Improving the throughput of IoT devices for smart cities

K. Ramalingham
ramu.k301@gmail.com
Sri Vidya College of Engineering and Technology, Virudhunagar, Tamil Nadu

Dr. P. Murugeswari
sankram301@gmail.com
Sri Vidya College of Engineering and Technology, Virudhunagar, Tamil Nadu

ABSTRACT

Device-to-Device (D2D) communications and applications are expected to be a significant part of the Internet of Things (IoT). However, conventional network gateways reported in the literature are unable to provide sustainable solutions to the challenges posed by the massive amounts of D2D communications requests, especially in the context of the IoT for smart cities. In this paper, we present an admission control model for D2D communications. The model differentiates all D2D requests into delay-sensitive and delay-tolerant first and then aggregates all delay-tolerant requests by routing them into one low-priority queue, aiming to decrease the number of requests from the various number of devices in the IoT for smart cities. Also, an admission control algorithm is devised on the basis of this model to prevent access collision and to improve the quality of service. Performances are evaluated by the network calculus, numerical experiments, and simulations show that the proposed model is feasible and effective.

Keywords: Device-to-device (D2D) communications, Internet of things (IoT), Smart city, Hybrid admission control, Performance analysis

1 INTRODUCTION

The development of the Internet of Things (IoT) together with smart cities bring us not only new opportunities but new challenges as well. As part of the IoT, machine-to-machine (D2D) communications either broadly refer to the entire communications among man, machine, and system, or narrowly refer to the communications among machines (devices) only. D2D communications can be understood as an automated process which requires minimum human interventions

[1] Tasks such as remote surveillance
[3] Home and traffic security, and intelligent transportation
[4] Instances of D2D communications. Personal navigation, e-pay, and industrial automation are also expected to be benefited from D2D communications.

Recent data collection suggests that the percentage of D2D connections in all Internet connections will grow from 24% to almost 43% by the year 2019. However, it has been observed that massive and concurrent machine accesses and radio signals generated by D2D devices (machine type communication devices (MTCDs)) in smart cities may cause unacceptable communication delays, data packet losses, and even service interruptions in human-to-human (H2H) communications. Also, the sporadic and diverse nature of D2D traffic calls for a new design of wireless networks to deal with it. One of the major challenges in the IoT for building sustainable smart cities resides in how to effectively handle the dramatic explosion of the number of connected devices. For example, the total number of mobile devices in 2014 was around 7.4 billion, nearly the same as the world population. According to data predictions, by 2019 there will be almost 11.4 billion mobile devices in the world. D2D connections are also expected to increase from 495 million in 2014 to almost 3 billion by 2019. Furthermore, the portion of D2D devices in cellular networks is expected to increase from 1% in 2014 to over 20% by 2019. As such, building sustainable smart cities needs to consciously consider not only the aggressive growth of mobile devices (including MTCDs) but also the consistently increasing traffic demand per device. With the recent activities of 3GPP, ETSI and IEEE standardization bodies intending to provide protocols and standards for D2D applications, note that most existing 4G base stations are designed to provide broadband services to regular H2H subscribers and that D2D communications typically transmit small-sized packets in a frequent manner by using sporadic radio resources. So D2D communications are unable to effectively take advantage of the 4G Communication channels. Although the idea of random accesses
from MTCs to channels may mitigate this problem, the huge number of MTCs congested on channels will lead to a significant rise of collisions, a higher packet loss rate, and a performance degradation for both D2D and H2H services.

Incidentally, the co-existence of D2D and H2H communications is essential for both service providers and users. As such, to enhance the network performance, the spectrum utilization efficiency needs to be maximized while the D2D random accesses need to be minimized. Although extensive studies have been conducted for D2D communications in various aspects, the primary challenge in D2D communications lies in how to deal with the frequent and massive amounts of access requests sent from the exponentially increasing number of D2D devices in smart cities. Given the huge amounts of access requests raised by MTCs in smart cities, the traditional network gateway is no longer able to handle these requests satisfactorily. An effective admission model designed for D2D communications can not only reduce the number of collisions caused by MTCs’ random accesses but also ensure an effective exploitation of the wireless resources.

It is well known that network calculus is an important and effective mathematical tool for the quantitative study of network system performances, and has been widely used in the modeling and analysis of the quality of service (QoS) of networks. In the network calculus (NC) calculates the delay bound, backlog bound, and other service quality parameters by using arrival curves and service curves. Compared with the traditional queueing theory, DNC is able to provide a determined boundary analysis for system performance, and offer a strict service guarantee by computing the worst-case scenarios. Taking the advantage of network calculus being a systematically structured theory, we, in this paper, propose a priority-based admission control model for D2D communications in smart cities and analyze its performance by leveraging network calculus.

The Main Contributions Of This Paper Are As Follows:
• We present a new IoT architecture for smart cities, through which the network control and data transmission are separated. This architecture follows the same design philosophy of SDN, and thus enhances the manageability and the controllability of the entire network. Under this architecture, we propose an admission control model which first differentiates D2D access requests as delay-sensitive and delay-tolerant, and then aggregates all delay-tolerant requests into a low-priority queue. The proposed admission control model can effectively reduce the number of needed connections from MTCs to base stations, and mitigate the possibility of collisions on channels generated by MTCs’ random requests.
• We present an admission control algorithm for massive D2D requests in smart cities. The algorithm can effectively prevent access request congestions, thereby improving the quality of D2D access connections. We also apply the network calculus to analyzing the performance of the proposed algorithm. Performance bounds including the worst-case delay and backlog bounds, which provide design guidelines for building sustainable smart cities, are derived.
• We evaluate the proposed model and examine the theoretical results by conducting extensive experiments. The validness and effectiveness of the developed theory are further confirmed. The idea of aggregating access flows of a massive number of devices/machines in the context of smart cities can effectively enhance their sustainability.

2. ARCHITECTURE
Standardization of D2D communications, together with related requirements and architectures, have already been proposed developments in home D2D networks, summarizing the architecture D2D communications and positioning some related to be supported by an existing architectural framework. Only under such an architectural framework support, can ubiquitous MTCs be effectively allowed to access base stations. There is an urgent need for the IoT to become an open, challenges. These challenges were about the large-scale maintenance of devices and remote management, explicated the D2D communication architecture and performance in LTE-advanced networks. Solutions to D2D support and LTE resource management were proposed. The existing D2D solutions can be divided into two categories:
(i) Radio resource optimization
(ii) Co-operation among devices.

Viewed from the networking perspective, radio resource management is of the utmost importance in terms of maintaining a certain level of QoS. Massive access management and reliable resource pooling schemes were proposed to ensure the QoS and the reliability of D2D operations proposed a scalable hybrid MAC protocol for machine type communications within heterogeneous networks. A batch data model was suggested to reduce the updating frequency of D2D core networks. A self-adaptive access barring parameter was used to optimize the system performance by changing resource blocks. IEEE 802.11 ah MAC for D2D communications was enhanced with the mechanism of self-adaptive Restricted Access Windows (RAW). Various MAC protocols for D2D communications were surveyed to support more machine accesses, MTCs, just like the base stations, can be grouped together to collaborate with one another toward load balancing or resource sharing. A co-operative access class barring protocol to balance the number of MTC requests in overlapping macro- and micro-cell coverage determines the optimal durations for data transmissions from MTC devices to gateway and from gateway to server.

**Fig. 1: The architecture of the D2D-supporting IoT system for smart cities**
In wireless sensor networks (WSNs) and other wireless networks, there are some notable work to reduce the network load caused by data transmissions. Although the technique of self-adaptive data compression has been used in WSNs for bandwidth management and location updating, it has not been addressed in the context of D2D communications, which consists of enormous concurrent transmission requests from a myriad of devices. This observation motivates us to explore new architectures and techniques in designing novel and more efficient strategies to handle the admission control in D2D gateways.

3. D2D HYBRID ADMISSION CONTROL MODEL

3.1 System Model

The successful design and implementation of D2D communications in the IoT for sustainable smart cities must complete, standardized, and universal architectural framework, which facilitates various newly developed techniques, including D2D communications, to be included in a consistent and effective manner.

Unfortunately, such an expectation is not met by the current IoT. Much like its concept, the current IoT lacks a widely-agreed, uninformed, and normalized architecture to support most conceivable functions in the world. In this section, we sketch a new IoT management and control architecture that intends to support intelligent D2D communications in smart cities. As shown in Fig. 1, the proposed architecture comprises 3 layers: a device layer, an access layer, and a control layer. The device layer is located at the bottom of the system and consists of a variety of wireless terminals. The access layer is located at the middle of the system and consists of cellular base stations and/or Wi-Fi access points, which allows the bottom layer devices to access networks through cellular and/or Wi-Fi technologies. The top layer is the control layer which provides administration and control over the access requests and data transmission activities at the lower layers and is connected to the Internet for real-time analyses and feedbacks.

It is worth mentioning that the devices and infrastructures in the device and access layers only forward data traffic, while the control mechanisms located in the control layer are dedicated to network control or management. This design follows the same philosophy of the Software-Defined Networking (SDN) paradigm that decouples the network control and data forwarding functionalities in order to enhance the controllability and manageability. Also, the interactions between the control and access layers are consistent with operations specified in the SDN protocol, this interaction-formed protocol is assumed given in this work.

3.2 Hybrid Admission Control Model

We now introduce a request aggregating Hybrid admission control model for D2D communications, which is depicted in Fig. 2. In the access layer depicted in Fig. 1, the incoming D2D requests for each base station or access point are classified either as delay-sensitive or as delay-tolerant. In conformity with this classification, each base station or access point is equipped with two queues inside: one is of high priority and is used to queue delay-sensitive requests; the other is of low priority and is used to queue delay tolerant requests. As indicated in Fig. 2, when a D2D request arrives at a base station or an access point, it will be buffered first, and then be routed to the high priority queue if it is delay-sensitive, and the low priority queue otherwise. In this case, all delay-tolerant D2D requests will be aggregated into one low priority queue waiting to be batch-processed. This would not be the case without the request aggregation idea: different delay-tolerant D2D requests would be routed to different queues to be processed separately, depending on the extent of their delay-tolerance.

It should be noticed that the starvation of the low priority flow would be an issue that cannot be avoided in the above model. However, as the main focus of this work is to develop a priority-based model for D2D request access in smart cities to guarantee the delay performance of D2D communications, the issues of how to avoid the starvation and to ensure the fairness of multiple flows (no guarantees regarding delays can be provided in this case), are not discussed here. Also, the above model is suitable for smart city scale applications. For example, consider a public safety scenario in a smart city where a severe flood warning is being issued. In this case, the data update on water levels of surrounding bayous, rivers, ditches, and culverts in various areas of the city is of critical importance and forms a delay-sensitive request flow. Other requests, such as air pollution monitoring, vehicle parking, transportation scheduling, etc. are aggregated as the delay-tolerant flow. Note that the criteria for flow classification vary from one case to another. For other scenarios, the criteria for determining delay-sensitive or delay-tolerant would be different. This issue is out of the scope of this paper but is worth of studying in future investigations.

Network calculus in R-Tool will be used to analyze the performance of this setting. We assume that all access request flows are regulated by the common technique of token bucket so that the request flows can be processed smoothly. The basic working mechanism of token bucket is as follows. Let r be the rate of adding tokens to the bucket, i.e., one token will be added to the bucket per 1/r second, and b be the maximum number of tokens the bucket can hold. If a token arrives when the bucket is full, then that token will be dropped. When a flow containing n requests arrives, n tokens will be removed from the bucket. The flow will be sent
to the network if the bucket has more than n tokens available; otherwise, the flow has to wait, until the bucket acquires a sufficient number of tokens, to be transmitted further. As such, the input function of a flow, after being shaped by the token bucket, is \( f(t) = rt + b \), where \( r \) is the rate and \( b \) is the initial burst traffic.

4. THE HYBRID ADMISSION CONTROL ALGORITHM

Algorithm 1 HYBRID Admission Control Algorithm

Input:
\( r_i, r_h, r_l, r_i \in \{1, 2, ..., 1\} \)

Output:
Accept/ reject request \( i \in \{1, 2, ..., 1\} \): For each current request
3: if \( r_i \notin r_h - r_l \) then
4: else
5: reject request;
6: end if
7: update \( r_h, r_l \);
8: Repeat the process for the next due request;

Due to the enormous amounts of D2D communication requests in smart cities, an acceptance and/or rejection algorithm with regard to these requests must be in place to avoid traffic congestion in the network. We thus devise such an algorithm to handle the large-scale D2D access requests on the basis of their arrival rate and the service capability of the service node.

One of the critical conditions to ensure a regular troublefree running of a network is that the arrival rate of the data flow cannot be larger than the service rate (or capability) of the network. Otherwise, the flow will tend to encounter an infinite delay, causing a malfunction of the network. If the current service capability of the network is able to handle the incoming access request, then the request will be accepted; otherwise, it will be rejected. The notations used in the algorithm and the algorithm itself are given in and Table 1 and Algorithm 1, respectively.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>The total service rate (capability) of the system</td>
</tr>
<tr>
<td>( r_i )</td>
<td>The arrival rate of the ( i )-th D2D device</td>
</tr>
<tr>
<td>( r_h )</td>
<td>The arrival rate of the unfinished high-priority flow</td>
</tr>
<tr>
<td>( r_l )</td>
<td>The arrival rate of the unfinished low-priority flow</td>
</tr>
<tr>
<td>( I )</td>
<td>The total number of D2D devices</td>
</tr>
</tbody>
</table>

5. PERFORMANCE ANALYSIS

By the model in Fig. 2, all D2D communication requests will be routed into a delay-sensitive high-priority queue or a delay-tolerant low-priority queue. In this section, we use \( \text{fh} \) and \( \text{fl} \) to denote the flows generated by the high-priority queue and the low-priority queue, respectively; \( R_h \) and \( R_l \) to denote the input and output functions of \( \text{fh} \), and \( R_l \) and \( R_l \) to denote the input and output functions of \( \text{fl} \). We assume that the order of arrivals of \( \text{fh} \) and \( \text{fl} \) is completely arbitrary without any timing constraints, that the service curve " offered by the system follows the rate latency function, i.e., \( r(T(t) = r(t - T) \), and that \( \text{fl} \) and \( \text{fh} \) have \( e_l(t) = r(t) + b_l \) and \( e_h(t) = r(t) + b_h \) as their arrival curves, respectively. (Note that the reason of having these linear arrival curves was explained at the end of Section 3.2.) In the sequel, we analyze the performance of the system by the order of arrivals of \( \text{fl} \) and \( \text{fh} \) and by considering cases of preemptive scheduling and non-preemptive scheduling.

5.1 \( \text{fh} \) Arrives Earlier than or Simultaneously with \( \text{fl} \)

In this case, preemptive scheduling or non-preemptive scheduling will make no difference. We assume that some requests exist in the system before the D2D requests arrive and are pushed into the queues. These pre-existing requests in the system are called unfinished requests. The system will first serve those unfinished requests at hand, and then start serving requests from \( \text{fh} \) and \( \text{fl} \) in order. Let \( l_{max} \) be the amount of unfinished requests in the system, \( r \) be the total service rate of the system. Then in the time interval \( (s, t] \), the amount of requests formed in \( \text{fh} \) is

\[
R_{\text{fh}} (t) = R_{\text{fh}} (s) \quad \text{and} \quad r(t - s) - l_{max}
\]

5.2 \( \text{fh} \) Arrives Later than \( \text{fl} \)

In this case, preemptive scheduling or non-preemptive scheduling yields different results, and needs to be considered separately.

5.2.1 Non-Preemptive Scheduling

By virtue of the nature of the non-preemptive scheduling algorithm, the system will serve its unfinished requests first, and then start processing requests from \( \text{fl} \) and \( \text{fh} \) in order. As such, the amount of requests formed in \( \text{fl} \) in the time interval \( (s, t] \) is

\[
R_{\text{fl}} (t) = R_{\text{fl}} (s) \quad \text{and} \quad r(t - s) - l_{max}
\]

5.2.2 Preemptive Scheduling

In this case, with the assumption that the current \( \text{fh} \) is empty, the system will serve requests from \( \text{fl} \) until the request from \( \text{fh} \) arrives. At that time, the system will stop processing requests from \( \text{fl} \) and start processing requests from \( \text{fh} \) until all requests from \( \text{fh} \) have been processed, and then resume the processing of requests from \( \text{fl} \). As such, the amount of packets transmitted by \( \text{fh} \) in the time interval \( (s, t] \) is

\[
R_{\text{fh}} (t) = R_{\text{fh}}(s) \quad \text{and} \quad r(t - s)
\]
6. EXPERIMENTAL RESULTS (Using R-Tool)

In this section, we continue the theoretical performance analysis carried out in the previous section together with the parameters are set as follows:

\[ r \in [300, 500] \text{ requests/second (r/s)}, \; r_l = r_h = 50 \text{ r/s}, \; b_l = b_h = 10, \text{ and } l_{\max} = 10. \]

Note that these parameters used here are intended to show the geographical features of theoretical results derived from previous section. Other parameters can also be configured to conduct the same analysis. The delay bound and backlog bound of \( f_h \) with respect to preemptive scheduling and non-preemptive scheduling. It can be seen clearly that preemptive scheduling delivers a superior performance than non-preemptive scheduling. This observed result can also be derived by purely analyzing the relevant bounds obtained in the previous section.
(a) Two flows/queues (Q1 = high-priority, Q2 = low-priority).

```python
description = "multi-priority queues"

network = MultiQueue
**.stopTime = 300s
**.Flow1.jobType = 1
**.Flow2.jobType = 2
**.Flow1.interArrivalTime = exponential(0.02)
**.Flow2.interArrivalTime = exponential(0.02)
**.serviceTime = exponential(0.002s)
**.fetchingAlgorithm = "priority"
```

(a) Parameter settings.

Fig. 6: Setup of the Devices experiment using Arduino IDE, Blynk, Thinks Speak Software

Fig. 7: Output of RFID Traffic System on Thinks Speak Cloud

Fig. 8: Output of LDR Lights Automation on Thinks Speak Cloud

Fig. 9: Output of PIR Security on Thinks Speak Cloud
7. CONCLUSION

The exponential increase in the amount of D2D devices introduced by the rapid development of the IoT in smart cities can no longer be adequately handled by the traditional network gateway techniques. How to effectively deal with such an enormous amount of D2D access requests and the ensuing data transmissions in IoT thus becomes the bottleneck hindering the development of sustainable smart cities. Considering that most D2D requests are delay tolerant, we have presented an architecture for IoT control and management in smart cities and proposed a priority-based model for D2D communications, which can reduce the collision possibility caused by random D2D accesses on wireless channels. The performance of this model is subsequently analyzed and evaluated by using R-Tool, Things Speak Cloud, C# & ASP.net applications. The consistency of results in R-Tool, Things Speak Cloud, C# & ASP.net applications validates the effectiveness and correctness of the proposed model.

For future work, we plan to investigate the following problems based on the current work.

- The multi-priority model would cause the starvation of the low priority flow, although the model can guarantee the delay performance. How to resolve this starvation issue and assure the fairness of the flows are interesting topics worthy of further investigations.
- The network calculus used in this paper is deterministic. It would be interesting to extend the current work by using stochastic network calculus in R-Tool to analyze the proposed model as it offers a more generic treatment for real-world IoT for sustainable smart cities.

REFERENCES


