Distributed Gateway Selection for M2M Communication in Cognitive 5G Networks

Muhammad Naeem, Waleed Ejaz, L. Karim, Syed Hassan Ahmed, A. Anpalagan, Minho Jo, and Houbing Song

Abstract

M2M communication is an important component for future wireless networks. M2M systems consist of a large number of devices that can operate with minimum or no human intervention. However, spectrum demand rises exponentially with the increase in the number of connected devices. Cognitive 5G networks are key to address the issue of spectrum scarcity. Further, use of multiple gateways in cognitive 5G networks for M2M communication can increase system throughput, coverage, and energy efficiency. Nevertheless, using multiple gateways for the secondary M2M devices may cause interference to the primary M2M devices. Existing gateway selection protocols for cognitive M2M communication mostly use single channel CSMA, and thus are not efficient in terms of reducing the interference. Thus, in this article, we propose a DGAP based on multi-channel CSMA for M2M communication in 5G networks. Further, we propose a Lo-DGAP, where each gateway transmits only the worst primary M2M device information rather than transmitting all neighboring primary M2M device information. The proposed Lo-DGAP increases the throughput of the system by reducing the message header payload and is also energy-efficient. Simulation results demonstrate the effectiveness of the proposed schemes in terms of network lifetime and energy consumption.

Introduction

Machine-to-machine (M2M) communication is a revolutionary paradigm in the future technology world. M2M has been one of the most popular protocols in IoT recently. M2M communication networks comprise a large number of devices in different applications such as smart homes, smart building management, smart meters, healthcare, and intelligent transport systems [1]. These devices have the capability to sense and communicate with little or no processing. M2M devices are continuously growing and it is expected that there will be 50 billion connected devices in M2M networks and tens of billions of Internet-connected devices by the end of 2020. According to Cisco, the number of wireless devices will increase at the rate of 10 times from 2014 to 2019 [2]. This tremendous increase in the number of M2M devices will result in several challenges, including energy consumption, data rate, quality-of-service (QoS), security/privacy, and traffic congestion.

Many communication technologies (e.g., WiFi, Zigbee smart, Bluetooth smart, Zwave, long-term evolution (LTE), LTE-advanced (LTE-A), 5G networks) have been considered for M2M communications [3]. Despite the fact that LTE and LTE-A offer high bandwidth, ubiquitous coverage, and mobility support, they are not able to fully support M2M communication requirements [4]. The frames of LTE/LTE-A hinder M2M support due to their design, which was originally proposed for broadband human-to-human (H2H) communications. On the other hand, 5G has been envisioned as heterogeneous networks that can provide access to a wide range of M2M applications and access technologies [5]. Moreover, M2M communication is considered to be one of the disruptive technology directions for 5G networks. However, the demand for spectrum raises exponentially with the increase in the number of M2M devices. Therefore, opportunistic spectrum sharing is indispensable in order to achieve the stringent requirements of M2M applications in 5G networks.

Spectrum sharing in 5G networks ensures the coverage of M2M services everywhere and anytime. It can also support a large number of M2M devices with diverse applications and services. Further, it is spectrum-efficient as it uses all the available non-contiguous spectrum. Cognitive radio technology is the key to employing opportunistic spectrum sharing. The architecture of cognitive M2M communication in 5G networks is demonstrated in Fig. 1. This architecture has the main components: a central control unit (CCU) or cloud computing framework, heterogeneous core network, (key technologies massive MIMO, cognitive radios, small cells), and M2M area network. The M2M area networks mainly comprise a large number of M2M devices having sensors with technologies such as WiFi, Bluetooth smart, and RFID. M2M gateways are used as an interface between M2M area and core networks. Data from M2M devices are transmitted to the distant CCU or cloud through M2M core networks/Internet.

In cognitive 5G networks, M2M devices can be classified into primary and secondary devices, where primary devices are legitimate to operate in the spectrum. On the other hand, secondary M2M devices use the spectrum opportunistically. M2M communication in cognitive 5G networks can improve spectrum utilization by allowing unlicensed/secondary M2M devices to use underutilized licensed frequency bands [6]. In geo-location/database schemes, the licensed/unlicensed M2M devices have a location-sensing device (e.g., GPS receiver). The locations of primary and secondary M2M devices are stored in a central database or cloud, as demonstrated in Fig. 1. The central controller (also known as the spectrum manager) for the secondary M2M devices has access to the location database.

The use of multiple gateways can increase the
performance of M2M communication in cognitive 5G networks. A well designed multiple gateway assignment scheme can be helpful in reducing the interference induced to the primary M2M devices in multi-user cognitive 5G system. The use of gateways can also reduce the overall transmission power of the system, which can be helpful in reducing global warming by minimizing CO2 emissions. Further, the use of multiple gateways can increase the performance of a cooperative communication system rather than a single gateway, which conveys the same information to the destination. However, using all the available gateways in the system all the time for the secondary M2M devices in a cognitive 5G system may not be feasible because the interference caused by the gateways to the primary M2M devices may exceed the prescribed limit. The use of multiple gateways in a network comprising a single source M2M device and multiple receiving devices brings the issue of how best to assign the gateways to the secondary M2M devices.

In the centralized gateway assignment, the central station or controller has to know the channel information, primary device information, and available power of gateways, resulting in a large number of control message transmissions. Hence, centralized gateway assignment algorithms are not energy efficient. Therefore, in this paper, we introduce an interference-aware, decentralized multiple gateway assignment protocol (DGAP) for M2M communication in cognitive 5G networks. The objective is to maximize the sum-capacity of M2M systems while minimizing the interference to the primary M2M devices. In DGAP, each gateway device transmits and receives all the licensed/primary user information. However, instead of sending all the primary user information, each gateway transmits only its worst primary user (WPU) information to the destination. By doing so, we either improve DGAP as low overhead DGAP (Lo-DGAP). Both DGAP and Lo-DGAP protocols use the multi-channel carrier sensed multiple access (CSMA) protocol to reduce the number of data collisions in an M2M network having thousands of devices. Following are the main benefits of the DGAP and Lo-DGAP schemes:

- DGAP and Lo-DGAP achieve energy efficiency by reducing the number of control message transmissions among primary and secondary M2M devices and gateway nodes.
- DGAP and Lo-DGAP utilize the concept of rewards for gateway nodes. The reward increases whenever a gateway node co-operates.
- Lo-DGAP increases throughput by reducing the message header payload.
- Lo-DGAP decreases the control channel traffic by only sending the information of the worst M2M device instead of sending the information of all primary M2M devices.

**RECENT GATEWAY SELECTION SCHEMES**

Several gateway assignment schemes have been proposed in the literature that comprise single gateway assignment schemes and multiple gateway selection schemes. Single gateway assignment schemes in the literature include nearest neighbor selection scheme [7], where the gateway that is the nearest to the base station cooperates, and single gateway selection scheme [8], where a path with the maximum signal to noise ratio (SNR) is selected.

Multiple gateway selection that maximizes the SNR is introduced in [9] through an exhaustive search over all gateways, which is computationally inefficient. Thus, multiple SNR-suboptimal (i.e., error rate-suboptimal) gateway selection schemes based on gateway ordering functions are introduced. These schemes achieve full diversity and low error rates. In [10], an auction-based relay assignment scheme in cooperative communication is proposed for centralized and decentralized networks. An optimal relay assignment scheme for cooperative networks is proposed in [11]. The authors proposed a payment mechanism for using relay services to ensure that the collected amount is no less than the amount paid for relay nodes. However, these mechanisms are unrealistic due to a large number of devices involved in M2M communication in 5G networks. Moreover, the centralized algorithm is very complex since it requires maximizing the number of admitted M2M devices while providing QoS.

In [12], a gateway selection mechanism for strengthening inter-cluster coordination in cognitive radio ad hoc networks is proposed. In [13], the authors investigated quantum particle swarm optimization for relay selection in cooperative relay networks. A utility optimal cross-layer resource allocation in distributed wireless cooperative networks is discussed in [14]. In [15], a framework is proposed for the selection of source nodes, relay assignment, and power allocation. However, these algorithms require the devices to learn about the network environment to make real-time decisions on gateway selection, spectrum and power allocation. This can be done by extensive spectrum sensing that consumes more power. Conventional CSMA/CA mechanisms cannot meet the radio sensitivity requirements.

Thus, multi-channel cooperative CSMA schemes can be used that provide accurate spectrum information to the nodes. To the best of the authors’
TABLE 1. Comparison among existing gateway selection protocols.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Distributed</th>
<th>Multiple gateway</th>
<th>M2M applicability</th>
<th>Cognitive capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>[7]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>[8]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>[10]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[11]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[12]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[13]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[14]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lo-DGAP</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1. Comparison among existing gateway selection protocols.

Further, only the proposed DGAP and Lo-DGAP protocols with existing protocols in the literature. Table 1 compares the proposed DGAP and Lo-DGAP protocols with existing protocols in terms of distributed approaches, use of multiple gateways, M2M applicability, and cognitive capability. Table 1 shows most of the work that considered decentralized approaches, since real-world scenarios are always decentralized and the overhead in the centralized network is much more than in a distributed multiple gateway assignment. Further, only the proposed DGAP and Lo-DGAP are considering all the attributes given in Table 1.

**SYSTEM MODEL**

We consider a cognitive 5G network for M2M communications that consists of a large number of M2M devices, as illustrated in Fig. 1. We use a multi-channel CSMA/CA protocol (MC-CSMA/CA) to overcome the system efficiency limitations caused by random signaling structure in CSMA/CA. The MC-CSMA/CA avoids the collision in transmitting data to/from a large number of M2M devices in a complex system. It enables M2M devices to contend with each other for channel access in orthogonal dimensions, that is, both in time and frequency domains. A two-dimensional back-off mechanism and multi-channel sensing are used for this purpose. The MC-CSMA/CA protocol allows splitting the available channel bandwidth into multiple orthogonal narrow-band random access channels. The transmission probability is adjusted according to multi-channel activity in a flexible manner.

We consider that the network comprises $M$ primary M2M devices and $K$ secondary M2M devices, where the transmission of secondary M2M devices is limited. We assume that the total bandwidth available is $B$, which is divided into $N$ sub-channels of equal bandwidth. The spectrum overlay approach is used in this network. A secondary M2M device needs cooperation to avoid interference to the primary M2M devices. Decode and forward (DF) cooperative relaying is used in this protocol. DF relaying is executed in two time slots. In the first time slot, the M2M device transmits its data, and in the second time slot, the gateway first decodes the data received in the first time slot and then retransmits the data again to the receiver. We assume that a gateway can only cooperate with one source-destination pair at one time and any M2M device in the network can act as a gateway node. However, which device should act as a gateway for any particular M2M device needs a protocol that can take care of spectrum and interference issues.

The energy consumption for M2M device for transmission and reception of a packet size $n$ over a distance $d$ is given by $E_{\text{tr}} = n \times e_{\text{elec}} + n \times e_{\text{fs}} \times d^a$ and $E_{\text{rx}} = n \times e_{\text{elec}}$, where $e_{\text{elec}}$ and $e_{\text{fs}}$ represent energy spent in transmitter electronic circuitry and energy spent in RF amplifiers. The constant propagation loss exponent $a$ depends on the surrounding environment. For free space without any obstruction in the line of sight, $a = 2$.

**DISTRIBUTED GATEWAY ASSIGNMENT PROTOCOLS**

In this section, we present the working principle of the proposed DGAP and Lo-DGAP schemes for M2M communication in cognitive 5G networks.

**DGAP Protocol**

In the DGAP protocol, each source node will compete to get an idle channel by using a multi-channel CSMA protocol. Figure 2 illustrates the frame format of the DGAP protocol. The message format of each protocol comprises a number of control messages. All the frames have a preamble at the start and postamble at the end to get the channel state information. To differentiate each control message in the protocol, we insert different identifiers at the start of each frame. These identifiers include source help request (SHR), destination information request (DIR), reply to SHR (RSHR), reply to DIR (RDIR), and source reply message (SRM). Each of these control messages has a number of fields.

Once a source (M2M device) obtains the channel successfully by using a multi-channel CSMA protocol, the source initiates DGAP. Both signaling and data transfer in DGAP have the same sub-band for operations. The signaling protocol is used to get the information about available gateways and their associated parameters, for example, channel information, PU informa-
Decentralized gateway assignment protocol.

Each M2M device receives a positive or negative reward based on its activity. If the M2M device cooperates with the other nodes it receives a positive reward. Each M2M device will only cooperate if the interference induced by this device is less than a certain threshold set by the primary network. On the other hand, if the M2M device receives cooperation from other nodes, it receives a negative reward. For instance, we denote the RS of the kth M2M device at time t by $\Gamma_k(t)$ and $K$ as the maximum number of M2M devices that can simultaneously cooperate with any other M2M device. Let $\phi$ be the set of M2M devices that can cooperate with the kth M2M device at time $t + 1$, then reward of the kth M2M device will decrease at time $t + 1$, which is denoted as

$$\Gamma_{k,t+1} = \Gamma_k(t) - (\alpha_1 \sum_{i \in \phi} p_i + \beta_1 |\phi|),$$

where $p_i$ is the power of the ith M2M device in the set $\phi$. The constants $\alpha_1$ and $\beta_1$ are set by the wireless design engineers. $\alpha_1 \sum_{i \in \phi} p_i + \beta_1 |\phi|$ restricts the source M2M device to utilize the minimum power of the gateways and a minimum number of M2M devices to get its required data rate. Each gateway will get its RS for cooperation, for example, the ith gateway will get the cooperation reward $\Gamma_{g,i} = \Gamma_g(t) + a_2 p_i + \beta_2 (1/|\phi|)$, where $a_2$ and $\beta_2$ are weights for $p_i$ and $(1/|\phi|)$, respectively. The gateway nodes that have more rewards have more cooperation with other M2M devices and thus dissipated more energy. Thus, a gateway node with more reward is less likely to be selected to cooperate with the source M2M device in the next round.

A flow diagram of source operations is given in Fig. 4(a). The gateway nodes that are available for cooperation will listen to the SHR message. A gateway node can get an SHR request from multiple sources in multiple sub-bands. The gateway node then chooses the source for its cooperation whose channel and RS are the best for it. After sending the SHR message, the source M2M device waits for a certain period of time. If no response is received in a specified time, the source node again sends the SHR message. Otherwise, if a reply for the SHR message is received, then the source determines the gateway power and sends an SRM to the gateway. After setting up the connection, the source M2M device can now transmit data and wait for the acknowledgment. The source M2M device will release the channel when it completes data transmission.

Multiple gateways can offer their services to the source M2M device for data transfer. A flow diagram of gateway operations is given in Fig. 4(b). Each gateway has its SNR threshold for cooperation. After receiving an SHR message from the source M2M device, the gateway selects the source M2M based on the reward and SNR level. If the SNR level of the gateway is less than a threshold, it will not cooperate. The gateway then sends a DIR to the destination. There are a number of sub-channels in the sub-band. Each gateway will randomly choose its sub-channel for the DIR. If there is a collision in the sub-channel, then the gateways that are involved in that collision will set a back-off timer and again retransmit when the back-off time is zero. In the DIR frame, the gateway sends the GID, DID, PU, the SNR received at the gateway, GAP, and RS. After sending the DIR, the gateway will wait for the RDIR from the destination M2M device, and if received successfully from the destination it will send an RSHR to the source M2M device. Otherwise, it will again select the source M2M device based on the reward level and its own SNR level.

A flow diagram of the destination operations is
shown in Fig. 4(c). Once the destination receives the DIR message from multiple gateways in a predefined time window, the destination selects the best gateway according to their RS and SNR level and sends an RDIR message to the selected gateway. This message contains the GID of the gateway that will resend an NI message to the source. We call it a selected gateway for the source (SGS). The destination will select the gateway for the SGS that has the best channel with the source and destination. The advantage of a single gateway reply to the source is simple implementation and low overhead. The selected gateway from the RDIR message sends an SHR message in reply to the source. After getting the RSHR message, the source determines the power of the selected gateways and sends the power information to the gateways via an SRM message. After this signaling

FIGURE 4. Flow charts for the DGAP a) for source M2M device, b) for gateway node, and c) destination device.
protocol is followed, the source starts sending data to the destination with the help of selected gateways.

**Lo-DGAP Protocol**

We can improve the DGAP to further decrease the traffic of the control channel. In the case of DGAP (Fig. 2), each gateway transmits and receives the information about all the PUs via DIR and RDIR messages. In contrast, in Lo-DGAP, each gateway only transmits its worst primary user (WPU) information instead of sending all the primary user information. By sending only the WPU information, Lo-DGAP can decrease control channel traffic since the WPU represents the closest primary M2M device to the source node that is affected.

Let us assume that the SID, DID, GID, SHR, DIR, RDIR, RSHR, SRM and SGS devices. For simulations, we assume that the SID, DID, GID, SHR, DIR, RDIR, RSHR, SRM and SGS require three bits and all other elements require four bits.

We compare the network lifetime of M2M devices in cognitive 5G networks using DGAP and Lo-DGAP over a number of rounds, where each round comprises a network setup and steady (routing) phases. We assume that there are 10 primary M2M devices among a total of 100 devices. The number of gateways G is set to 5 and 10 in the simulation. Figure 5(a) compares the performance of DGAP and Lo-DGAP in terms of network lifetime. Figure 5(a) shows the network lifetime versus number of rounds for the different number of gateways. Since Lo-DGAP is transmitting a smaller number of data bits compared to DGAP, the lifetime of the network using DGAP is shorter than that using the Lo-DGAP scheme. Moreover, the network lifetime is less in the case of G = 10 compared to G = 5.

Figure 5(b) shows the network energy consumption versus the number of primary M2M devices for different numbers of gateway nodes. It is observed that Lo-DGAP consumes less energy as compared to the DGAP because the increasing number of primary M2M devices also increase the number of payloads in DGAP. However, the payload does not increase in Lo-DGAP when increasing the number of primary devices since each gateway only transmits information of its WPU in Lo-DGAP. Since in Lo-DGAP only WPU information is used in the header, it reduces the header overhead significantly. For example, let there be 25 PUs, with 2 bytes required to carry the information of one PU. In this case for 25 PUs, there is a need for 50 bytes in the header. In the worst PU scenario, the protocol only requires two bytes. In a nutshell, Lo-DGAP not only reduces energy consumption but also reduces the transmission header overhead significantly.

Figure 5(c) compares the message header payload of DGAP with Lo-DGAP when varying the number of gateways G and primary users M. In this simulation, we assume that the SID, DID, GID, SHR, DIR, RDIR, RSHR, SRM and SGS...
is achieved by transmitting only the worst primary user information rather than transmitting all primary user information in DGAP. Simulation results reveal that the Lo-DGAP scheme is energy efficient and requires fewer header bits compared to DGAP.

**CONCLUSION**

The demand for radio spectrum increases exponentially with the massive number of connected devices, particularly with the emergence of M2M communication. This paper discussed M2M communication in cognitive 5G networks. We proposed DGAP for the gateway assignment for M2M communication in cognitive 5G network based on multi-channel CSMA. This protocol is efficient in terms of data communication overhead compared to its centralized counterpart. The low overhead DGAP (Lo-DGAP) further improves the DGAP by reducing control channel traffic, which is achieved by transmitting only the worst primary user information rather than transmitting all primary user information in DGAP. Simulation results reveal that the Lo-DGAP scheme is energy efficient and requires fewer header bits compared to DGAP.

**ACKNOWLEDGMENT**

This research was supported in part by a grant from the NSERC Canada Discovery Grant. The research was also supported by the National Research Foundation of Korea Grant funded by the Korean Government (MOE) (Grant:2016R1D1A1B03932149).

**REFERENCES**


**BIographies**

Muhammad Nazem received his Ph.D. degree from Simon Fraser University, BC, Canada. From 2012 to 2013, he was a postdoctoral fellow at Ryerson University, Toronto. Since 2013, he has been with the Department of Electrical Engineering, COMSATS Institute of IT, Wah Campus, Pakistan. His research interests include optimization of wireless systems and non-convex optimization.

Waleed Ejaz [S’12, M’14, SM’16] received the Ph.D. degree in information and communication engineering from Sejong University, South Korea. He is currently a senior research fellow with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada. His current research interests include Internet of Things, energy harvesting, 5G cellular networks, and mobile cloud computing.

Lutfurul Karim received his Ph.D. degree from the University of Guelph, ON, Canada. Currently, he is working as a professor in the School of ICT, Seneca College of Applied Arts and Technology, Toronto, Canada. His research interests include wireless and mobile sensor networks, ubiquitous and pervasive computing, and communication protocols and algorithms.

Syed Hassan Ahmed [S’13, M’17] received his B.S. in computer science from Kohat University of Science and Technology, Pakistan. Later, he completed his masters combined Ph.D. in computer engineering from the School of Computer Science and Engineering, Kyungpook National University (KNU), Korea in 2017. In 2015, he was a visiting researcher at the Georgia Institute of Technology, Atlanta, USA. He has published over 70 articles in international journals and conferences, as well as two Springer brief books. From 2014 to 2016, he won the Best Research Contributor award in the workshop on Future Research in Computing of Computer Science and Engineering, KNU. In 2016, he also won the Qualcomm Innovation Award at KNU. He is also an IEEE and ACM member while serving several conferences and journals as a TPC and reviewer, respectively. His research interests include sensor and ad hoc networks, cyber-physical systems, vehicular communications, and Future Internet.

Aalagan Anpalagan is a professor in the Department of Electrical and Computer Engineering, Ryerson University, Canada. He directs a research group focusing on network management, spectrum management, and radio access and networking (RAN) within the WINCORE Laboratory, with research interests in 5G wireless systems, energy harvesting and green communications technologies, M2M and sensor communication, small cell, and heterogeneous networks. He is a Fellow of the Institution of Engineering and Technology.

Minho Jo [M’07, SM’16] received the Ph.D. degree from the Department of Industrial and Systems Engineering, Lehigh University, USA, in 1994. He is currently a professor with the Department of Computer Convergence Software, Korea University, Sejong. His areas of interest include LTE-Unlicensed, cognitive radio, Internet of Things, mobile cloud computing, machine learning in IoT, network security, and massive MIMO.

Houying Song [M’12, SM’14] received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in August 2012. In August 2017, he joined the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL, where he is currently an assistant professor and the director of the Security and Optimization for Networked Global Laboratory (SONGLab, http://www.songlab.us/). He served on the faculty of West Virginia University from August 2012 to August 2017. He has served as an associate technical editor for IEEE Communications Magazine since 2017. He is a senior member of IEEE. In 2016, he was the first recipient of the Golden Bear Scholar Award, the highest faculty research award at West Virginia University Institute of Technology (WVU Tech), in 2016.