3

SMNAT as Enabler of Device-to-Device Communication

D2D communication for WISDOM based 5G would be a significant aspect, especially as a mechanism to minimize the latency incurred for accessing a network that would be unacceptable for the WISDOM based 5G. Smart Mobile Network Access Topology (SMNAT), an enabler for D2D in WISDOM, has been discussed in this chapter.

The 5th and 6th generation of mobile networks are envisioned to realize the bandwidth required for WWWW (Wireless Worldwide Wide Web). They should be highly dynamic in nature and actuate self-optimization of resources to support the bandwidth for ever increasing customer base. Spectrum crunch is one of the major issues. Hence these networks should thrive to achieve very high spectral efficiency and less Carrier to Interference ratio (C/I), relative to the existing technologies like LTE and WIMAX. Data speed per user should also not be compromised. Ubiquitous, instantaneous and always connected are some of the key prerequisites of a modern mobile device. It has been partly complied by LTE and some Network based applications based on the IMS (IP Multimedia Subsystem) [1] framework like Rich Communication Suite (RCS) [2]. The data requirement per user is facing an exponential growth with the high proliferation of the applications on the UEs available today. M2M and IoTs today play an important role in shaping up the data traffic patterns which can be sporadic or distributed over time. As the new device technologies usher in, the data throughput requirement per user soars. The smart phone and tablets with three dimension (3D) screens and cameras will be bandwidth hungry. Screen resolution will further increase leading to higher data rates. Online multiplayer gaming applications running on 5G gaming consoles with 3D displays will require higher throughput both for the uplink and downlink. Ultra High Definition (UHD) voice and UHD video realized through Rich Communication Services (RCS) [2] will contribute to the growth of data traffic. More dedicated data bearers will be required at the access and core to guarantee the QoS promised to the user. Due to the swarm of M2M and IoT devices in the network, the mobility management related processes will be more intense and will consume more network resources. The nano nature of the devices with limited RF power will imply additional small-cells/microcells which will increase the intricacy of the access and core network.

The key question is how can one realize a network that can achieve the following:

1. Cater the bandwidth requirement of next generation mobile devices meant for human users as well for machines.
2. High Data rate over large coverage areas and dense demographies.
3. Reduce cost of network infrastructure and operation
4. Meet the bandwidth, latency, QoS requirement for supporting the next generation network and device applications.
5. Reduce power consumption in the network and the device
6. Lessen the complexity in the Access (radio) and core network layers

Living through the advancements in mobile network technology from 2G/3G–UMTS towards LTE (4G), one may notice that the primary emphasis is to enable higher bandwidth per user required to drive the next generation network and device applications. With the evolution of the mobile networks from 2G to 4G [3], there is an attempt to inject more symbols (data) per unit time. This has been made possible by the evolution of the modulation schemes from Binary Phase Shift Keying in case of 2G, to 64 Quadrature Amplitude Modulation for LTE-Advanced. In the current Mobile Radio Access Network there is a host of network processes, tightly synchronized and orchestrated by intelligent network elements, such as:

- Handover management
- Location management
- Call drop off management
- Interoperability and downward compatibility management
- Service control (like roaming control)
- Feature management

These processes involve complex signalling operations across the radio and the core network. The attempt to simplify the network in IMS [1] and LTE [4] is focused to make the core and access networks all IP. Following the evolution trail from 2G to 4G, a significant philosophical drift towards simplification in terms of the service logic related to mobility and location management is not discerned. A similar process for handover, frequency reuse, location updates and cancellations in 4G as compared with 2G [3] is witnessed. 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 signalling operations. This has a knock-on effect on the handsets too. It involves considerable amount of processing power, thus leading to more battery drainage. So it is imperative to look beyond the state of the art and conceive a technology which is devoid of these constrains and is ideal for the next generation mobile devices. The same goes for core network design. Mobility and Location management in the core network necessitates a substantial volume of signalling interaction between the HLR/HSS (Home Location Register/Home Subscription Service) and the MME (Mobility Management Entity) in LTE core [4] and MSC/VLR (Mobile Switching Center/Visited Location Register) in UMTS Core. Inefficient and un-optimized engineering and design of these network elements impact the service assurance, and finally degrade to the overall QoS offered by the network. Also, a substantial amount of resources is used up for addressing the wireless node in the cellular network, both from the perspective of available bandwidth and also the associated processing activities carried out in the different network elements at the Radio Access Network (RAN) and Core Network (CN).

Managing a mobile network is hugely resource and knowledge intensive, primarily because of the inherent complexity of the network architectures. For example, during cell planning one needs to consider the traffic demand to cover a specific region, availability of base station sites, available channel capacity at each base station, and the service quality at various potential traffic demand areas. The allocation of the right frequency at the cells to get an optimum frequency reuse factor is crucial to achieve smooth handovers, avoidance of call drops during handover, and overall elevation of the performance of the network. These are part of the overall cell planning activity and it requires a lot of resources, in terms of FTE (Full Time Equivalent) and hence it increases the operational cost. The Switching (Time/Space) matrix of a Mobile Switching Center of a 2G/3G network has a finite limit to make and break the number of calls. The Busy Hour Call Attempt handling capacity of a Switch depends much on engineering and dimensioning of this matrix. With the CSCF (Call Session Control Function) for IMS (IP Multimedia Subsystem) [1] and the MMEs (Mobility Management Entity) for the LTE [4], one does not need the Time and Space switching Matrix. The CSCF/MME acts more as SIP (Session Initiation protocol) router for IMS and SIP/GTP/DIAMETER router in case of LTE. However, for routing a call, these network elements still need to involve directly the Layer 7, which utilizes lot of resources and processing power of these network elements. This in turn makes the network elements expensive and increases the capital and operational cost of the mobile network. As a consequence of the convergence of heterogeneous mobile applications catered by the expanse of the mobile devices preordained for diverse deployment scenarios, the signalling inter-processes between the network and the user equipment become more intricate.

3.1 5G Communication Landscape

With the present day device capabilities, the mobile devices are in constant interaction with each other. The human-centric mobile device interacts with various wearable devices, like smart watches, wearable computers (e.g., Sixth Sense), SOS devices and health equipments. As the user moves around, these devices become more dynamic in nature. The devices, categorized as M2M or IoTs, may be as below (though this is an example and not an exhaustive list):

1. Environment sensors
2. Connected cars
3. Smart objects and robots
4. Health equipments
5. Small cells not owned by the mobile operators

According to the newly formed group called METIS (Mobile and wireless communications Enablers for the Twenty-twenty Information Society), which is a consortium of mobile Original Equipment Manufacturer, operators, academic institutions and automotive companies a 5G Network should have 1000 times higher mobile data rate volume per area, 10 to 100 times more connected devices, 10 to 100 times higher typical user data rate, 10 times longer battery life for low power devices and 5 times smaller end-to-end latency. The main objective of the project is to lay the foundation of 5G, the next generation mobile and wireless communications system. The aim is to let people seamlessly bridge the virtual and physical worlds offering the same level of all-senses, context-based, rich communication experience over fixed and wireless networks. Apart from these basic needs, one need systems which can work with the same performance level in the crowds. It should offer the same QoS and throughput as in office, or home as on the move. It should render low end-to-end latency and reliability to enable machine type applications.

3.2 Related Work on 5G

To meet these demands and to conceive a network which incorporates all these enhancements over the 4G, one may need to analyse the potential of new access technologies and architectural concepts and then identify the ones which can be the eligible contenders to serve as 5G technology

Some of them are as follows:

a) Multi-tier 5G Networks
b) PHY layer based Network Coding
c) Generalized frequency division multiplexing (GFDM)
d) Non Orthogonal Multiple Access (NOMA)
e) Cognitive Radio
f) D2D communication
g) Beam forming technology
h) In mm-wave technology
i) Massive MIMO

However, the scope of this chapter is not to investigate on all these options and not to provide a comparative analysis. Rather, some new paradigms which have the potential to play pivotal roles in shaping up the 5G network architecture and the D2D Communication Framework are staged. In D2D communication, user equipments (UEs) exchange information among themselves peer-to-peer over a direct link using the cellular resources instead of that through the base station or eNodeB. This is markedly different from small cell (femto cell) communication where UEs communicate with the help of small low-power cellular base stations. D2D users communicate directly while remaining controlled under the Cellular Access network. This optimizes resource utilization in a cellular network and boosts spectral efficiency.

This chapter will primarily delve in the following domains.

a) D2D communication: a study on the existing approaches with the new approach, the Smart Mobile Network Access Topology (SMNAT).
b) Integration of SMNAT with LTE-Advanced and 5G Core.
c) Security aspects in light of cooperative communication between two devices.

3.3 Cellular Device-to-Device Communication

Cellular D2D communication is meant to reduce the cellular traffic load by actuating a breakout from the UE itself and directly establish traffic towards the other paired UE using the cellular channel. As a collateral impact, this can significantly contribute to slacken the processor intensive signalling process as explained above. It leverages the benefit from the proximity between two devices and increases the overall resource utilization of the cellular network.

But it is imperative that one needs to come up with new methodologies for device discovery and pairing. Direct D2D technologies have already been developed in several wireless standards, aiming to meet the needs for efficient local data transmission required by variant services in personal, public and industrial areas. Some of the existing contenders are Bluetooth, Zigbee and direct Wi-Fi. With D2D communication, the aim is to find a method which is tightly integrated with the cellular network and uses the same spectrum as cellular operations.

Imparting D2D capability in a mobile device impacts the whole of the network framework and is not an inconsequential addition. Issues like authentication, real time billing, fraud control will crop up and the devices will directly interact with each other bypassing the network.

But on a positive note, the benefits are bountiful.

1. As the paired devices are in the same cell sector, or in the same cell (different sector) or in adjacent cells in physical proximity, high data rates with low latency can be achieved.
2. Depending on the proximity of the devices, the radio power level will be reduced. This will result in better battery life.
3. As the same radio resources are used for cellular and D2D communication, hence the average frequency reuse factor will be better.
4. For a traditional cellular network, one needs two distinct channels for uplink and downlink between the UE and the base station (BS). But in case of D2D communication, a single channel can be used for both directions. Hence the overall spectral efficiency is better.
5. The spectral efficiency can be further enhanced if the cognitive radio communication is used wherein the unused spectrum holes may be utilized for establishing direct communication between the two nodes.
6. As the D2D communications has limited dependence on the network infrastructure the devices could be used for instant communications between a number of devices within a range.
7. D2D on 5G would use licensed spectrum and this would enable the frequencies to be used to be less impacted by interference.
8. In times of natural calamities where some essential components of core and access network have failed, D2D communication can ensue.

A practical implementation scenario of an Over the Top (OTT) application meant from smartphones (for D2D communication) is from Google. A new mobile messaging application called FireChat is empowering nearby smartphone users to stay in touch even when there’s no cellular service or

Internet connection. The messaging application harnesses a technology called wireless mesh networking, which might someday allow a myriad of devices to connect like links in a chain. The technique might someday be used to tie together thousands of devices with built-in radios and make it possible to be online without having to pay for the access. It could also enable online communications in remote areas or disaster zones without Wi-Fi or cellular signals.

D2D has been proposed as a Rel.12 3GPP feature. D2D Study Item had an approval in 3GPP SA1 (Services working group) in 2011, called ProSe (Proximity based Services) [5] which identifies the use cases and envisage the requirements including network operator control, security, Authentication, Authorization and Accounting (AAA), regulatory aspects, public safety requirements, integration with current infrastructure, network offloading. The ongoing discussion by ProSe includes evaluation requirements, D2D channel model, resource use, ProSe discovery and ProSe communication, etc.

ProSe Communication between two UEs in proximity is established by means of a communication path established between the UEs.

- The ProSe Communication path is established
  - by Direct communication between the UEs
  - or routed via the local eNB
  - by ProSe Discovery.
- Communication Process identifies that a UE is in proximity of another UE.
  - by Open [ProSe] Discovery
- ProSe Discovery without permission from the UE being discovered.
  - by Restricted [ProSe] discovery
  - ProSe Discovery that only takes place with permission from the UE being discovered.

D2D communication for M2M type devices is a topic of interest for many telecom researchers. This is because of the huge volume of the devices which may eventually clog up the Mobile network and jeopardize the human-to-human services. The study on D2D requirements for MTC Device to MTC Device scenarios covers

- The identification and functionality needed to set up a connection towards a MTC Device.
- The IMS domain may provide a solution for this required functionality.
- MTC Devices often act as a gateway for a MTC capillary network of other MTC Devices or non-3GPP devices

D2D group for Machine type communications (M2M) Study on Enhancements for MTC is a 3GPP specification – TR 22.888 [6] has been crafted in purview of these requirements.

As depicted in Figure 3.1, the scenarios that are covered are

- Devices Communicating directly
- Devices Communicating via MTC Server
- Devices Communicating with assistance from a Name resolution server

D2D Communication can be broadly categorized as:

1. Out of band
2. In-band

Out of band D2D essentially implies that the devices use a radio technology in the standalone mode to actuate communication with a paired device. The pairing method is in the purview of that specific radio technology.

Some existing out of band D2D communication methods are:

- Bluetooth
- ZigBee
- Near Field Communication
- Direct Wi-Fi

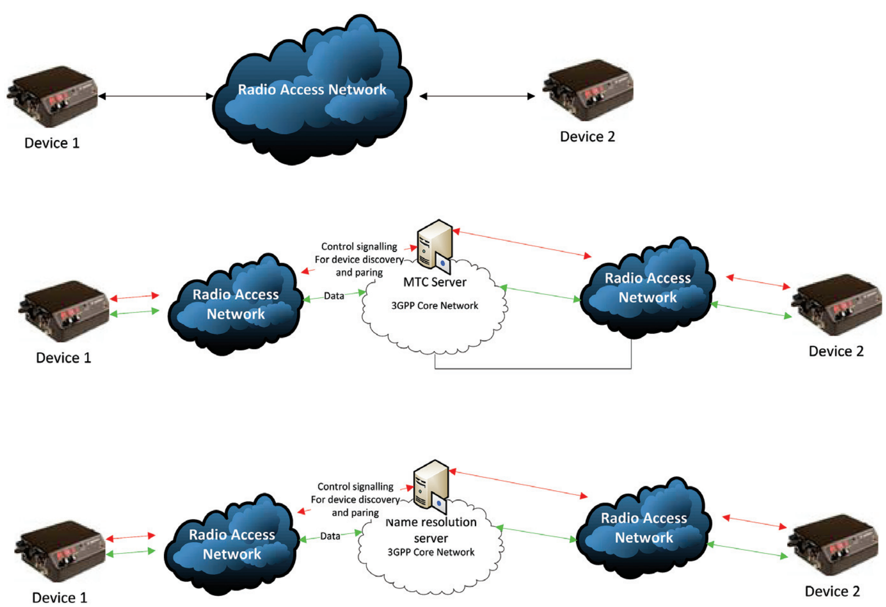

Figure 3.1 Three methods of D2D communication for M2M devices.

In-band D2D communication implies that the devices use the cellular spectrum and the cellular technology.

In this chapter a new concept for the realization of in-band D2D not by using new resource allocation methodologies but by using a new mobility management and addressing concept called SMNAT (Smart Mobile Network Access Topology) [7–13] has been presented.

The subsequent Sections 3.4–3.6 will bring out the various radio resource management technologies that can be employed in the in band D2D communication, as seen in the state of the art. These technologies are based on the underlay cellular network, hence they fall in the category of in-band D2D communication methods.

3.4 D2D Using Physical Layer Network Coding

In this technique [20], there are two cooperating mobile nodes which can relay network codes over the channel codes in two different paths, as shown in Figure 3.2. First it is being sent to the candidate mobile, and subsequently

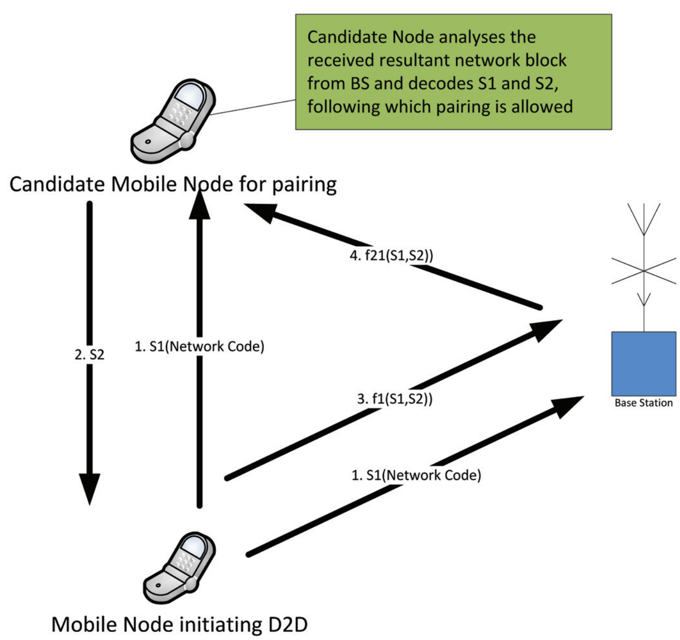

Figure 3.2 Interaction between the initiating device, candidate (pair) device and base station for realizing network assisted D2D communication.

it is being sent to the BS. The candidate mobile node responds back with a new code. The code received from the candidate mobile is combined with the code that was originally sent to that mobile. The combined code is sent to the BS which modifies it with a mathematical expression and relays to the candidate mobile. The candidate mobile who is aware of this mathematical expression deciphers the code that it had sent/received to/from the mobile that has initiated the pairing. When it is successful, the channel is identified and the traffic path between the two nodes is established.

The random selection of two pairs may not alleviate the performance of the network and optimize the system capacity. This is because of the fact that channel conditions between the mobile node and the BS can vary to quite some extent. This may lead to disproportionate SINRs between the nodes and the BS, entirely depending on the channel quality. Hence User Grouping based on proportionate parameters is done by the network based on some Cost Function so as to optimize the performance.

3.5 D2D Using Fractional Frequency Reuse (FFR)

Interference is a significant issue in D2D communications. The interference is caused by devices in D2D mode with that in the normal cellular mode and the eNodeBs. It also depends a lot on the position of the D2D device in the cell, whether it is towards the exterior or towards the interior.

In [21], the radio resource allocation scheme using FFR is proposed. Different resources are allocated to the D2D UEs according to their location in the cell. If the D2D UE resides in the inner region of a cell, then they can use the frequency band that is not used by the eNodeB for relaying to the D2D UEs [24]. The ratio of different frequency reuse factor and the corresponding power level are optimized or adjusted adaptively according to the traffic load and user distribution. D2D and eNodeB relaying UEs located in the same cell do not interfere because radio resources from another frequency band are orthogonally allocated to the D2D and the eNodeBs relaying to the UEs.

3.6 D2D Using Cognitive Radio

In reference [22], there is an overview on how D2D communication can be realized over secondary users. With this concept, the primary users can only transceive via the BS. The secondary user can avail both D2D plus the BS transmission mode.

In reference [23] a joint subcarrier and power allocation method CR-D2D-MC for cognitive multicast with D2D communication coexisting with cellular networks have been proposed. The impact of imperfect spectrum sensing is considered in the proposed problem, which results in the capacity decrease of the cognitive multicast. The simulation results show that the proposed algorithm improves the spectrum efficiency and maintains a better trade-off between capacity and fairness for cognitive networks in a low algorithm complexity. Therefore, employing cognitive multicast based on D2D is able to explore more potential spectrum resources adequately to improve the system performance, and make it possible to satisfy the requirements of multiple kinds of high rate transmission.

3.7 Introducing SMNAT

SMNAT as a concept has been introduced in references [8, 12], which offers a global solution for various categories of cellular network. In SMNAT, the PHY layer is redesigned to take part directly in addressing and mobility management [13]. Similarly, in reference [10], it has been discussed how to implement SMNAT for a mobile network which serves M2M and H2H (Human-to-Human) users. In reference [11], SMNAT has been proposed for Vehicular communication network where D2D communication can be achieved at a group level. In reference [9], implementation scenario for a nano mobile network was brought in. In references [16–18] it has been demonstrated how an intelligent PHY layer can be conceived by using colour synthesis and conveyed electronically for use in mobile or fixed networks. In reference [19] a system model was developed showing how to use the smart PHY layer realized by colour coding for the purpose of joint source channel coding.

Of late, the focus of the telecom researchers is to make the PHY layer more generic, intelligent and adaptable for higher bandwidth. As an example, GFDM [25] implements a generalized Ortho Frequency Division Multiplexing (OFDM), which introduces additional degrees of freedom when choosing the system parameters. A technique called tail biting is employed to eliminate the need for additional guard periods that would be necessary in a conventional system, in order to compensate for filtering tails and prevent overlapping of subsequent symbols.

The purpose of SMNAT is to realize a smart and flexible PHY layer which can directly take part in some activities which are performed by layer 7 in the state-of-the-art networks. Take the example of mobility management itself which entails activities like device location tracking across cells, handover and channel management. All these activities demand some processing power required to actuate associated service logic at layer 7.

SMNAT aims to simulate layer 5 and 7 processes related to addressing and identification of a mobile node (UE) in the RAN, at the PHY layer [8, 13]. A mobile node is identified by a symbol located at a fixed phase and amplitude in the complex plane, a time slot in the time frame and a physical channel which are determined during provisioning of the UE. Figure 3.3 shows one Time Division Multiple Access (TDMA) frame, which includes T_{n+1} time slots (numbered from T_0 to T_n). The data rate required for the traffic burst is denoted by $f_d f_p$ denotes the sampling rates of the time slots before it is used to convey the traffic data. f_r denotes the refresh frequency for the given frame.

$$f_d >> f_p >> f_r$$

The modulation scheme implemented is a blend of M1 PSK (for outer ring) and M2 PSK (for inner ring), where M1<M2. Figure 3.3 shows the proposed constellation in the case of M1=8 and M2=4. The constellation diagram in Figure 3.4 pertains to a specific time slot, T1, of the frame. The frame corresponds to a specific subcarrier. Symbols of the outer ring are conveyed at a rate of f_r and the symbols of the inner ring are exchanged at a much higher rate, f_d.

In particular, one could set: $f_d = 8 \times f_r$. The outer ring comprises M1 symbols for user traffic, and the inner ring comprises M2 that will be used for the purpose of addressing the users. In the proposed multiple access scheme, a user is identified in the network with respect to the symbol coordinate in the

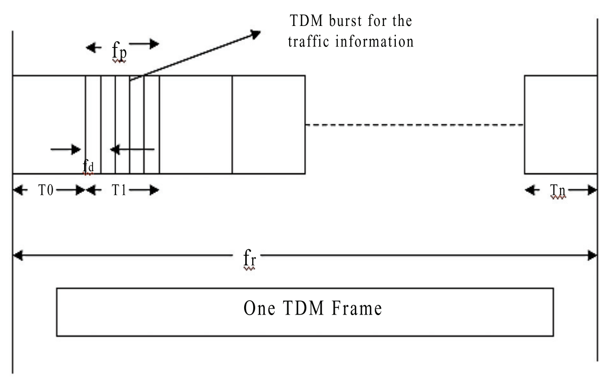

Figure 3.3 TDMA frame.

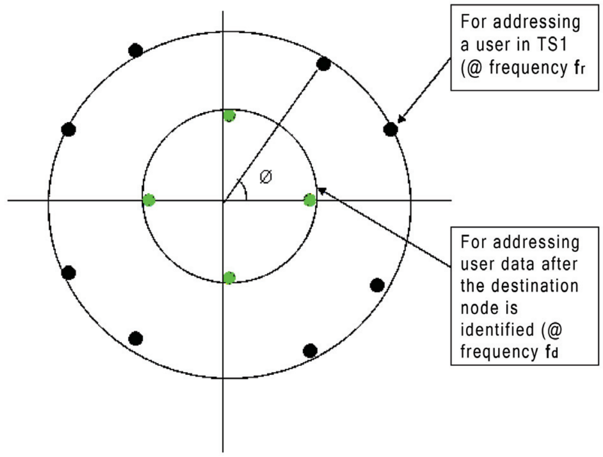

Figure 3.4 Constellation diagram.

complex plane. The 8PSK symbols are continuously rotated with 3/8 radians per symbol. The rotated symbols are defined in Equation 3.1.

$$\hat{s}_i = s_i \, . \, e^{ji3\pi/8} \tag{3.1}$$

The modulated RF carrier is therefore given in Equation 3.2:

$$x(t') = \sqrt{\frac{2E_s}{T}}\,\mathrm{Re}[y(t').e^{j(2\pi f_0 t' + \varphi_0)}] \tag{3.2}$$

Where, E_s is the energy per modulating symbol, f_0 is the centre frequency and ϕ_0 is a random phase and is constant during one burst.

A single time slot can carry M1 symbols, used for addressing M1 users and later, the same time slot will be used to convey the symbols for data traffic employing M2 symbols of the inner ring of the complex plane after the mobile node is identified in the network.

Therefore, M1 users, which have been allocated M1 different symbols of the outer ring could use the same time slot and frequency sub-bands.

3.8 Network Architecture and the Processes

The network consists of access points which are Wireless transceivers/ repeaters supporting the Smart modulation scheme. As shown in Figure 3.5, multiple access points are interfaced to the local multiplexers. For the uplink,

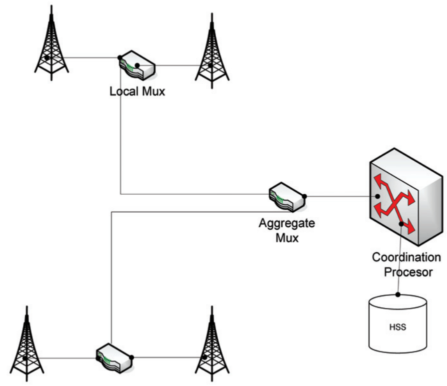

Figure 3.5 The multiplexer unit in the access network.

the local multiplexers aggregate the symbols in the time slots of the time frame pertaining to the different Absolute Radio Frequency Channel Number (AFCRNs). For the downlink, it acts as a transit node between the Coordination Processor and the Mobile Station (MS) to convey the broadcast signal related to all the AFCRNs/time frames to the MSs served by the network. Multiple Local Multiplexers converge to an Aggregate Multiplexer which transceives the time frames on both directions of the Coordination Processor [8] in the uplink and downlink. The high level architecture and the internal functional blocks of the Coordination Processor as in Figure 3.6 are summarized as below.

1. It receives the timeframe in the uplink from the Aggregate Multiplexer and analyses the PHY layer parameters for processing the response message and commence the termination leg of the call or SMS.
2. It formulates the aggregate frame for the downlink communication and relays it to the Aggregate Multiplexer.
3. It interfaces with the HLR/HSS and the VAS network elements to actuate user related tasks.
4. It takes part in the layer 7 signalling to enable authentication, support supplementary services and roaming.
5. It is responsible for CDR (Call Data Record) generation and interacts with the Mediation and Billing Systems.
6. It does dialled number analysis and actuates a TDM/VOIP breakout when needed.
7. It can actuate the functionality of a gateway router to receive calls from a state-of-the-art network. This is alike the functionality of Gateway MSC of 2G/3G networks.

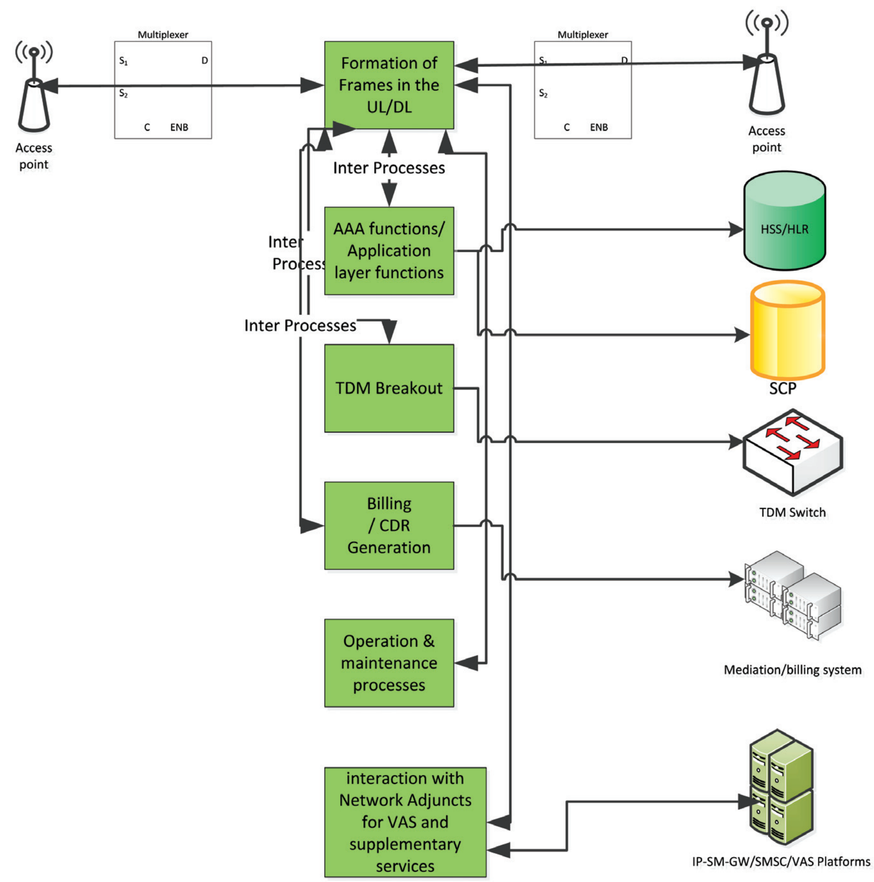

Figure 3.6 High level functionality of Coordination Processor.

3.8.1 Frame Formation in the Uplink

The control signalling messages are invoked by the mobile device only when it initiates a network operation, like Mobile origination Calls, SMS, etc. But no messages are generated for location update unlike the state of the art. The device never gets attached to the network. The network does not keep a track of the location and presence of the device continuously; hence one may eliminate the need of the continuous exchange of mobility management messages between the network and the device.

Figure 3.7 Illustrates how the frame is formed in the uplink for the addressing part. Multiple UEs (mobile devices) as in this diagram invoke

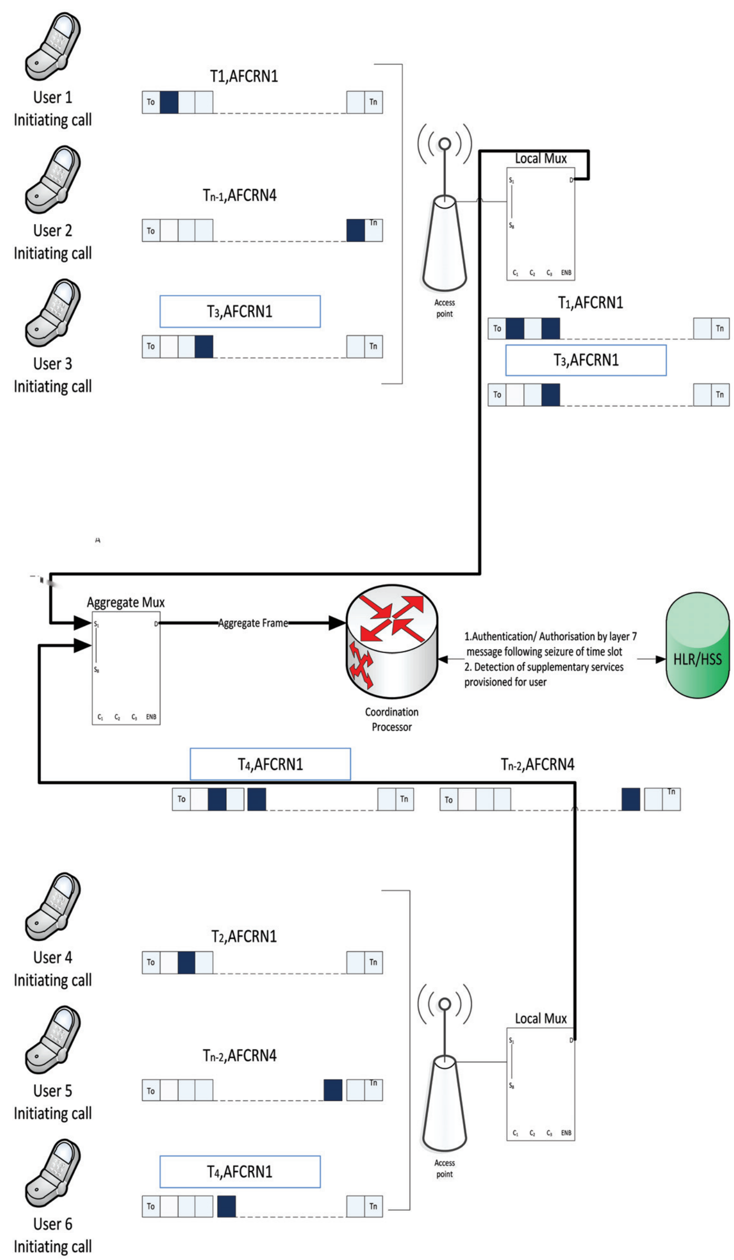

Figure 3.7 Aggregate frame formation in the uplink and user (A number) identification by Coordination Processor.

call requests. The UE first hunts for an available AFCRN where the time slot in the allocated position (as assigned to the user during the provisioning process) is identified. The symbol is then populated in the time slot and the UE generates a time frame by injecting only the specific symbol in that time slot in the time frame. Other time slots in the time frame are empty (as in Figure 3.7). The time frame pertaining to the available AFCRN is marked in black in Figure 3.7 which implies that the user equipment has successfully populated the assigned symbol (assigned to itself) in the time slot predefined for the user. The access point transfers the information to the local Multiplexer which aggregates the symbols ingressing from the user equipments and formulates a single time frame for each AFCRN. The Local Multiplexer formulates a time frame by injecting the symbols from the different users in the appropriate slots as per the time frames arriving from the user equipment. This time frame is transported to the Aggregate Multiplexer. The Aggregate Multiplexer synthesizes an aggregate time frame with respect to each AFCRN and forwards it to the Coordination Processor for analysis and call processing.

At this stage, it is impossible to identify the address of A party address (calling number of a particular user) from the symbol coordinate and timeslot, because the AFCRN is variable. Hence the Coordination Processor cannot commence authentication. It first needs to secure the time slot in both uplink/downlink directions between the MS and itself. Hence it generates a response message in downlink towards the MS using the same AFCRN number that was used by the A party in the during the uplink layer 1 message for call initiation. However, the particular AFCRN may not be available because it may already be in use as traffic channel by other users. In this case, the Coordination Processor does not generate the response message towards the MS. The MS waits for the timer expiry (in milliseconds) and restarts the scan process and re-initiates a new message with another available AFCRN. In this process, it skips the AFCRN that was previously determined as unavailable. Following a successful layer 1 handshake between UE and the Coordination Processor, the time slot pertaining to the AFCRN is used further as traffic channel. The UE generates a layer 7 message towards the Coordination Processor encapsulated in the time slot that was allocated following the layer 1 handshake process as described earlier. The data conveyance materializes following M2 PSK modulation on the inner ring of the constellation diagram (Figure 3.3). UE populates A party addresses (calling address in E.212 format), authentication parameters, dialled numbers, supplementary service information in this message, so that the Coordination

Processor can liaise with the HSS/HLR to actuate authentication, authorization for A party. Subsequently, it analyses the B Party (dialled) number and figures out the outgoing channel by carrying out dialled digit analysis. The above process deviates from mobile originating call procedure followed by the state-of-the-art Mobile Networks. Generally in conventional networks, the serving MSC/MME/CSCF generates an ISUP/SIP message directly for call termination towards the B party address. There is no interrogation towards the HLR//HSS. Only in case of Home routing scenarios, the call is routed towards the Home network where the SCP (Service Control point) is queried to generate a "Connect To" number.

3.8.2 Frame Formation in the Downlink

The downlink process of how the call matures at the downlink towards B Party (called address) UE is shown in Figure 3.8. When the call arrives at the Coordination Processor, it first analyses if the B party address is served by the same mobile network, i.e., the smart network. If yes, then it interrogates the HSS/HLR for obtaining the symbol coordinate, time slot and the primary AFCRN allocated for the B Party. The supplementary and the tele-services assigned for the B party are also downloaded by the HSS/HLR towards the Coordination Processor. Subsequently, the Coordination Processor injects the symbol of B party in time slot (preconfigured for B number) in the time frame for the particular AFCRN in the downlink. In case the primary AFCRN is busy, then alternative AFCRNs are tried according to the predefined frequency scanning schemes. Note that this time frame also carries the symbols from other users in the respective time slots. Aggregate frame is formed and is

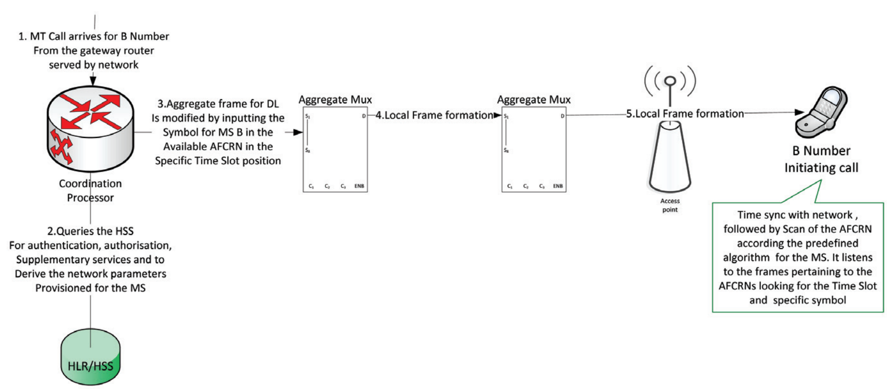

Figure 3.8 Frame formation in the downlink

routed to the Aggregate Multiplexer. The Aggregate Multiplexer broadcasts the time frames pertaining to all the AFCRNs in the Downlink via the Local Multiplexer and access points. B party scans the AFCRNs and looks out for the specific symbol contained in the specific time slot assigned for B number. If it finds one, then it formulates a response message by injecting the same symbol in the same time slot position in the UL. If the time slot (pertaining to the AFCRN) is busy, there is no response back from B party. The Coordination Processor has a time out and tries to page B party again with an alternate AFCRN according to the frequency scanning logic. As AFCRN is not fixed, one cannot ascertain at this stage seeing the radio layer PHY layer message whether the page message is intended for the particular B party address or some other users. The page message can indeed be meant for another user provisioned with the same symbol coordinate and time slot position. Hence, multiple users can simultaneously try to generate the response message. But once the Coordination Processor receives the layer 1 page response, the time slot related to the AFCRN will be reserved for use by the Coordination Processor. The time slot/AFCRN is also reserved at the UE end. The subsequent layer 7 message generated by the Coordination Processor will bear the B party address (E.212) as well as the A party address (E.164 for the CLI) and the supplementary service information. The specific UE (the original B party) will respond to this layer 7 message. The other UEs which reserved the time slot will release it when it ascertains from the layer 7 message carried by the time slot that the page was intended for another user (i.e., the B party). This completes the addressing process in the downlink.

3.8.3 Integration of SMNAT with 5G Cores

It is imperative that the success behind any new technology lies in its interoperability and integration possibility with the existing or evolving networks. As a reference, in [14], the primary challenges for interworking between LTE core network and legacy core network are discussed. SMNAT is oriented on the PHY layer to implement the processes of addressing, mobility management and data exchange. On the contrary, 3GPPs vision on mobility management on 5G is founded on the application layer, more specifically the NAS (Non Access Stratum). The NAS defines the basic processes for mobility management for EPC (Enhanced Packet Core) between the UE and the MME. The two topologies are ideologically converse. Due to this reason it is recommended to follow a two stepped approach for integration of the two technologies.

Short-term approach

WLAN can be integrated by the EPC core as it permits integration of non 3GPP untrusted network via the ePDG (evolved Packet Data Gateway). In Figure 3.9, the end-to-end integration between SMNAT and EPC is shown.

The UE establishes a communication with HSS through ePDG for actuating EPC authentication. At SMNAT end, RADIUS based authentication is followed alike the EAP-AKA mechanism. The conversion rules from RADIUS to DIAMETER (EPC) follow the GSM IR 61 recommendations.

The data transfer between the UE and the IMS core is established via the S2b interface between the ePDG and the PGW.

The call flow is shown in Figure 3.10. It can be seen that two processes are covered.

1. Authentication
2. Data transfer

It may be interesting to note that mobility management messages are not exchanged with EPC. In a state-of-the-art (LTE) network, the UE actuate NAS signalling with the EPC. NAS, which is a layer 7 process, is not implemented in SMNAT. Rather SMNAT is dependent on a process where location management is realized via layer 1. This is the reason, the mobility management messages like Location Update/Location Cancellation are not seen (Figure 3.10). SMNAT can directly fit in the architecture which has been realized for integrating Wi-Fi with EPC.

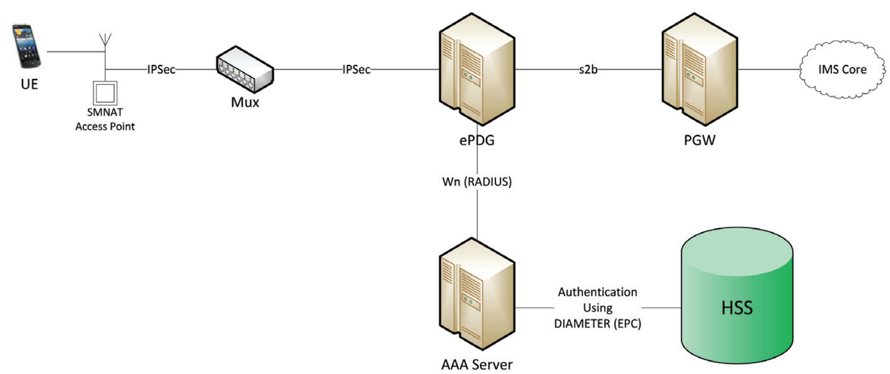

Figure 3.9 Integration of SMNAT with EPC.

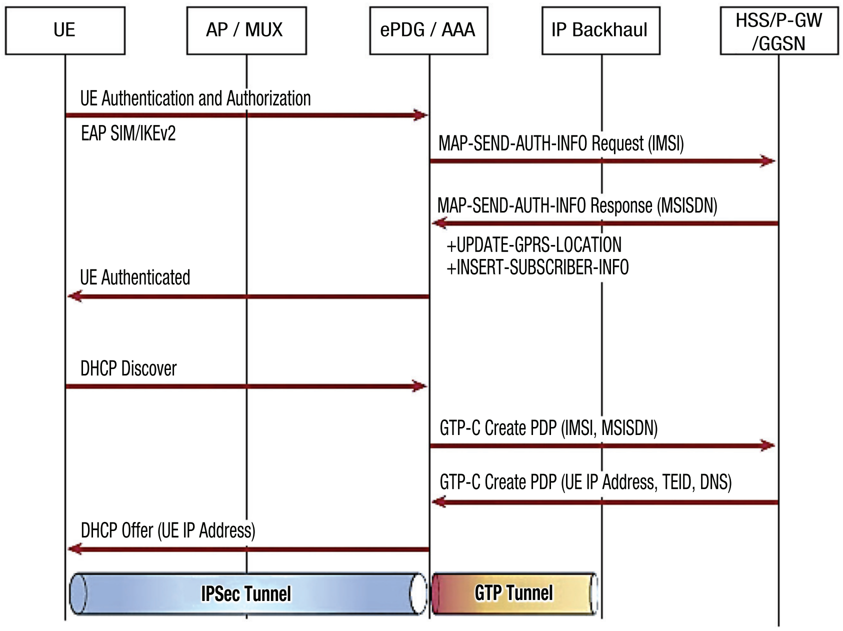

Figure 3.10 Call flow between UE, SMNAT access and EPC core.

Long-term approach

The long-term approach is to evolve the eNodeBs and MMEs as to be compatible with SMNAT architecture. As the mobility management part is simplified, these smart eNodeBs will thrive on PHY layer addressing, rather than the communication on NAS. The MMEs will be closer to the definition done for Coordination Processor, which can directly engage itself in PHY layer addressing and mobility management. Hence eNodeBs/Home eNodeB and MMEs will be lighter in terms of power consumption and processor capacity than the state-of-the-art networks.

3.9 Implementation of SMNAT for In-Band–D2D and Interoperability with WISDOM

This section brings out how SMNAT can be used to realize D2D communication, which can solve some issues and achieve some specific objects which the state-of-the-art D2D technologies cannot attain. Contrary to the work done so far on D2D where the peering is only possible between the two devices in physical proximity, SMNAT offers a more versatile solution where the D2D leg also could be established between two UEs which are not in vicinity. The following section discusses how this can be achieved.

SMNAT can be deployed as an access and core network layer (Figure 3.11), located on the same network plane as the WISDOM 5G [26–28] based on cognitive communication. In such a scenario, SMNAT will be dedicated for D2D communication, while WISDOM will be responsible for global cellular roaming within the network area or beyond.

The available spectrum will be allocated between SMNAT and WISDOM. The ratio of the spectrum allocation can be determined through the process of network planning and can vary according to the traffic characteristics and the business requirement of the particular operator.

To decide whether D2D leg will be established or not, there are two possible options.

1. Decision by the user
2. Decision by the network based on cooperative communication

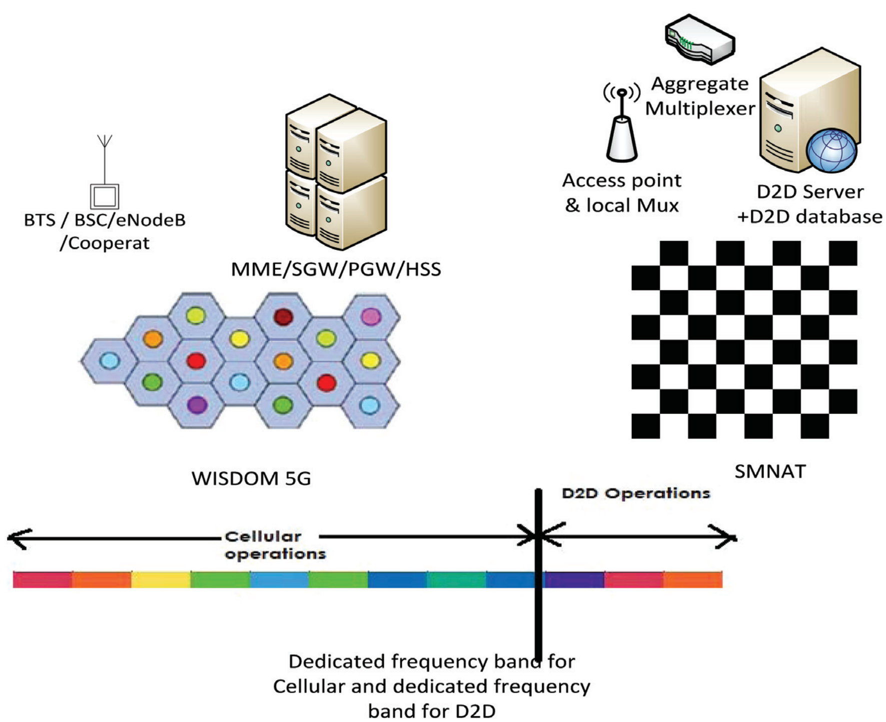

Figure 3.11 Coexistence of WISDOM 5G and SMNAT.

When any device initiates a communication process on data, it will have two options. It will be given a chance to try a D2D communication with the incentive of a faster communication process or lesser data charges (if offered by the network operator).

If the device does not opt for the D2D option, then the session will be established via the normal process of EPC signalling. Subsequently, the EPC can further decide whether to terminate the session to the candidate device via LTE/LTE-A, or the network will send instructions to the device rather to automatically initiate a D2D session.

3.10 Description of network elements of SMNAT and the Call Flow for Session Establishment

SMNAT [12] [8] has a unique design of the cellular network which represents a checkerboard. There are access points in each cell, which via the Frame Aggregator can communicate with the D2D server. The D2D server is

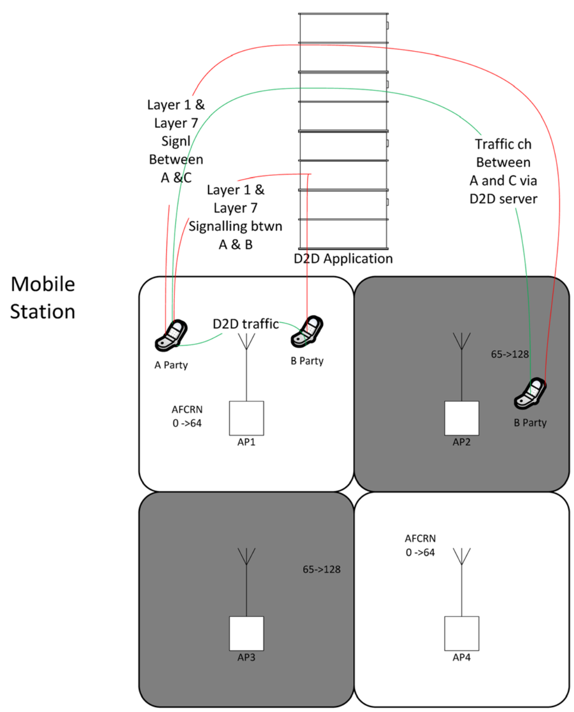

Figure 3.12 D2D Channel establishment based on proximity of the paired device.

interfaced to a D2D database which holds the mapping between the E.164 address of the mobile device and the SMNAT coordinates namely the symbol coordinate, time slot and AFCRN.

As evident from Figure 3.12, D2D transport channel can be established directly between two nodes in case they are in proximity (same or different cell), but facilitated by the SMNAT Access and core network (for signalling). In case the devices are not in the direct signal range of each other, then the traffic channel will be established via the SMNAT Access and core (access point Frame Multiplexer D2D Server). But this process uses the addressing and mobility management methods pertaining to the PHY layer as defined by SMNAT.

In Figure 3.13, the end-to-end call flow for D2D channel establishment is shown, which is self-explanatory. D2D server is a core network application platform which is responsible in establishing the pair, either directly between

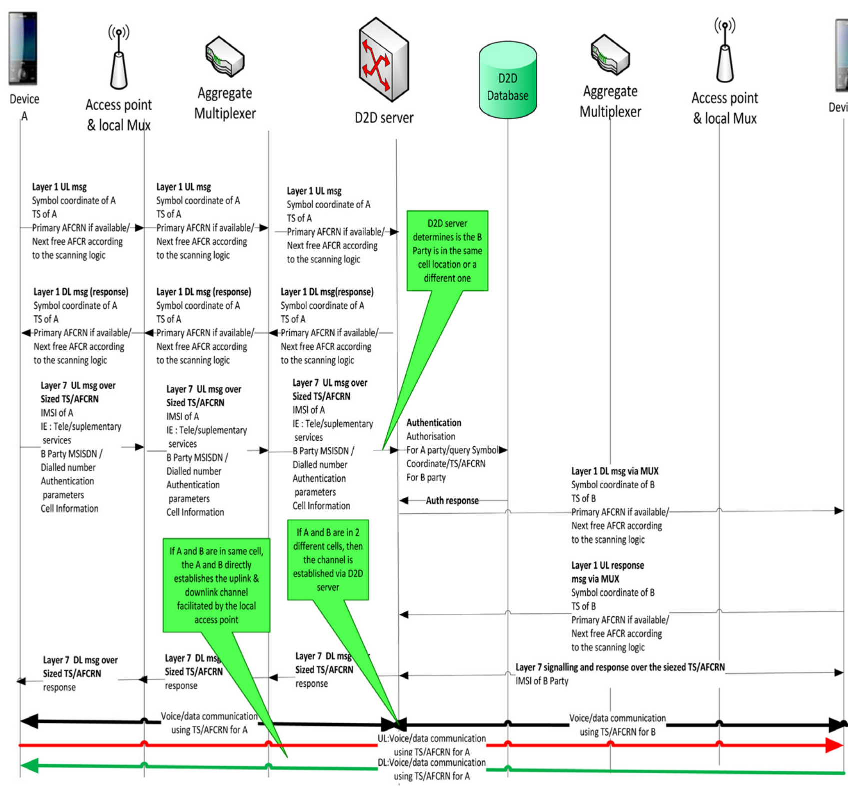

Figure 3.13 End–to-End communication channel establishment for D2D.

the two devices or via the core network using PHY layer addressing and mobility management.

3.11 Decision by the Network to Initiate D2D Based on Cooperative Communications

In case the user attempts (by default) to establish communication through BS via the WISDOM 5G [26–28] network, then the 5G Network will first check if the session can be terminated by the D2D channel. This implies that the network will check if it can offload from cellular to D2D for capacity and optimization reasons. This can be done with the help of Diameter interaction between the HSS and the D2D server. A high level process is shown in Figure 3.14.

3.12 Security Aspects in Light of Cooperative Communication between 2 Devices in Cellular D2D Mode

As explained before, D2D communication can be established in the direct mode or via the network depending upon the proximity of the devices,

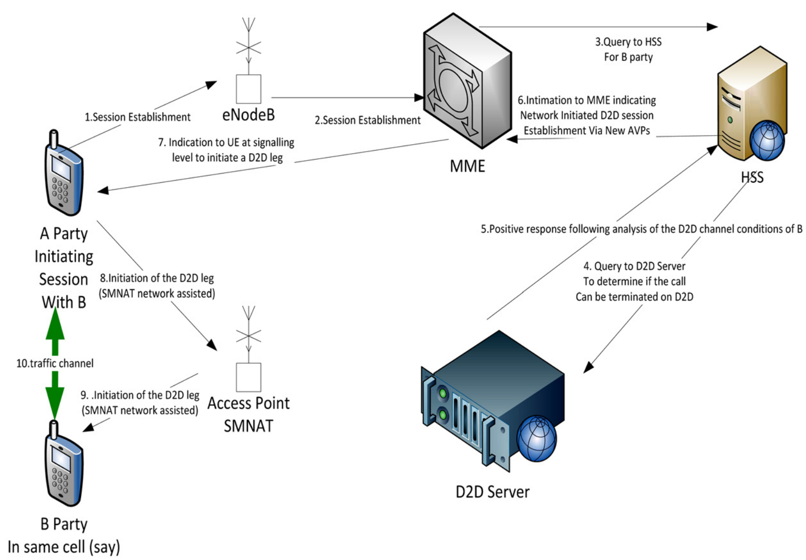

Figure 3.14 Network initiated D2D channel establishment.

channel conditions and location (both devices in the same cell or different cells). Security has become a primary concern especially for the D2D communication.

In case of the network assisted D2D communication, the network can enforce the security methods that are in place for cellular communication. The integration of SMNAT with the EPC has been shown in Figure 3.9, where the conventional EAP AKA method meant for traditional WLAN network can be used. The ePDG (Evolved Packet Data Gateway)/AAA platform modifies the RADIUS based authentication messages to AKA based EPC authentication parameters over DIAMETER directly with the HSS. The EPC authentication vectors including ciphering keys that are downloaded from the HSS will be retranslated and compared for the purpose of authenticating the UE.

However, for the direct mode of D2D communication, where two devices need to pair with each other using the cellular channel, the procedures become intricate. Generation of symmetric keys for the communicating peers without key exchange is challenging. One of the methods to generate the keys is based on random variations of wireless channel like Channel State Information (CSI). Most of these methods usually use the measurement results of individual subcarriers. But this is not robust, given the fact that the subcarriers in proximity can have similar channel conditions and have strong correlation dependence. Due to this fact there can be repeated segments in the key, which may be easy to crack.

Another approach is to use RSS key generation mechanism based on channel measurements. But the main issue is that it is not designed to meet the requirements of a system demanding key generation at massive scale. Bit generation rate is low as a single sample can provide a single RSS value. This also becomes vulnerable if the UE is static and the channel conditions are predictable and the channel variations are predictable. Hence the endeavour of the researchers is to realize a method of key generation based on CSI, but not based on a single subcarrier. In reference [29], a new algorithm termed as KEEP has been proposed. KEEP aims to actuate the key generation process by incorporating the following sub-processes.

1. Dropping the inconsistent bits by exploiting the correlation of CSI measurements from multiple carriers. This is executed by a federated filtration method.
2. Implementation of Universal has key for validating consistency of bit streams between two communicating channels. A pair of devices

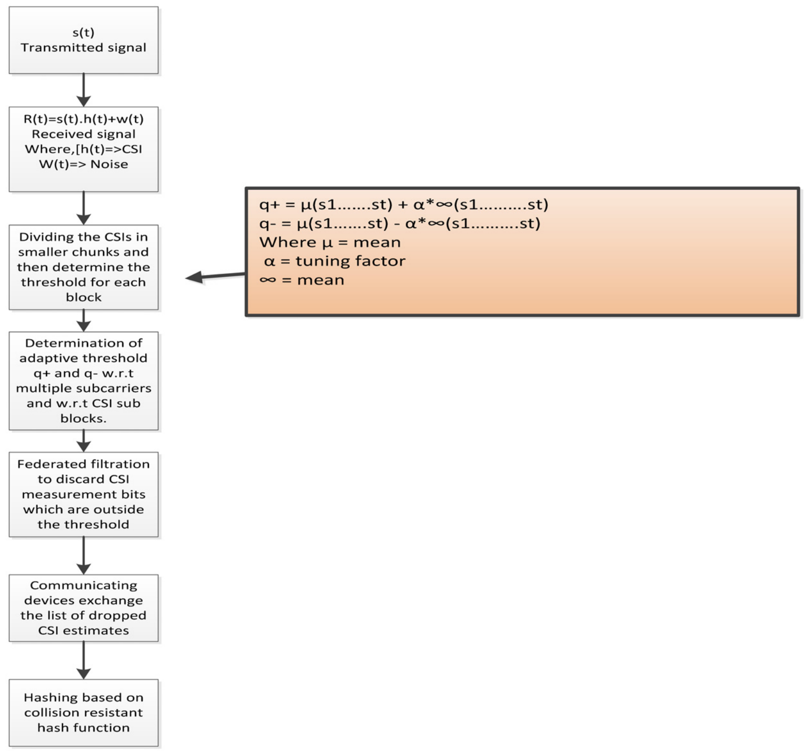

Figure 3.15 Key generation process.

exchanges only a part of the hashed keys, which makes it difficult for the attacker to guess and formulate the whole key.

3. Using methods of key recombination and adaptive quantization to generate secret bit from a large number of subcarriers. The bit mismatch rate with the process is low. Moreover, it eliminates the correlation of bit streams from multiple subcarriers and reduces predictivity.

The key generation process based adaptive quantization procedure, CSI sub-blocking and hashing is captured in the following algorithm in Figure 3.15.

3.13 Summary

At present, mobility management is explicitly in the domain of the application layer. This chapter introduces SMNAT where it is proposed to shift from this paradigm and knit this functionality with the PHY layer. This entails simplification of network architecture and associated processes and helps in attaining a network which has a degree of location and presence agnosticism. The interaction between the device and the network occurs only when one of them initiates a network operation for a call, SMS or other services. The handover operation is also simplified. Further, a novel D2D communication concept based on SMNAT has been proposed. Unlike the existing researches in D2D, engrossed in redesigning the radio resource allocation process, a parallel network stratum has been proposed which can actuate D2D communication based on addressing and mobility management processes closer to the PHY layer. The processing cycles in the network will diminish, which creates extra room to cater to new customers. This is aligned with our endeavour to realize a network which is cleaner, greener and leaner.

References

[1] 3GPP 23.228 Architecture and main flows for a IMS system. Rel 11, version 11.5.0, 2012-06-22.

[2] Rich Communication Suite 5.2 Advanced Communications Services and Client Specification, GSMA.

[3] Amit Kumar; Yunfei Liu; Jyotsna Sengupta; Divya; Evolution of Mobile Wireless Communication Networks:1G to 4G, IJECT Vol. 1, Issue 1, December 2010, pp. 68–72.

[4] 3GPP TS 24.301: Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3. Rel 11, version 11.3.0, 2012-06-27.

[5] 3GPP TR 22.803 V12.2.0 (2013-06), Feasibility Study for Proximity Services (ProSe), (Release 12).

[6] 3GPP TR 22.888 V12.0.0 (2013-03), Study on enhancements for Machine-Type Communications (MTC) (Release 12).

[7] Rajarshi Sanyal and Ramjee Prasad, Enabling Cellular Device-to-Device data exchange on WISDOM 5G by actuating cooperative communication based on SMNAT, paper submitted.

[8] Sanyal, R., Cianca, E. and Prasad, 'Smart mobile networks with intelligent physical layer actuating mobility management', Int. J. Mobile Network Design and Innovation, Inderscience, 2012 Vol.4, No.2, pp. 91–108.

[9] Sanyal, R., Cianca, E. and Prasad, R 'Nano mobile network based on Smart Location Management and Addressing', Int. J. Mobile Network Design and Innovation, Inderscience, Vol. 4, No. 3, 2012.

[10] Sanyal, R., Cianca, E. and Prasad, R, ‘Challenges for Coexistence of Machine-to-Machine and Human-to-Human Applications in Mobile Network: Concept of Smart mobility management, International Journal of Distributed Sensor Networks, Hindawi publishers, Volume 2012 (2012), Article ID 830371, 12 pagesdoi:10.1155/2012/830371.

[11] Sanyal, R., Cianca, E. and Prasad, ‘Closed User Group Automotive communication network based on addressing at physical layer’, International Journal of Interdisciplinary Telecommunications and Networking. IGI Global publishers USA, Volume 4, Issue 4, 2012.

[12] R. Sanyal, E. Cianca, and R. Prasad, “Beyond LTE: next generation multiple access technology with intelligent lower layers,” in Proceedings of ICWN 2011, the World Congress Computer Science, Computer Engineering and Applied Computing and International Conference, 2011, p. 474.

[13] Rajarshi Sanyal, Ernestina Cianca, Ramjee Prasad “Rendering Intelligence at Physical Layer for Smart presented at IARIA conference on wireless networks 2011, France, published at IEEE digital library.

[14] Rajarshi Sanyal, ‘Challenges in Interoperability and Roaming between LTE - Legacy core for mobility management, Routing, Real Time Charging’ ITU WORLD 2011, Geneva, 26th October 2011, published in IEEE Explore.

[15] Sanyal, R., Cianca, E. and Prasad, ‘Novel WLL architecture based on Colour Pixel Multiple Access implemented on a Terrestrial Video Network as the overlay’, International Journal of Interdisciplinary Telecommunications and Networking, IGI Global publishers USA Volume 5, Issue 1, 2013.

[16] Rajarshi Sanyal, Ernestina Cianca, Ramjee Prasad, Colour Pixel Multiple Access- A multiple Access Technology for Next Generation Mobile Networks’, International Conference on Wireless Networks (ICWN 09), 2009 World Congress Computer Science, Computer Engineering and Applied Computing at USA.

[17] Rajarshi Sanyal,‘Framework for Realizing mobile and computer communication through color signals’, SAM 2006, World Congress in Computer Science, Computer Engineering and Applied Computing, USA.

[18] Rajarshi Sanyal, Mobile and Computer Communications Through Colour Signals - An Approach Note. WINSYS 2006: 73–78. Portugal.

[19] Rajarshi Sanyal, Transmission of Audio signals from multiple sources riding on single lossless streaming video channel using Phase Vocoder by

applying Joint Source Channel Coding technique, (CIC07): 2007 World Congress in Computer Science, Computer Engineering and Applied Computing in Las Vegas USA.

[20] Afif OSSEIRAN, Klaus DOPPLER, Cassio RIBEIRO2, Ming XIAO, Mikael SKOGLUND, Jawad MANSSOUR, Advances in Device-to-Device Communications and Network Coding for IMT-Advanced, ICT-MobileSummit 2009 Conference Proceedings.

[21] Hyang Sin Chae, Jaheon Gu, Bum-Gon Choi, and Min Young Chung, Radio Resource Allocation Scheme for Device-to-Device Communication in Cellular Networks Using Fractional Frequency Reuse, 2011 17th Asia-Pacific Conference on Communications (APCC), 2nd–5th October 2011.

[22] Peng Cheng; Dept. of Electron. Eng., Shanghai Jiao Tong Univ., Shanghai, China; Lei Deng; Hui Yu; Youyun Xu Resource allocation for cognitive networks with D2D communication: An evolutionary approach, Wireless Communications and Networking Conference (WCNC), 2012 IEEE.

[23] Yueyun Chen1, Xiangyun Xu1, Qun Lei1, Joint Subcarriers and Power Allocation with Imperfect Spectrum Sensing for Cognitive D2D Wireless Multicast, KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS VOL. 7, NO. 7, Jul. 2013.

[24] Tarun Bansal, Karthikeyan Sundaresan, Sampath Rangarajan, Prasun Sinha, R2D2: Embracing Device-to-Device Communication in Next Generation Cellular Networks, Proc. of IEEE INFOCOM, Toronto, Canada, Apr 2014.

[25] G. Fettweis, M. Krondorf, and S. Bittner, “GFDM – Generalized Frequency Division Multiplexing,” in Proc. IEEE 69th Vehicular Technology Conf. VTC Spring 2009.

[26] Cornelia-Ionela Badoi, Neeli Prasad, Victor Croitoru, Ramjee Prasad, 5G Based on Cognitive Radio, Wireless Pers Commun (2011) 57: 441–464.

[27] Ramjee Prasad, Global ICT Standardisation Forum for India (GISFI) and 5G Standardization, Journal of ICT Standardization, River Publishers, Vol: 1 Issue: 2, June 2013.

[28] Ramjee Prasad and Albena Mihovska, Challenges to 5G standardization. ITU News, Issue N° 10 December 2013.

[29] W. Xi, X. Li, C. Qian, J. Han KEEP: Fast Secret Key Extraction Protocol for D2D Communication, S. Tang, K. Zhao, Proceedings of IEEE/ACM International Symposium on Quality of Service (IWQoS), 2014.