MAC PROTOCOL FOR
M2M COMMUNICATIONS

M.TECH SEMINAR REPORT

by

MEET HARESH HARIA

(153079029)

under the guidance of

PROF. GAURAV KASBEKAR

Department Of Electrical Engineering

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

Contents:

1. Abstract

There is a tremendous demand for the use of autonomous computing, sensing and actuating devices used in multi-faceted applications such as utilities (smart grid), vehicular telematics, home networking, healthcare (M-Health), smart cities etc. M2M communication has become the most important networking paradigm. There needs to be adequate support provided from all the layers in the network stack to fully exploit the applications facilitated by M2M communications and also to meet the service requirements. The report presents various MAC protocols for enhancing and motivating the development of M2M communications in a more efficient, reliable and secure manner. MAC protocols specific to M2M communications are discussed. Also the current Standard development organization's efforts to standardization and future research scope are discussed.

2. Introduction

A rapid expansion is observed by the wireless connectivity beyond the traditional mobile devices used by humans. The future will observe many wireless devices like sensors and actuators connected in the framework of IoT (Internet of Things). In cellular networks, hundreds and thousands of devices can exist in one cell. Therefore the concept of Machine-to-Machine (M2M) has been introduced to handle the transmission of a number of devices the network. M2M can also be referred to as Machine-Type-Communications (MTC) involving mobile nodes over a network with minimal or without human intervention. The goal of M2M communication is to enhance the system automation level so that devices can exchange and share data efficiently. The service facilities delivered by M2M communications covers private, public and professional spaces and scenarios of interest which includes intelligent spaces, smart power grids, smart cities, industry automation and health care to name a few. The increasing M2M traffic and the associated revenue have created an interest among telecom operators as well as regulatory and standardization bodies to facilitate M2M communications.

The fundamental issue is the efficient management of network resources. The network has to consider the traffic characteristics, scalability and cater to QoS requirements. Several M2M devices are battery operated and hence efficient power management is a key concern in M2M devices. Also, M2M devices are “handsoff”, hence must be self-capable in various aspects such as configuration, organization and healing. All these requirements and characteristics affect all the layers in the network stack. This probes a challenging area of research at different levels.

In this report, the MAC layer issues related to M2M communications is considered. The MAC layer is responsible for channel access in a shared medium. The MAC layer difficulty for M2M communications lies in providing channel access to a large number of devices parallel supporting varied service conditions and unique traffic characteristics of devices in M2M networks. Also, MAC protocols should be scalable, power efficiencu must be high, have low latency and implementable using low cost hardware. This report reviews several MAC layer protocols proposed for M2M networks.

MAC Protocol Classification for M2M Communications:

1> Contention-Based Protocols

2> Contention-Free Protocols

3> Hybrid Protocols

3. Requirement of MAC Protocols For M2M Communications

A. Data throughput and efficiency:

The demand for channel resources by vast number of M2M devices simultaneously leads to collisions and a considerable amount of time is wasted. Throughput needs to be very high so as to incorporate vast number of devices. Collisions are critical to understand and needs to be resolved for contention-based systems which pose a negative impact on throughput and efficiency. In contention-free systems, overhead of control data and slots which go empty are important issues that affect throughput and efficiency.

B. Scalabilty:

The node density grows as the deployment of application scenarios with M2M communications become more dominant. Also, the network conditions may be dynamic with nodes entering or leaving the network which in turn necessitates the MAC protocol to be scalable and adjustable to changing node densities with very less or no control information exchange.

C. Latency:

The delays during channel access or network congestion are serious Issues which are important to be considered in M2M networks. It is important in applications like control of vehicles in real-time and applications related to e-health to make communication reliable and fast.

D. Energy Efficiency:

The main factors that make energy efficiency, one of the most important design considerations are: 1) the fact that many devices are operated using battery and are hence power constrained. 2) the economical impact like cost of operation and profit margins. 3) Environmental impact of the power consumed. Thus all M2M communications must be optimized to consume very low power. Collisions during channel access are major cause of power consumption. Also, at high loads, control overhead takes away 50% of the total energy in IEEE 802.11 MAC protocol. Methods to decrease the energy consumption by MAC layer is 1) reducing collisions, 2) Sleep Scheduling, 3) power control, 4) Reduce idle listening.

E. Cost Effectiveness:

The M2M devices must be cost effective so that it is affordable to deploy them. MAC protocol has many desirable properties but depends on the use of complex and costly hardware which is impractical. MAC protocol must be designed to work effectively on simple hardware.

F. Coexistence:

Many M2M devices operate in unlicensed bands due to significant spectrum cost associated with licensed bands. M2M devices need to coexist with devices operating under unlicensed band like WiFi or Bluetooth and licensed band like cellular networks. Problems related to interference in these scenarios and bandwidth sharing are addressed at both physical and MAC layers whereas issues such as collisions due to hidden terminals from the neighboring networks have to be addressed at the MAC layer.

4. General wireless MAC Protocols

A. Contention-Based MAC protocols

The nodes contend for the channel in multiple ways in order to get the channel and transmit the data. Scalability is not arranged in such protocols, particularly due to increasing number of collisions between concurrent transmissions from different nodes.

Aloha and S-Aloha transmit packet as soon as it arrives which increase probability of collisions and decreases throughput. CSMA sense the carrier before transmitting and hence suffer less collisions. Hidden and exposed terminals problems further degrade the throughput. The most widely deployed protocol for Random Access (RA) is CSMA/CA. It performs well for small network sizes but degrades quickly as the number of nodes increase.

These protocols are largely unsuited for M2M communications due to collisions as node density increases. Busy tone based protocols such as Dual Busy Tone Multiple Access (DBTMA) offer better performance at the cost of additional hardware cost and complexity. CSMA/CA are most widely deployed but they are unable to scale as the network size increase and that leaves much to be desired. Other concerns are wastage energy due to collisions, idle listening and the overhead of control packets which consume more energy than data packets.

B. Contention-Free MAC Protocols:

Contention-free protocols eliminate issues of collisions by pre-allocating transmission resources to the nodes of the network. Common contention-free protocols are TDMA, FDMA and CDMA. In TDMA, entire bandwidth is allocated to the user for a fraction of time. In FDMA, a fraction of bandwidth is allocated to the user all the time. CDMA operates by assigning orthogonal codes to each user which are then used to modulate bit patterns. These protocols have fixed number of resources which are time slots, frequency bands, and orthogonal codes. Disadvantages are that they have limited flexibility in presence of dynamic network conditions and are not very efficient at low loads. Hence dynamic resource allocation methodologies needs to be applied.

These protocols are TDMA based. They reallocate the slots based on the number of active nodes and their traffic intensity. Eg. in the Unifying Dynamic Distributed Multichannel TDMA Slot Assignment Protocol (USAP), nodes first observe and then select the slots in the TDMA frame to transmit their data. Considerable control information needs to be exchanged. Frame length and frame cycles change dynamically based on network conditions. Although these improvements enhance the average channel utilization, it is still low and control overhead is significant due to frequent information exchange between nodes.

The main advantage is better channel utilization at high loads. However, the protocols are difficult to adapt as the number of nodes in the network changes. TDMA based protocols have stringent time synchronization requirements which are difficult to implement resulting in extra bandwidth and energy consumption. Also average packet delays is considerably higher. CDMA-based protocols are not suitable for low-cost M2M devices due to their complexity. They require strict power control which increases overall system cost. It also requires computationally expensive operations for encoding and decoding messages. FDMA is a low priority option for operations with low-cost devices because FDMA capable nodes require extra circuitry to communicate and switch between radio channels. The band-pass filters required for this operation are expensive.

C. Hybrid MAC Protocols:

Contention based protocols easily adapt to changing network conditions and are better suited for networks with low loads. On the other hand, contention-free protocols eliminate collisions and have better channel utilization at high loads. Hybrid MAC protocols take the advantages of both these protocols.

These protocols combine CSMA with TDMA, FDMA and CDMA. Protocols that combine CSMA with TDMA, operate as CSMA at low contention levels and switch to TDMA at high contention levels. HyMAC protocols combine CSMA with TDMA and FDMA where nodes are assigned a frequency as well as a time slot to transmit data once the successfully send a bandwidth request using contention based mechanism.

These protocols address some performance issues that arise with contention-based and contention-free protocols. Although these protocols avoid the degraded throughput and collisions of random access protocols at high loads and low channel utilization of scheduled access at low loads, the problem of scalability still exist. Also the overhead associated with reconfiguring the system settings in order to accommodate varying traffic conditions cannot be ignored.

Hybrid protocols have better scalability compared to pure FDMA and CDMA. TDMA based Hybrid protocols are more efficient than FDMA or CDMA based protocols.

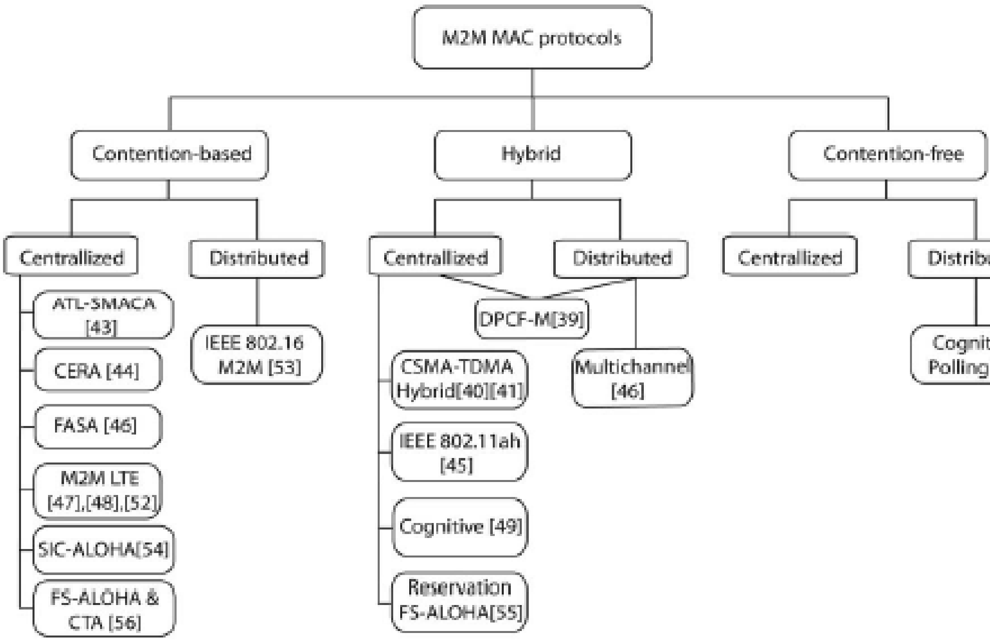


Fig. 1. Taxonomy of M2M MAC protocols.

Ref: A. Rajandekar and B. Sikdar, "A Survey of MAC Layer Issues and Protocols for Machine-to-Machine Communications"

5. MAC Protocols specific to M2M communications

A. *DPCF-M: A MAC Protocol for dense M2M area networks with dynamic gateways.*

DPCF-M (Distributed Point Co-ordination Function-M) is a novel energy-efficient MAC protocol for short range interface of densely populated M2M area networks.

M2M connectivity can be provided by a combination of:

(i) Cellular Networks. (GSM,UMTS, LTE, LTE-A)

(ii) Short-range Radio networks or Capillary networks (IEEE 802.11, IEEE 802.15.4, IEEE 802.15.6).

Low cost end devices are equipped with low-power short range Radio Access Technology (RAT). More powerful devices may be equipped with dual-RATs, one for short range and other for cellular connectivity. Dual RAT devices operate as dynamic and spontaneous M2M gateways, providing single RAT devices with cellular connectivity. The role of M2M gateway must be rotated dynamically among the available dual-RAT devices in order to balance the energy consumption in fair manner.

DPCF-M is a hybrid MAC protocol for M2M communications. DPCF is a combination of DCF (Distributed Co-ordination function) and PCF (Point Co-ordination function) of IEEE 802.11. DPCF-M is an energy efficient derivative of DPCF. The lack of duty cycled operation and the fact that periodic broadcast of poll messages is energy consuming, make DPCF not suitable for energy constrained M2M networks. DPCF-M rotates the role of M2M gateway dynamically among different devices and no means of polling is included in order to reduce overhead and conserve energy. This protocol uses CSMA/CA as a distributed co-ordination function for local communication among the neighboring nodes. When a M2M node needs to contact an external server through the cellular network, it uses one of the gateway capable nodes (dual-RAT) to send the data.

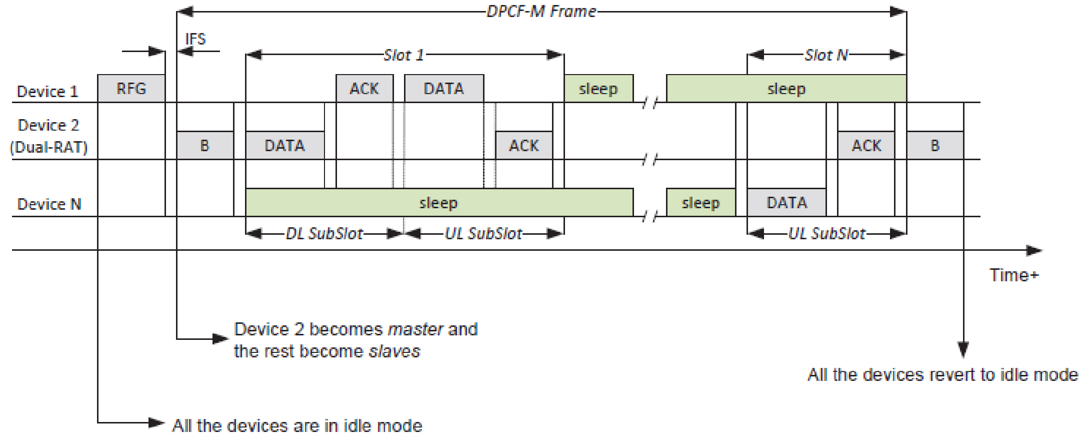


Fig2. Ref: Azquez-Gallego et. al., "DPCF-M: A medium access control protocol for dense machine-to machine area networks with dynamic gateways"

Device 1 is an M2M node and wishes to send data to external data through cellular link. Following events happen:

1) Device 1 obtains access to local channel using CSMA/CA (non-beacon mode of IEEE 802.15.4 MAC protocol)
2) Sends Request for Gateway (RFG) packet to selected gateway.
3) On receiving RFG, device 2 assumes role of master and starts a temporary cluster by periodically transmitting a beacon.
4) Device overhearing beacon enter in to slave mode and suspend their CSMA/CA-based operation and transmit only when permitted by master.
5) Time between successive beacons is divided in to slots which is further divided in to uplink and downlink slot
6) Devices in cluster are asigned individual slots by master during which the devices can transmit if they have the packet else enter in to sleep mode for the entire beacon duration.

M2M Gateway selection can be done ineither of the following ways:

1)At random
2) By using some channel quality indicators.
3) Based on some traffic load of gateways.
4) Based on remaining energy of gateways.

A cluster ends in any of the following two situations:

1) *Network Inactivity*: Inactivity counter is incremented every time an allocated slot goes idle. Counter is reset to zero when a device transmits packets. If the counter overflows, the cluster is ended and the CE (Cluster End) flag is set to 1 in the next scheduled beacon.
2) *Master timeout*: Maximum time a device can operate as master

under heavy traffic conditions in order to avoid the drain-out of batteries. Beacon counter is incremented by 1 unit after the transmission of 1 beacon frame. The master sets CE flag to 1 when the counter reaches the maximum value.

The beacon packets contain following information:

1) Maximum duration of cluster. (Maximum number of beacons).
2) Transmission allocation vector (TAV) indicating which time slot is allocated to each of the devices.
3) A flag (1 bit per slot) indicating whether the time frame structure will allow both U/L and D/L transmissions in each slot.

DPCF-M outperforms CSMA/CA based protocols in terms of throughput and energy efficiency but at the cost of additional hardware for gateway nodes that require two radios. The energy efficiency is obtained by exploiting the duty-cycle between two consecutive beacons. By using the information of TAV, the slaves only switch their radio transceiver on at their allocated time slot and remain in sleep mode in other time slots dedicated for other slaves.

B. A Scalable Hybrid MAC Protocol for Massive M2M Networks

A hybrid MAC protocol for M2M networks which combine the benefit of contention-based (p-persistent CSMA) and reservation-based (TDMA) scheme is proposed. The following is the frame structure:

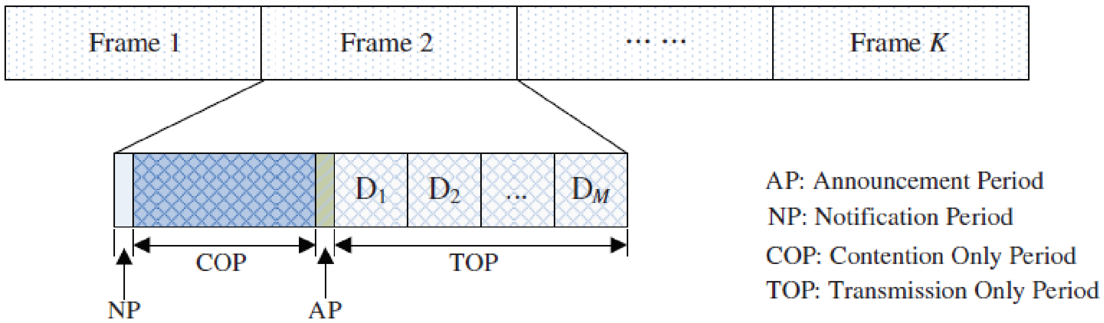


Fig 3. Ref: Y. Liu, et. al. "A Scalable Hybrid MAC Protocol for Massive M2M Networks"

It consists of a single Base-Station dominating over K number of devices. Out of K devices, M devices wants to transmit during one frame. The frame is divided in to 4 slots:

1) NP (Notification Period)
2) COP (Contention Only Period)
3) AP (Announcement Period)
4) TOP (Transmission only period)

The COP is based on p-persistent CSMA access method and is used to contend for transmission slots in TOP. Only successful contending devices are allowed to transmit the data that provides TDMA type of data communication. For a costant Frame duration, if COP duration is increased:

1) Number of successful contending devices increase.
2) TOP duration gets shortened which results in decrease in transmission slots.

Thus, there is a need to optimize the COP duration. Contending probabilty and optimal number of devices to transmit during TOP so as to maximize the throughput.The values of K and L vary from frame to frame and K-L devices can be put in to sleep mode as they have no data to transmit. Let M no. of devices succeed in contention during COP and secure a transmission time slot in TOP.

The BS broadcasts notification messages at NP to all devices to notify the beginning of contention. The L devices will contend for channel during COP. After contention is finished, the BS initiates and broadcasts the announcement message to all of the contending devices which consist of (i) Successful devices' ID and (ii) The transmission schedule. If the device verify its ID in the message, the device should send the data in the assigned transmission slot else the device should go in to the sleep mode and wait for the next frame. Such an arrangement keep the wake up time at minimal and conserve energy. Higher M can be obtained if T_{COP} is longer but T_{TOP} will decrease. Hence there is a tradeoff between TOP and COP.

The expression for the average value of T_{COP} as a function of M(Successfully contended devices) and p(contention probablity) i.e. T_{COP}(M,P) is given by:

$$\mathcal{T}_{COP}(M,p) = \sum_{i=1}^{M} \left\{ \frac{(1-p)^{L-i}}{(L-i)p(1-p)^{L-i-1}} \cdot \delta_{idle} + \left(\frac{1-(1-p)^{L-i}}{(L-i)p(1-p)^{L-i-1}} - 1 \right) \cdot \delta_{coll} + \delta_{succ} \right\}$$

It is also constricted by: $T_{COP}(M,P) + T_{TOP} \leq T_{frame}$. To balance this tradeoff, an optimization problem is formulated to maximize the aggregate throughput obtained by all devices which are allocated transmission slots in each frame. Let T_{tran} and R denote the transmission time slot and data rate of each device which are constant. Then,

$$\begin{aligned} \{M_{opt}, p_{opt}\} = \quad & \max_{M,p} \quad C_{total} = \max_{M,p} \quad MRT_{tran} \\ s.t. \quad & \mathcal{T}_{COP}(M,p) + MT_{tran} \leq T_{frame} \\ & 0 \leq p \leq 1 \end{aligned}$$

It can be calculated that both the equations are convex and the optimization problem is a convex programming problem and can be solved easily with off-the-shelf toolbox. An the optimal period of COP,

$T_{COP,opt} = T_{COP}(M_{opt}, p_{opt})$.

During the notification period BS broadcast notification message to all K devices and estimates K by some estimation method. Based on above optimization solution, BS broadcasts the duration of contention period $T_{COP,opt}$ and the contending probability p_{opt}. If Tran-REQ message (asking for channel) is successfully received by BS , it increments a counter to control the number of successful devices. If the counter exceeds M_{opt} calculated from above optimization, then the contention period ends. If no Tran-REQ message is received for some reason, the other counter of BS that monitors each time slot in COP will eventually exceed a value corresponding to T_{COP} and the contention period will end. Following is the flowchart summarizes the above discussion:

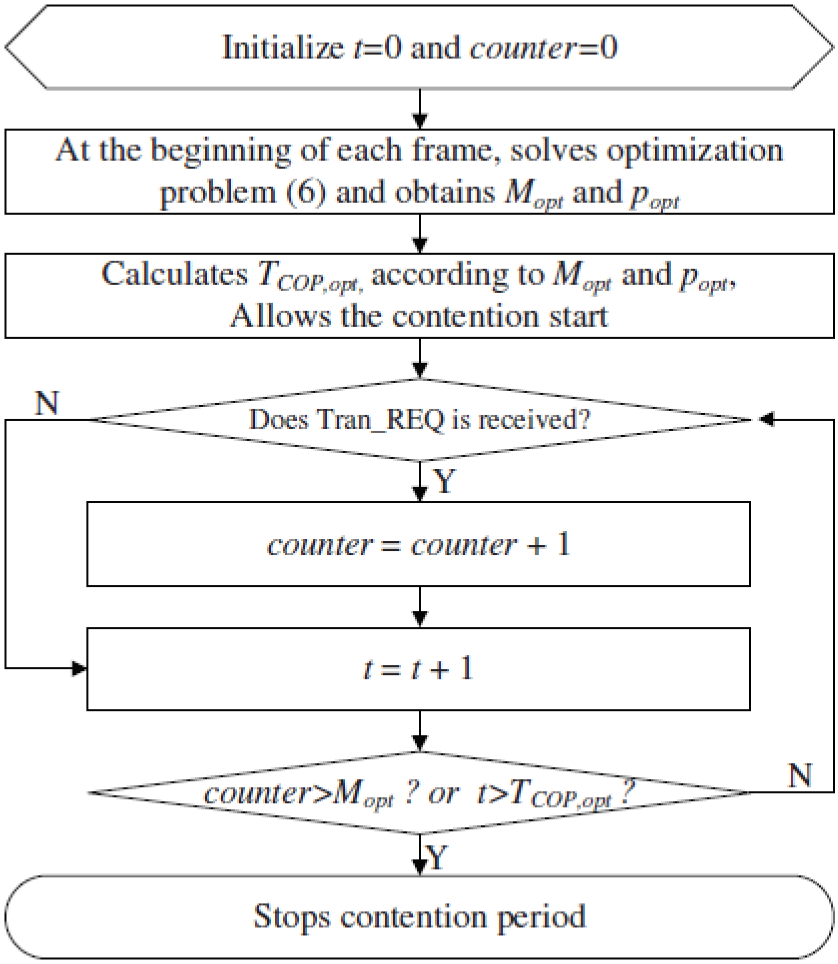


Fig 4. Ref: Y. Liu, et. al. "A Scalable Hybrid MAC Protocol for Massive M2M Networks"

C. An Adaptive Multichannel Protocol for Large Scale M2M networks

A contention-FDMA hybrid MAC protocol based on the use of common control channel is proposed for use in large scale M2M networks. The available bandwidth is split in to number of channels, with one of them used as the control channel.

Multichannel operation can be supported in a centralized or distributed manner. Problems with centralized approach is signaling over-head and single node failure problem. Distributed approach can be divided in to two schemes: Channel hopping based and CCC (Common Control Channel) based. Advantage of hopping based scheme is no signaling overhead but cannot guarantee frequent rendezvous for communicating machines which could lead to longer delays. In CCC-based protocols, one of the channel is used as control channel over which machines negotiate with each other to reserve channels for data transmission. Since negotiation is done in advance, data transmission will be collision free. Therefore CCC-based protocols achieve higher channel utilization than hopping based protocols and are also immune to single node failure problem.

Many CCC based protocols have been proposed. One is the dedicate control channel protocol where each machine must employ two transceivers, one for control and other for data so that they can operate concurrently and channels will be more effectively utilized. Drawback is hardware complexity and more power consumption. To relieve the burden on hardware requirement another protocol is a split-phase multichannel protocol in which time is divided in to periodic intervals which is further divided in to estimation phase, negotiation phase and data transmission phase.

$$T_{total} = T_e\ (estimation) + T_n\ (negotiation) + T_d\ (data\ transmission)$$

1) The estimation phase consists of a number of time slots in which nodes transmit busy tones on the common control channel if they have data to send or if they hear a busy tone from other nodes, with decreasing probability in each time slot. Based on the total number of busy tones sent and heard, a methodology for statistically estimating the number of active nodes(M) is presented.

2) In negotiation phase, all machines switch to the control channel to negotiate the channel reservation. The request message is transmitted at the beginning of each slot with a probability p. After receiving request message, receiver waits for 1 IFS and sends back a reply message. After that rest of the machines wait for another IFS to send their request message with probability p. If collision happens on CCC, all machine wait for 1 IFS to resend their request message with probability p. T_{req}= 18 time slots and T_{rep}=15 time slots and each slot is 20us.

3) In data transmission phase, machines switch to data channels and start their data transmission. Here since concurrent transmission is not possible, all channels other than the control channel will be wasted during the negotiation phase.

If T_n is very short, only few machines can complete negotiation before data transmission phase starts and many channels will be left unused. If T_n is sufficiently long, all machine complete negotiation but T_d will be shorter and hence less time will be left for data transmission. Thus there exist optimal T_n that maximizes channel utilization.

Impact of T_n and p on channel utilization:

Channel utilization is defined as the ratio of channel time used for data transmission and given as:

$$U = \frac{T_d}{T_n + T_d} \times \frac{N_{used}}{N}$$

(Assume $T_e = 0$)

N= no. of channles available and N_{used} = no. of reserved channels for data transmission. U depends on Tn and there exist an optimal Tn that maximizes channel utilization for any given M (number of machines that intend to negotiate). There also exist an optimal value of p for a given value of Tn and M. The dynamic adaptation of p and Tn according to M is impotant in determining the efficiency of a CCC-based multichannel network. M cannot be determined offline and needs to be estimated in real time. Light weight estimation algorithm is used which reiles on "busy tones" for machines to advertise their intention for negotiation.

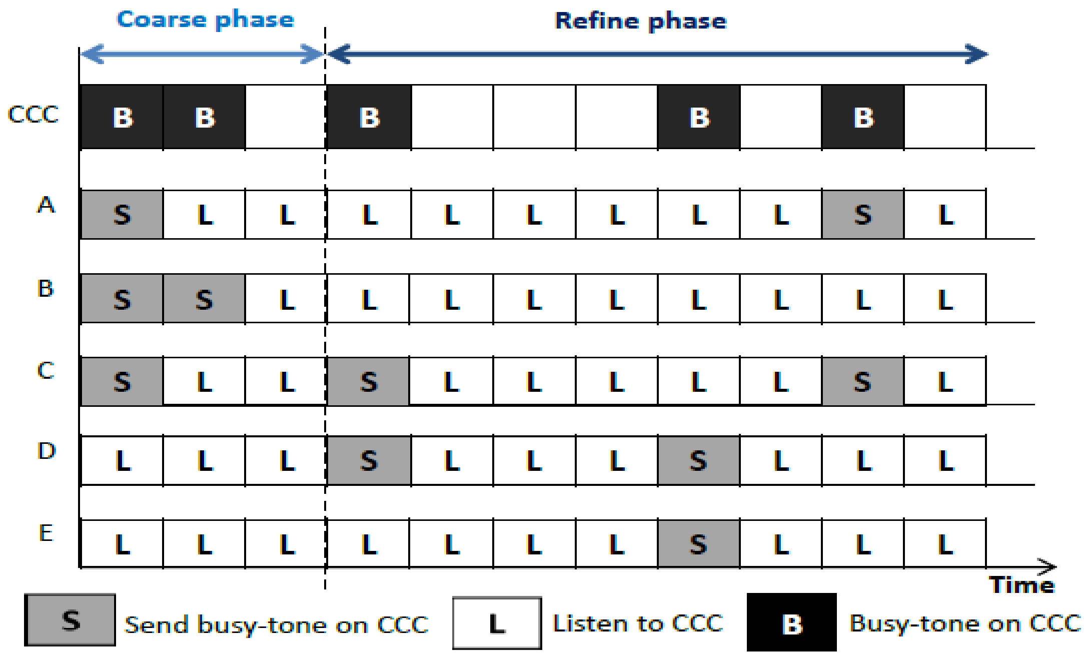


Fig 5. Ref: C. Hsu, et. al. "An Adaptive Multichannel Protocol for Large-Scale Machine-to-Machine (M2M) Networks"

The algorithm is divided in to two phases: coarse phase and refine phase. In the coarse phase, each device transmits a busy-tone on the CCC in the 1st time slot with a probability of p1 = 1/2. It will send a busy-tone in the second slot with a probability of $p_2 = 1/2^2$. This continues with $p_i = 1/2^i$, where i is the index of slots in the negotiation phase. If a device does not have a packet to send then it stops sending busy tone and it listens during the slot and finds if some busy tones are detected. If a busy tone is detected in slot i, the device sends a busy-tone with a probability of $p_{i+1} = 1/2^{i+1}$ in slot i+1. Otherwise, the coarse phase is considered completed for the device. Given that each device reduces the probabilty by 2 in every time slot, the average length of the coarse phase is $\log_2 M$, which is acceptable especially when M is large.

In the refine phase, each device transmits a busy-tone in every time slot. The probability with which the device sends the busy tones is constant and equal is to that probability when last busy tone was sent in coarse phase. The length of the refine phase (L_r) is determined depending on the requirement of accuracy. When refine phase ends, each machine estimates the number of devices that intend to negotiate in the negotiation phase by:

$$\hat{M} = \frac{log(1 - B_r/L_r)}{log(1 - p_b)}$$

where Br = the total number of slots with busy tones on the CCC during the refine phase and p_b = transmission probability of busy tones in each slot.

The optimal p that maximizes the number of machines that complete negotiation for a given Tn is given by:

$$p_{i,opt} = \arg\min_{p_i} \frac{T_{req} - T_{req}(1 - p_i)^i + 1}{i \times p_i(1 - p_i)^{i-1}}$$

The optimal Tn that maximizes the channel utilization is given by:

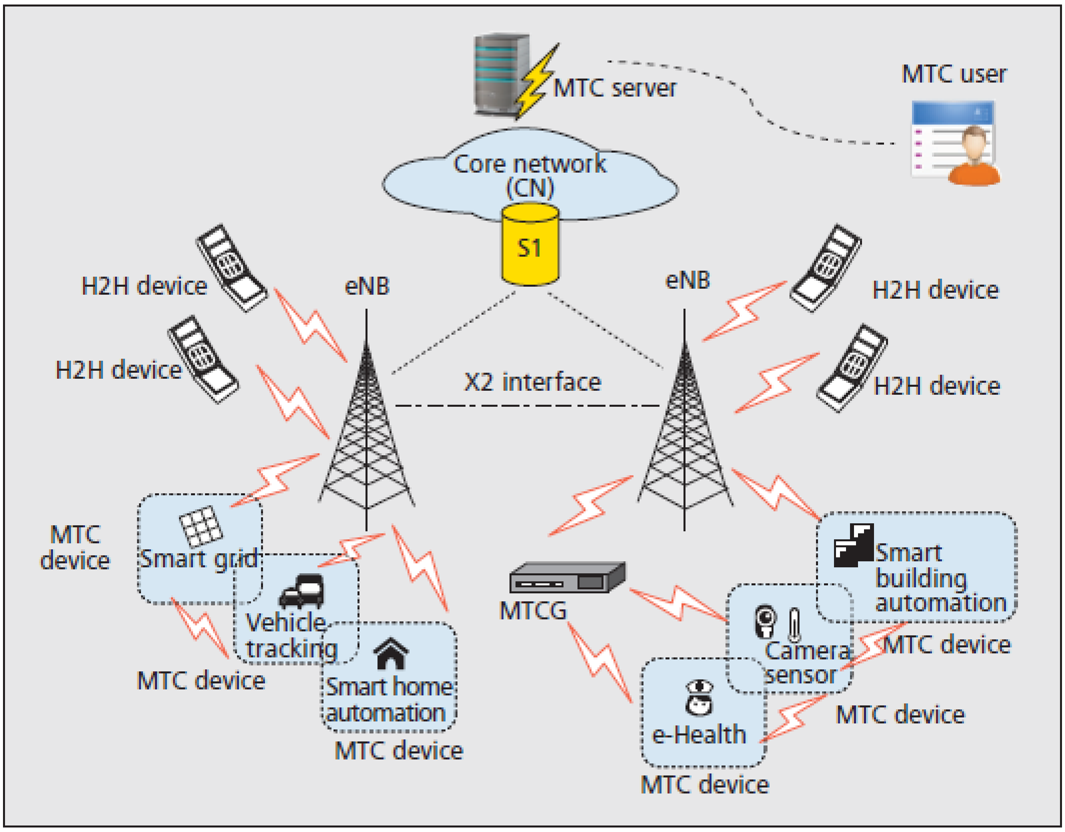
$$\begin{aligned} T_{n,M,opt} &= \arg\max_{T_n} E[U] \\ &= \arg\max_{T_n} \begin{cases} \frac{T_d}{T_n + T_d} \times \frac{E[m_{T_n}]/2}{N}, & \text{if } \frac{E[m_{T_n}]}{2} < N \\ \frac{T_d}{T_n + T_d}, & \text{if } \frac{E[m_{T_n}]}{2} \geq N \end{cases} \end{aligned}$$

where E[mTn] represents the expected value of the number of devices that complete their negotiation in Tn and Nused is replaced by $E[m_{Tn}]/2$. We assume that all m_{Tn} devices fully utilize the data transmission phase. Therefore, in the case of $E[m_{Tn}]/2 \geq N$, at most N pairs of devices is allowed to reserve data channels. $E[m_{Tn}]$ can be computed recursively and finally $T_{n,M,opt}$ can be obtained numerically from above equation.

D. Random Access for M2M communications in LTE-Advanced Networks

Higher layer connection to MTC devices are provided by configuring them with the existing cellular infrastructure [LTE-A]. MTC and H2H devices perform RA using PRACH (Physical Random Access Channel). Though data size is small, when large number of M2M device communicate, it creates a network overload problem. Also base station selection is one of the important issue when M2M device performs an RA in LTE-A networks. MTC devices contend for RBs (Resource Blocks) using RA which uses contention resolution method based on a uniform back-off algorithm. MTC device connects to the infrastructure in either wired or wireless manner. Each have its own advantages and disadvantages but among the available cellular solutions, LTE-A for M2M communication is an appropriate choice due to its long term deployment of infrastructure, scalability and low service cost.

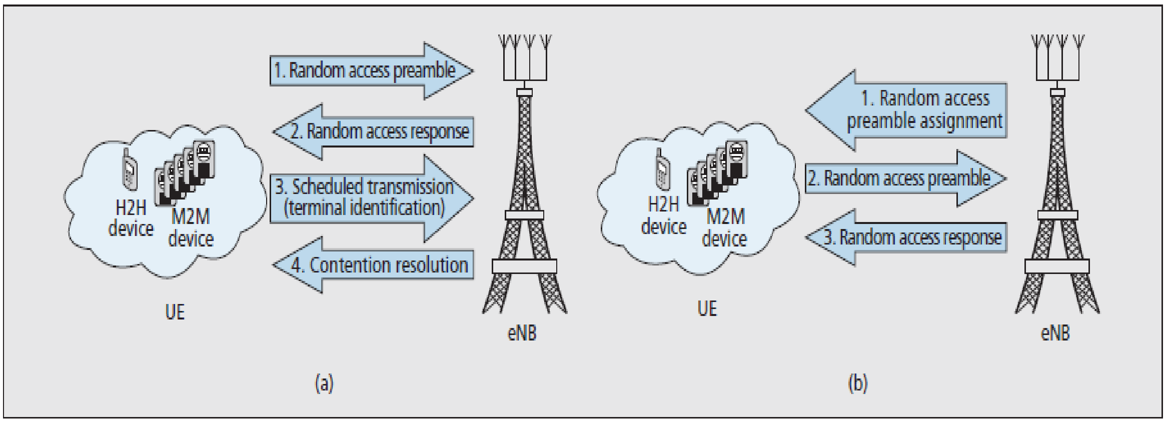
Fig 6. LTE-A Network Architecture for M2M communications:



Ref: M. Hasan, et. al. "Random Access for Machine-to-Machine Communication in LTE-Advanced Networks: Issues and Approaches"

LTE-A comprises of CN (core network) and RAN (Radio access network) or eNB (evolved Node Base-Station). The mobile devices are User Equipments (UEs) which can either be M2M or H2H. Communication between eNBs take place through X2 interface whereas CN and RAN communication takes place through S1 interface. MTC devices are connected to the eNBs either directly or through MTCG (MTC Gateway). MTCG provides suitable path and local control to M2M communication. There can be a direct communication between MTC devices. The eNB-to-MTCG wireless link follow the 3GPP LTE-A specification whereas MTCG-to-M2M and M2M-to-M2M links follow either LTE-A or other wireless standards.

Whenever the device is turned on and has not acquired or lost the uplink timing synchronization or handover from one eNB to another eNB is performed, the RA process is initiated. RA is used for initial access to establish a radio link, for scheduling request if not configured, for resource request when no uplink radio resource has been allocated and for re-establishing a radio link in case of failure. RA procedure can also be contention-based and contention-free.



Random access procedures in LTE-A: a) contention-based RA procedure; b) contention-free RA procedure.

Fig 7. Ref: M. Hasan, et. al. "Random Access for Machine-to-Machine Communication in LTE-Advanced Networks: Issues and Approaches"

A UE device normally initiates RA in a contention-based manner and randomly chooses a preamble which is OFDM-based signal using narrower subcarrier spacing and therefore is not orthogonal to the PUCCH (Physical Uplink Control Channel) and PUSCH(Physical Uplink Shared Channel). Contention-free RA makes eNB to explicitly control devices using dedicated preambles. This approach is faster and can be employed mainly in handover situations where timing delay is crucial.

Contention Based RA Procedure:

1) UE device selects one of the available preambles and transmits it using PRACH (Physical Random access channel).
2) After receiving preamble, eNB transmits an RAR (Random Access

Response) on PDSCH (Physical Downlink Shared Channel). RAR contains useful parameters used for further communication between eNB and UE.

3) After receiving RAR, UE synchronizes its uplink transmission timings and transmits actual RA- message or a unique identity like IMSI using PUSCH.
4) The last step is of the contention resolution using PDSCH. If the eNB decodes any message from step 3, it replies with identifier like IMSI and only that UE device with the same IMSIacknowledges the message and rest of the colliding devices discards the message and try to initiate another RA procedure using random Backoff.

Contention Free RA Procedure:

1) eNBs allocate dedicated preamble signatures to a UE device
2) UE transmits the assigned RA preamble
3) eNB responds with RAR. This is the last step of contention free RA since there is no need to resolve further collisions.

Challenges for RA based M2M communication:

1) *PRACH overload control:* Large MTC devices trying to access the network simultaneously leads to low RA success rate, high network congestion in PRACH, unexpected delays and service interruption. Hence an efficient overload control mechanism must be implemented for RA based M2M communication.
2) *Mode selection and QoS provisioning:* Mode selection means selecting appropriate base station if an M2M devices has choices of selecting so as to enhance the QoS.

Other challenges involve efficient group management and opportunistic RA (Cognitive Radio).

PRACH Overload Control Mechanisms:

There are many solutions proposed for controlling PRACH over load. Access Class Barring (ACB) scheme, PRACH resource separation scheme, slotted access scheme, dynamic allocation of RA resources, Grouping or clustering of MTC devices, MTC specific backoff scheme, Pull-based scheme etc.

When the MTC device wants to transmit a packet, it first performs the ACB (Access Class Barring) check. If the MTC device passes it, then it randomly selects a backoff value from 0 to a maximum backoff counter value and decrements it in every time slot. The moment counter reaches 0, device transmits the packet and waits for acknowledgement from the eNB. If ack is not received, the device assumes that packet is collided and retransmits it using the same above procedure some maximum number of

times. When maximum number of retransmissions is reached, the MTC device discards the packet and proceeds to transmit the next one. Thus MTC device can measure the QoS performance i.e. the packet transmission delay and can choose a proper eNB for its communication.
An efficient method for eNB selection along with an effective overload control method will be required for RA- based M2M communication. Reinforcement Learning Based eNB selection is employed to enhance the QoS by minimizing the packet transmission delay. This will in turn enhance the throughput.

E. A Distributed Multichannel MAC protocol for Multi-hop Cognitive Radio Networks

Surveys show that up to 85% of the licensed spectrum remains unused at a given time and location and hence there is a need for more flexible allocation strategy. Cognitive radio is an intelligent radio that can be programmed dynamically and its transceiver is designed to use the best wireless channels in its vicinity. There is a need to search unused licensed spectrum band and communicate via these white holes. By the means of sensing and adapting, the CR's are able to fill these white holes without causing interference to licensed users. Difficulties that arise while spectrum sensing is that of the hidden terminal problem as it may fail to notice the licensed user and will use the licensed channel and cause interference to it. Spectrum sensing can be greatly enhanced by the use of co-operative partners as will be discussed later.

There are two types of MAC protocols designed for cognitive radio: Single Channel MAC and Multichannel MAC protocol (MMAC). MMAC have advantages over single channel MAC. They offer increased network throughput due to simultaneous transmissions, decreased interference among users and a reduction in the number of CR's. There are two types of MMAC protocols: Single Rendezvous (SRV) and Multiple Rendezvous (MRV). SRV protocols exchange control information using only one channel at a time whereas MRV uses multiple channels in parallel. SRV is further divided in to three classes: Common Control Channel, Common Hopping and Split-Phase approach. For procuring secondary spectrum access, sensing algorithm relies on two phases: low power inaccurate scan and high power accurate scan. Distributed sensing uses multiple CR's to detect a transmission by PU (primary or licensed user) which may be based on OR rule, AND rule or Optimal Fusion rule. Software Defined Radios (SDR's) are key enablers for implementing CR due to inherent flexibility. All radios are have a special interface dedicated for sensing and hence communication and sensing can be done in parallel. These radios have AFE (Analog Front Ends) and DFE (Digital Front Ends) for tackling the Radio-level sensing aspects. Sensing at system level is done using two mechanisms: (i) Fast Sensing (1ms/ch) (ii) Fine Sensing (25ms/ch)

CR's using Common Control Channel have several advantages. CCC reduces the overhead of control information as only one handshake is needed per connection during the beacon interval. Another advantage is the ease of deployment. Also broadcasting can be done with great efficiency and allows distributed sensing. It is assumed that PU's don't interfere with CCC. A dedicated band may be allotted for commercial CR's. CCC can also be configured in unlicensed ISM bands which implies sharing with other unlicensed devices leading to CCC starvation. Another option may be to select a new CCC for each beacon interval in licensed band using the hopping sequence.

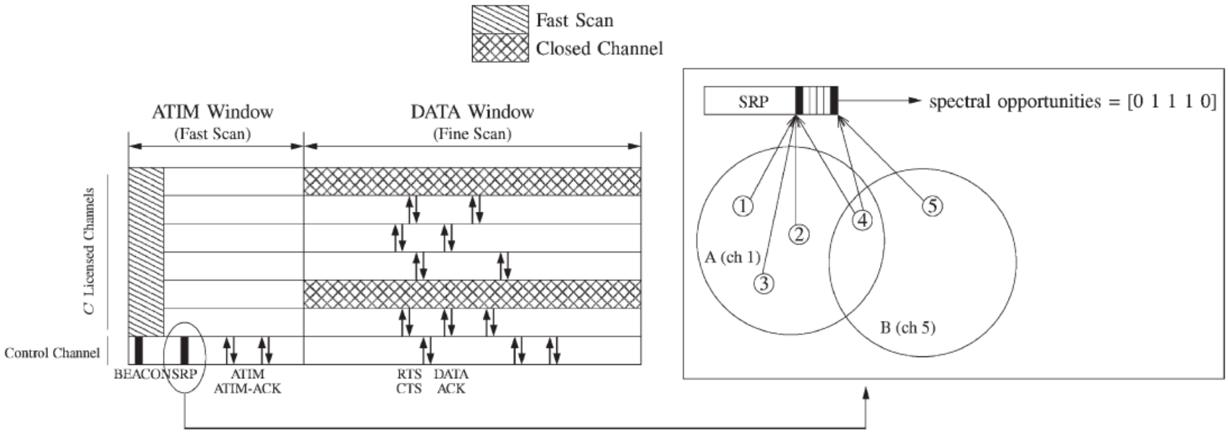


Fig.8. Ref: M. Timmers, et. al. "A Distributed Multichannel MAC Protocol for Multihop Cognitive Radio Networks"

Each CR has two data structures: spectral image of PUs (SIP) vector and Secondary Channel Load (SCL) vector. SIP views the spectrum and has 3 types of entries:

1) No PU channel active on channel c. (SIP[c]=0)
2) A PU channel active on channel c. (SIP[c]=1)
3) Spectral image of channel is uncertain (SIP[c]=2)

A fast scan is performed for every channel in ATIM window (Ad-hoc Traffic Channel Indication Message) and outcome is stored in SIP vector. SIP vector determines the channel availability for data communication and also determines the need of feature scan during data communication when spectral image of the channel is uncertain. SCL vector chooses the communication channel. It contains the expected load which the CR puts on each channel. So when a node wants to transmit data, it picks up the channel having lowest SCL. The global time is divided in to beacon intervals which is further divided in to ATIM window (fast scan) and Data window (fine scan). The ATIM window is used for synchronization among nodes, sensing the licensed channels, learns the network-wide spectral opportunities and performs two way ATIM handshake. Nodes that have exchanged ATIM frames remain awake till they have completed data exchange whereas the nodes that haven't participated in exchange remain in the dose state for the next beacon interval. Data window is used to exchange the data and fine sensing.

Co-operative Spectrum Sensing Optimization

Optimization is needed in the cooperative spectrum sensing to minimize the total error rate. It is necessary to determine the optimum detection threshold to minimize the error rate due to missed detection and false alarms. The main idea behind spectrum sensing is based on binary-hypothesis testing problem.

H0 = Primary User is absent

H1 = Primary User is in operation

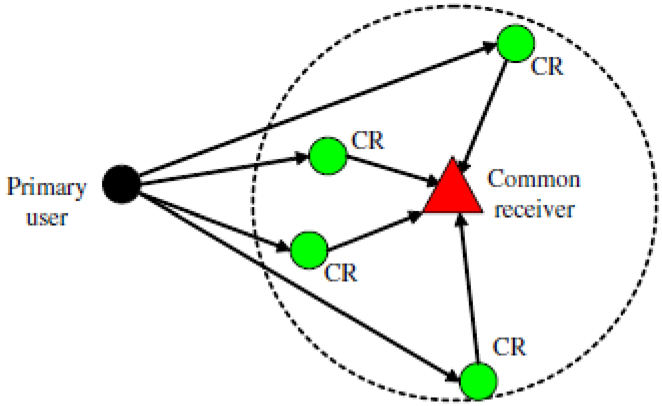


Fig.9. Wei Zhang, et. al. "Cooperative Spectrum Sensing Optimization in Cognitive Radio Networks"

Let there be K cognitive radios and one common receiver (Base Station). Local spectrum sensing is performed by each cognitive radio independently. To observe the working of energy detector only i^{th} cognitive radio is considered.

$$x_i(t) = \begin{cases} n_i(t)\,, & H_0, \\ h_i s(t) + n_i(t)\,, & H_1, \end{cases}$$

$x_i(t)$ = observed signal at the i^{th} cognitive radio, s(t)=PU signal, $n_i(t)$=AWGN

h_i = Complex gain of the sensing channel between PU and the i^{th} cognitive radio.

The energy acquired in the frequency domain is given by:

$$E_i \sim \begin{cases} \chi^2_{2u}\,, & H_0, \\ \chi^2_{2u}(2\gamma_i)\,, & H_1, \end{cases}$$

X^2_{2u} = central chi-square distribution with 2u degrees of freedom
$X^2_{2u}(2\gamma_i)$ = noncentral chi-square distribution with 2u degrees of freedom
$2\gamma_i$ = Non-centrality parameter
γ_i = Instantaneous SNR of the received signal at the i^{th} cognitive radio
u = Time-Bandwidth product.

Average probability of false alarm:

$$P_{f,i} = \text{Prob}\{E_i > \lambda_i | H_0\} = \frac{\Gamma(u, \frac{\lambda_i}{2})}{\Gamma(u)},$$

Average probability of detection:

$$P_{d,i} = \text{Prob}\{E_i > \lambda_i | H_1\} = Q_u\left(\sqrt{2\gamma_i}, \sqrt{\lambda_i}\right),$$

Average probability of missed detection = 1- $P_{d,i}$
λ_i= Energy Threshold

$\Gamma(a,x)$ = Incomplete gamma fumction = $\int_x^\infty t^{a-1} e^{-t} dt$
$Q_u(a,b)$ = Generalized Marcum function given by

$Q_u(a,x) = \frac{1}{a^{u-1}} \int_x^\infty t^u e^{-\frac{t^2+a^2}{2}} I_{u-1}(at) dt.$

$I_{u-1}(.)$ = modified bessel function of the first kind and order = u-1

Co-operative spectrum sensing is performed as follows:

1) After performing local spectrum measurements, CR makes binary decisions D_i ε {0,1} for i = 1,…,k.
2) These binary decisions are forwarded to base station(cellular n/w) or access point (WLAN).
3) The common receiver combines those binary decisions and makes a final decision H0(absence of PU) or H1(Precense of PU).

$$Z = \sum_{i=1}^{K} D_i \begin{cases} \geq n, & \mathcal{H}_1, \\ < n, & \mathcal{H}_0, \end{cases}$$

When atleast n out of K cognitive radios infer H1, BS infers that PU signal is being transmitted else H0. If all the cognitive radios have the same threshold λ, P_{fi} being independent of I can be now denoted as only P_f. Similarly, in case of AWGN channel, $P_{d,I}$ is independent of I whereas in case of Rayleigh fading channel, $P_{d,I}$ can be averaged over γ_i to get P_d. So for both kind of channels, P_m = 1- P_d.

The false alarm probabilty of cooperative spectrum sensing can thus given by:

$$
\begin{aligned}
Q_f &= \text{Prob}\{\mathcal{H}_1|H_0\} \\
&= \sum_{l=n}^{K} \binom{K}{l} [\text{Prob}\{H_1|H_0\}]^l [\text{Prob}\{H_0|H_0\}]^{K-l} \\
&= \sum_{l=n}^{K} \binom{K}{l} P_f^l (1-P_f)^{K-l},
\end{aligned}
$$

The missed detection probability of the cooperative spectrum sensing is given by:

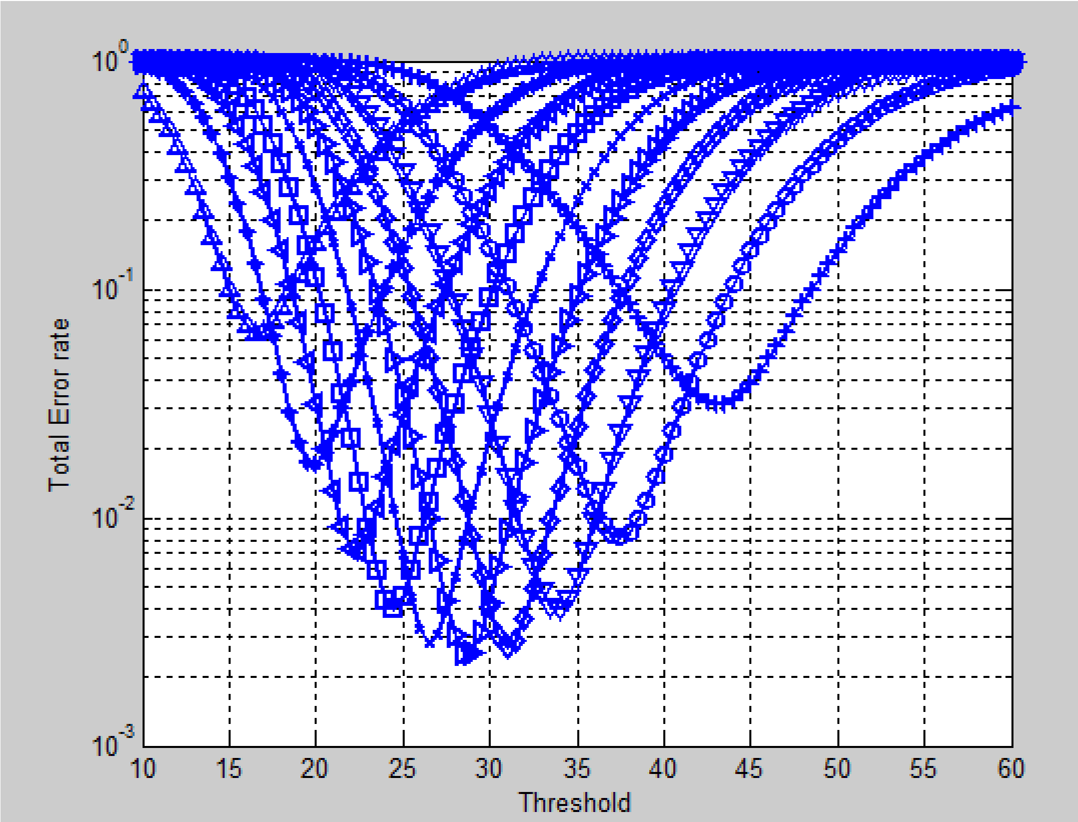
$$
\begin{aligned}
Q_m &= \text{Prob}\{\mathcal{H}_0|H_1\} \\
&= 1 - \text{Prob}\{\mathcal{H}_1|H_1\} \\
&= 1 - \sum_{l=n}^{K} \binom{K}{l} [\text{Prob}\{H_1|H_1\}]^l [\text{Prob}\{H_0|H_1\}]^{K-l} \\
&= 1 - \sum_{l=n}^{K} \binom{K}{l} P_d^l (1-P_d)^{K-l}.
\end{aligned}
$$

If K and SNR are known, then the optimal fusion rule i.e. the optimal n (n_{opt}) which minimizes the total error rate ($Q_f + Q_m$) is given by:

$$
n_{opt} = \min\left(K, \left\lceil \frac{K}{1+\alpha} \right\rceil\right),
$$

where $\alpha = \frac{\ln \frac{P_f}{1-P_m}}{\ln \frac{P_m}{1-P_f}}$ and $\lceil \cdot \rceil$ denotes the ceiling function.

The above deduction is verified by the simulation done in MATLAB and following is the graph of Total Error Rate v/s Threshold for multiple values of n ranging from 1 to 10.



It can be seen from the above graph that the optimal fusion rule over all the examined range of detection is n=5. Thus for a large cognitive radio network, this fast spectrum sensing algorithm requires only a few, not all, cognitive radios in cooperative spectrum sensing to get a target error bound.

The other MAC protocols specific to M2M communications are briefly described as follows:

F) ***Adaptive load slotted MACA(Multiple access with collision avoidance):***

If M2M nodes are not capable of sensing the carrier, an extension of slotted MACA protocol called the Adaptive Traffic Load Slotted MACA (ALT S-MACA) protocol is proposed. This protocol slightly modifies the RTS-CTS-DATA-ACK-based scheme of MACA. The motivation behind ALT S-MACA is that the slotted MACA reaches its maximum throughput at some G_{opt} and then decreases rapidly. The BS estimates traffic load G and then assigns probability of G_{opt} / G to every node for RTS contention. Thus offered traffic load is kept constant at G_{opt} and achieves maximum throughput.

G] *Code Expanded Random Access (CERA):*

This protocol modifies the dynamic random access channel (RACH) resource allocation used in LTE. The main purpose is to provide support to large number of devices without increasing resource requirements. In LTE, RA is performed by selecting one of the available preambles and then sending it over a sub-frame that is selected randomly in a virtual frame. When two or more nodes select the same preamble and same slot, collision occurs.

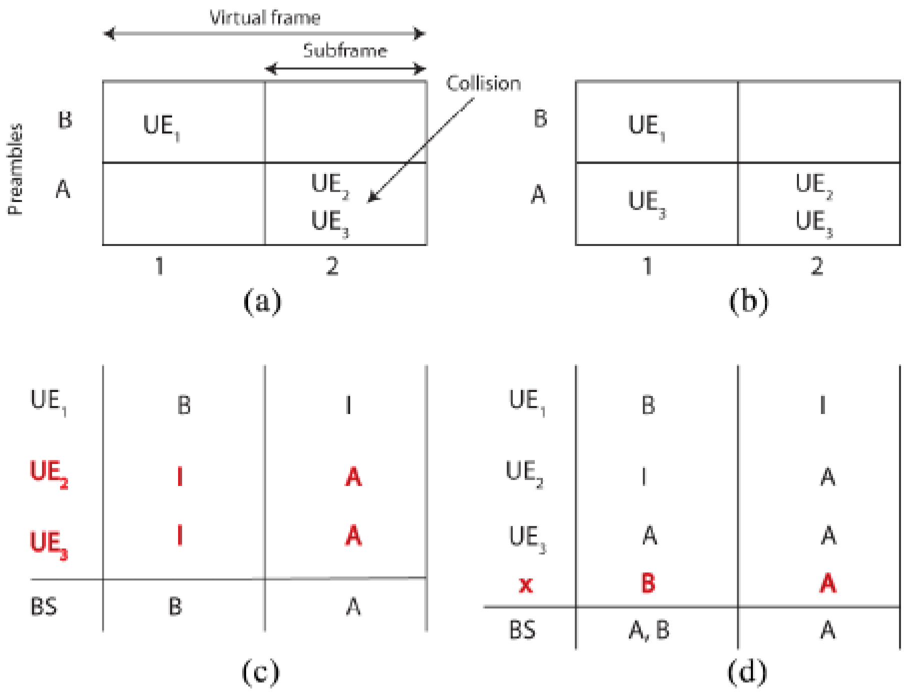


Fig10. Ref: N. K. Pratas et. al."Code expanded random access for machine-type communications"

In the modified protocol, the number of sub-frames is fixed in a virtual frame. Nodes can send the same preamble in multiple sub-frames. The sequence of preambles transmitted by a node in the virtual frame constitutes its code-words and BS identifies a node based on its code-word. This increases contention resources and reduces possibility of collision.
In above figure the code-words for UE1, UE2 and UE3 are BI, IA and AA respectively. This kind of code transmission leads to phantom code words like BA which are received by Base Station and it will incorrectly add nodes corresponding to that base station. Though when the traffic load is high, the possibility of transmitting such phantom code words goes down.

H] ***Enhancement of IEEE 802.11ah for M2M communications:***

802.11ah has the ability to offload cellular traffic as well as support M2M communication. It uses beacons to divide time in to frames which is further divided in to: 1) restricted access window (RAW) 2) Offload traffic (Common Window). Every RAW is divided in to slots and a slot may be randomly selected by the device or it may be allocated to an M2M device by an AP . Binary exponential back-off algorithm is used to poll for the channel. RAW length needs to be optimized for efficient channel access. RAW is divided in to RAW U/L and RAW D/L and the interesting part is U/L. It is necessary to determine RAW U/L size which can be initiated by estimating the number of devices wanting to transmit. This estimate is fulfilled by AP by utilizing the probability of successful transmission in the last frame. The estimated slot in RAW U/L is a linear function of the number of active nodes. This protocol increases the probability of successful transmission.

6. Standards for M2M communications:

Ref: G. Wu, et. al. "M2M: From Mobile to Embedded Internet"

1] *3GPP:*

Release 10: It identifies the requirements and optimizes radio and network for features such as congestion, low power,overload control, identifiers, addressing, subscription control and security.

Release 11 and beyond: It improves the network for device to device communication, M2M gateway, enhancements for M2M group and co-located M2M devices, network selection and steering, service requirements and optimizations.

2] *ETSI:*

M2M network architecture: It defines functional and behavioral requirements of each network element to provide an end-to-end view.

3] *GSMA:*

GSM operation for M2M: It defines a set of GSM based embedded modules that address operational issues, such as module design, radio interface, remote management, UICC provisioning and authentication, and basic elements costs. It also defines use-cases in vertical markets: health, utilities, automotive, and consumer devices.

4] *IEEE:*

802.16p (WiMAX): It optimize air interface for low power, mass device transmission, small bursts, and device authentication. It also defines M2M gateway, co-operative M2M networks, advanced M2M features.

802.11 (WiFi): It updates enables use of sub-GHz spectrum by updating air interface .

802.15.4 (ZigBee): It optimizes air interface for smart grid networks.

5] *WiMAX Forum:*

Network system architecture specification: It defines deployment models with low OPEX, usages, functional requirements which are based on IEEE 802.16 protocols, and performance guidelines for end-to-end M2M system.

6] *WFA:*

Smart grid task group: It promotes the adoption of Wi-Fi within the smart grid through marketing initiatives, government and industry engagement, and technical/certification programs

Healthcare task group: It maintains Wi-Fi as the preferred wireless access technology and increase adoption in the Home and Hospital Healthcare market segment.

7] *OMA:*

Device manageability: It defines requirements for the gateway managed object.

8] *TIA:*

M2M SW architecture TR50: It develops and maintain access agnostic interface standards for monitoring and bi-directional communication of events and information between smart devices and other devices, applications or networks.

9] *CCSA NITS:*

CCSA TC10: It works for the development of pervasive networks, including general requirements, applications, networking, sensing and related short range RF connectivity.

NITS WGSN: It works for the development on sensor network interface and data format, ID and security, vertical applications including airport and smart buildings.

7. Future Research:

1] Growing networks sizes and scalability:

As the number of nodes increases, scalability of the MAC protocol remains an important area of research. The current MAC protocols should have the ability to fulfill channel access request by the ever-growing number of M2M devices. Smarter physical layer signal processing technique that deliver multiple access communication can solve such problems.

2] Quality of Service support:

There are certain mission critical applications where there are stringent delay requirements for correct delivery of data at right time. The designed MAC protocols for M2M communications must guarantee the reliability for delay, throughput and loss requirements.

3] *M2M traffic characteristics:*

Traffic can be periodic, bursty, may have stringent timing requirements or may be elastic. Thus accurately designed MAC layer schedulers or research allocation mechanisms are necessary to exploit the traffic characteristics so as to enable a better utilization of available resources.

4] Extremely low power operation:

M2M devices may have limited or no access to power. Thus MAC protocols needs to be developed that can support low power communications such as on-demand query-initiated transmissions, co-ordinate sleep wake schedule across the nodes and controlling the power.

8. Conclusion

As the mobile internet is approaching towards the embedded internet, M2M pose challenges and opportunities to the industry. Development of new technologies that scale the growth of M2M markets and extensive standardization effort in system interfaces, network architecture and implementation platform is required for the embedded-Internet vision to materialize. The outgrowth of Internet-of-Things in near future is dependent on the growth and development of network support at all layers for supporting M2M communications. The report presented various MAC protocols for enhancing and motivating the development of M2M communications in a more efficient, reliable and secure manner. MAC protocols specific to M2M communications were reviewed. Also the current Standard development organization's efforts to standardization and future research scope were discussed.

9. References:

[1] G. Wu, S. Talwar, K. Johnsson, N. Himayat and K. Johnson, "M2M: From Mobile to Embedded Internet," IEEE Communications Magazine, pp.36-43, April 2011.

[2] A. Rajandekar and B. Sikdar, "A Survey of MAC Layer Issues and Protocols for Machine-to-Machine Communications," IEEE Internet of Things Journal, vol. 2, no. 2, pp. 175-186, April 2015.

[3] M. Hasan, E. Hossain and D. Niyato, "Random Access for Machine-to-Machine Communication in LTE-Advanced Networks: Issues and Approaches," IEEE Communications Magazine, pp. 86-93, June 2013.

[4] Y. Liu, C. Yuen, J. Chen and X. Cao, "A Scalable Hybrid MAC Protocol for Massive M2M Networks," in Proc. of IEEE WCNC, 2013.

[5] C. Hsu, C. Yen and C. Chou, "An Adaptive Multichannel Protocol for Large-Scale Machine-to-Machine (M2M) Networks," in Proc. of IWCMC, 2013.

[6] P. Karn, “MACA-A new channel access method for packet radio,” in *Proc. ARRL/CRRL Amateur Radio Comput. Netw. Conf.*, pp. 134–140, 1990.

[7] F. Azquez-Gallego, J. Alonso-Zarate, I. Balboteo, and L. Alonso, “DPCF-M: A medium access control protocol for dense machine-to machine area networks with dynamic gateways,” in *Proc. IEEE 14th Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 490–494, 2013.

[8] Y. Liu, C. Yuen, X. Cao, N. U. Hassan, and J. Chen, “Design of a scalable hybrid MAC protocol for heterogeneous M2M networks,” *IEEE Internet Things J.*, vol.1, no.1, pp. 99–111, 2014.

[9] N. K. Pratas, H. Thomsen, C. Stefanovic, and P. Popovski, “Code expanded random access for machine-type communications,” in *Proc. IEEE Global Telecommun. Conf. Workshop (GLOBECOM)*, pp. 1681–1686, 2012.

[10] C. W. Park, D. Hwang, and T. J. Lee, “Enhancement of IEEE 802.11ah MAC for M2M Communications,” *IEEE Commun. Lett.* vol. 18, no. 7, pp. 1151–1154, Jul. 2014.

[11] Mario E Rivero-Angeles, Domingo Lara-Rodriguez and Felipe A. Cruz-Perez, "A New EDGE Medium Access Control Mechanism Using Adaptive Traffic Load Slotted Aloha", *0-7803-7005-8/01/$10.00 © 2001 IEEE.*

[12] M. Timmers, S. Pollin, A. Dejonghe, L. V. Perre and F. Catthoor, "A Distributed Multichannel MAC Protocol for Multihop Cognitive Radio Networks," IEEE Transactions on Vehicular Technology, vol. 59, no. 1, pp. 446-459, Jan. 2010.

[13] Wei Zhang, Ranjan K. Mallik, Khaled Ben Letaief, "Cooperative Spectrum Sensing Optimization in Cognitive Radio Networks", *ICC 2008 proceedings, IEEE Communications Society.*