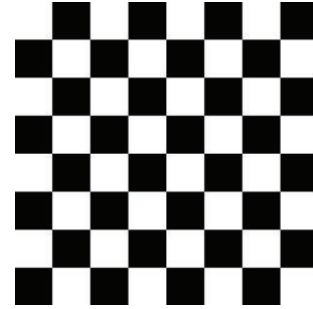


FIGURE 7: Cell pattern and AFCRN block assignment with proposed topology.

conservation. The more the signaling messages transceived, the more the power is consumed. Moreover, the network needs to be reengineered from to augment the capacity of the network elements. Requirements for additional spectrum requirements also need to be addressed.

With the proposed technology: we have a simple checker board design. The available bandwidth is divided in just 2 subbands (against 125 of them as in GSM). The cells are square shaped. Figure 7 defines the cell pattern. Handover is driven by the handset. The signaling interaction required for handover is lesser [26], compared to the hard and soft handoff schemes followed in the-state-of-the-art.

There are no periodic location updates invoked by the mobile station. There are no location updates traversing the RAN and reaching the core network for location changes.

5.2. Addressing. In state-of-the-art, when a mobile station acquires a network, an IP address is assigned. IP V4 addresses are 8 bits, while IPV6 IP addresses are 32 bits. So the

capability to address the mobile devices is limited by the number of devices present in the network.

With the new approach, the addressing is done by physical layer parameters. The subscriber coordinate per time slot is directly used for node addressing. So the constraint to reach the devices is limited by the time slot availability which again depends on the spectrum availability. So the physical limitation due to the IP addressing mechanism to reach a vast expanse of the M2M devices is resolved.

5.3. Dependence on Layer 7 (Application Layer according to OSI Model) Processes. Network tracks the handset continuously. Handset initiates a location update process in a given time periodicity. Network and handset manage the handover to another cell when location is changed. All these network activities are managed by application layer signaling processes at the radio network. So the network equipments and the handset need to be intelligent enough to compute and process continuously the signaling messages for mobility and location management. The network equipments should be able to actuate these processor intensive activities, thus increases the cost of overall network. For the handset as well, it is the same issue as above. Moreover, these processes drain a lot of battery power, just to maintain their "presence" in the network. This makes the mobile device more expensive. When a large number of devices are required (e.g., each device per sensor in a sensor network) the overall cost to set up the application takes an upward trail.

With the proposed technology, we render more intelligence to the lower layers so that it can simplify and actuate some processes (like addressing a mobile node), which are today layer 7 functionalities. We realise a smart modulation scheme which does not just aim to inject more symbols per unit time. It rather (also) attempts to identify the users at layer 1 (physical layer). This in turn enables us to realise the devices cheap, and when we have a swarm of the M2M devices, it makes a substantial difference from the perspective of capital expenditure.

6. Conclusion

The QoS offered by the-state-of-the-art for the Human to Human applications may be grossly impacted once the swarm of M2M devices commence to interoperate in the mobile network framework in full swing. In this paper, we identified the key issues related to this context. We attempted to redesign the mobility and the radio resource management part of the mobile networks to simplify the associated processes and hence reduce the network interaction between the device and the handset. This will help in reducing congestion and boosting the performance of both the M2M and the H2H devices by optimising the battery power, processing power and the bandwidth. As a result, the overall cost to deploy, maintain, and operate the devices and the network will reduce. This is aligned with our endeavour to make the mobile networks cleaner, greener, and leaner.

References

- [1] EXALTED: Expanding LTE for Devices:FP7 Contract Number: 258512. European Commission Information Society and media.
- [2] R. Gunasekaran, V. R. Uthariaraj, R. Rajesh, S. Kaarthikeyan, and S. Aravind, "Priority scheduling in mobile ad hoc networks with improved bandwidth utilization," in *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE '08)*, pp. 1955–1958, May 2008.
- [3] T. Arora, P. Singh, and S. S. Kang, "Performance evaluation and improving bandwidth utilization of AODV protocol by finding hidden terminals in wireless networks," *International Journal of Computer Science and Telecommunications*, vol. 2, no. 6, 2011.
- [4] T. Gowri, K. Raja, and G. T. Arasu, "Improving the bandwidth utilization by recycling the unused bandwidth in IEEE 802. 16 networks," *International Journal of Electronics and Computer Science Engineering*, vol. 1, pp. 623–631, 2012.
- [5] P. Omiyi, H. Haas, and G. Auer, "Analysis of intercellular timeslot allocation in self-organising wireless networks," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, September 2006.
- [6] M.-C. Chang and B.-H. Liao, "Improving bandwidth utilization of NG-SDH by dynamic bandwidth management," in *Proceedings of the 13th Asia-Pacific Network Operations and Management Symposium (APNOMS '11)*, 2011.
- [7] D. G. Jeong and W. S. Jeon, "Time slot allocation in CDMA/TDD systems for mobile multimedia services," *IEEE Communications Letters*, vol. 4, no. 2, pp. 59–61, 2000.
- [8] A. Kumar, Y. Liu, J. Sengupta, and Divya, "Evolution of mobile wireless communication networks: 1G to 4G," *The International Journal on Electronics & Communication Technology*, vol. 1, no. 1, 2010.
- [9] 3GPP 23. 228 Architecture and main flows for a IMS system.
- [10] 3GPP TS 24. 301: Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS), Stage 3.
- [11] V. Simon, *Mobility management algorithms for reducing signaling overhead in next generation mobile networks [Ph.D. thesis]*, Department of Telecommunications Budapest University of Technology and Economics, 2008.
- [12] F. Paint, P. Engelstad, E. Vanem et al., "Mobility aspects in 4G Networks- White Paper".
- [13] C. Castelluccia, "A Hierarchical Mobile IPv6 Proposal. Theme 1—Réseaux et systèmes," Rapport Technique 0226, Projet SIRAC.
- [14] 3GPP TS 36. 104 V8.4.0 (2008-12) Base Station Radio Transmission and Reception.
- [15] 3GPP TS 36. 331 V8.4.0 (2008-12) Radio Resource Control (RRC) Protocol Specification.
- [16] 3GPP TS 36. 211 V8.5.0 (2008-12) Physical Channels and Modulation.
- [17] Network Improvements for Machine Type Communications (NIMTC), Triggered by Rel-8 UID.7027 Study on Facilitating Machine to Machine Communication in GSM and UMTS (FS.M2M) TR 22. 868, created 2008-09-15, last updated 2011-02-28.
- [18] System Improvements to Machine-Type Communications (SIMTC), Enhancements to Rel-10 Feature UID.410031 (NIMTC) based on Rel-11 Study on enhancements to MTC, created 2010-05-28, last updated 2011-02-28.

- [19] G.-Y. Lin, H.-Y. Wei, Y.-S. Chen, and C.-C. Hsu, "Enhanced RACH design for machine-type communications," US Patent application 20120033613, August 2011.
- [20] F. Yu and V. Leung, "Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks," *Computer Networks*, vol. 38, no. 5, pp. 577–589, 2002.
- [21] S. Choi and K. G. Shin, "Predictive and adaptive bandwidth reservation for hand-offs in qos-sensitive cellular networks," in *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '98)*, vol. 28, pp. 155–166, October 1998.
- [22] G. Cao, "Integrating distributed channel allocation and adaptive handoff management for QoS-sensitive cellular networks," *Wireless Networks*, vol. 9, no. 2, pp. 131–142, 2003.
- [23] K.-R. Jung, A. Park, and S. Lee, "Machine-type-communication (mtc) device grouping algorithm for congestion avoidance of mtc oriented lte network," in *Security-Enriched Urban Computing and Smart Grid*, T.-H. Kim, A. Stoica, and R.-S. Chang, Eds., vol. 78 of *Communications in Computer and Information Science*, pp. 167–178, Springer, Berlin, Germany, 2010.
- [24] R. Sanyal, E. Cianca, and R. Prasad, "Rendering intelligence at physical layer for smart addressing and multiple access," in *Proceedings of the 9th International Conference on Networks*, Menuires, France, 2010, <http://www.computer.org/portal/web/csd/doi/10>.
- [25] R. Sanyal, E. Cianca, and R. Prasad, "Beyond LTE: next generation multiple access technology with intelligent lower layers," in *Proceedings of the World Congress Computer Science, Computer Engineering and Applied Computing and International Conference on Wireless Networks (ICWN '11)*, Las Vegas, Nev, USA, 2011, <http://www.world-academy-of-science.org/>.
- [26] S. Rajarshi, E. Cianca, and R. Prasad, "Smart Mobile Networks with intelligent physical layer actuating Mobility Management," *International Journal of Mobile Network Design and Innovation*, vol. 4, no. 2, pp. 91–108, 2012.
- [27] S. M. Razavi, "Tracking area planning in cellular networks-optimization and performance evaluation linkoping studies in science and technology," Licentiate Thesis 1473.
- [28] 3GPP TS 36. 331 V8.5.0 (2009-03) Evolved Universal Terrestrial Radio Access (E-UTRA) Radio Resource Control (RRC) Protocol specification.

