A case study for M2M traffic characterization in a smart city environment

Filippo Malandra
Dept. of Electrical Engineering
Ecole Polytechnique de Montreal
filippo.malandra@polymtl.ca

Steven Rochefort
Ericsson Canada
steven.rochefort@ericsson.ca

Pascal Potvin
Ericsson Canada
pascal.potvin@ericsson.ca

Brunilde Sansò
Dept. of Electrical Engineering
Ecole Polytechnique de Montreal
brunilde.sanso@polymtl.ca

ABSTRACT

This paper presents a case study to characterize Machine-To-Machine (M2M) traffic in a smart city environment. Real data on the position of machines and three different probability distributions (Poisson, Beta, and deterministic) are used to model the packet generation of some realistic IoT applications. A web application is presented and allows to perform a variety of analyses on M2M traffic characterization. The island of Montreal is used as study case: realistic data on the position of machines and of eNodeB stations in a real LTE network are employed to demonstrate the possibilities of the tool. In the numerical results, the traffic generated by different M2M applications is presented and some differences between M2M and human traffic and their impact on the LTE infrastructure are highlighted.

KEYWORDS

Machine-to-machine, M2M communications, M2M traffic characterization, Internet of Things, Smart City, LTE

ACM Reference format:
Filippo Malandra, Steven Rochefort, Pascal Potvin, and Brunilde Sansò. 2017. A case study for M2M traffic characterization in a smart city environment. In *Proceedings of International Conference on Internet of Things and Machine Learning, Liverpool, UK, October 2017 (IML 2017),* 9 pages.
DOI: 10.1145/3109761.3109809

IML 2017, Liverpool, UK

DOI: 10.1145/3109761.3109809

1 INTRODUCTION

In the last decade, a large number of machines have become *smart*: many enhanced features are available thanks to the possibility of connecting these devices over the Internet forming the so-called *Internet of Things* (IoT). The IoT is also linked to other well established research areas, such as the *Smart Grid* and the *Smart Cities*, encompassing a large number of different applications, such as building security, smart health-care, and vehicle recovery. The number of devices to be connected is expected to increase to several billions in the next few years. Therefore, it is important to find a communication infrastructure able to support Machine-To-Machine (M2M) communications. LTE, the $4th$ generation of cellular communications, is a good candidate solution because of its extensive coverage and wide capacity availability. 3GPP, the organization in charge of developing and maintaining LTE, is very active towards the introduction of massive M2M traffic in the cellular networks. In particular, in release 14 they presented the so-called Narrow Band Internet of Things (NB-IoT), which is intended to become the new standard to provide low power access to a large multitude of devices using the cellular infrastructure and is expected to be one of the major enablers of M2M communications.

However, the cellular infrastructure was originally conceived to support Human-to-Human (H2H) communications, which considerably differ from M2M communications. The main differences are:

uplink / downlink proportion: most data traffic on cellular networks flows in the downlink direction and bandwidth and resources are allocated accordingly. On the other hand, M2M traffic flows predominantly in the opposite direction;

data size: M2M traffic packets are mostly in the order of few hundred Bytes, considerably smaller than regular H2H data traffic;

traffic distribution: smart-phones activity times are strongly correlated with the human daily routine, which is quite well known and analyzed; M2M traffic, on the other hand, is mostly affected by other factors, such as the type of involved M2M application;

device mobility: M2M devices are often static, and in general less mobile than humans;

application usage: M2M devices often use custom application protocols, much more difficult to be predicted than standard protocols.

As a consequence, the massive introduction of M2M traffic on the cellular infrastructure might degrade the overall network performance. It is therefore fundamental to gain additional insight into the characterization of M2M traffic. This would assess the suitability of cellular networks to support M2M communications, and also the potential impact of these on H2H communications.

The Radio Access Network (RAN) is the component of the cellular infrastructure that is expected to be the most affected by M2M communications. This is due to the fact that wireless resources are limited and potential users need to compete for them through the so-called Random Access procedure, which differs depending on the type of technology which is used. The competition for the available resources might become considerably harsher if the number of users increases, as pointed out by several machine data forecasts. 3GPP presented in [1] the potential impact of M2M communications on the RAN, illustrating possible solutions. This work spurred a large deal of literature on the subject. Nevertheless, some gaps of literature need to be filled in order to keep pace with the expected growth of M2M communications.

First of all, there is a lack of studies considering the real position of machines. We believe that the geographic location of machines needs to be considered because it permits to better grasp the peculiarities of M2M traffic. This is particularly important in a *smart city* environment, where the density of machines can particularly jeopardize the network performance. Second, only few probability distributions are used to model the traffic generated by different M2M applications (i.e., uniform and Beta distribution). Moreover, we could not find any study simultaneously evaluating different M2M applications, although a single cellular infrastructure is expected to support a wide variety of different M2M applications at the same time.

In this paper, we propose a web-based framework, referred hereafter as *web application*, that takes into account the geographic location of machines, a thorough traffic generation for a broad set of realistic M2M applications, and a realistic LTE network infrastructure. Section 2 presents previous work on M2M traffic characterization; Section 3 describes our data collection on the position of machines; Section 4 illustrates the traffic generation used in the framework; Section 5 presents details on the implemented web application; Section 6 reports numerical results for a case study; Section 7 summarizes the work and provides the reader with some comments on possible future studies.

2 STATE OF THE ART ON M2M TRAFFIC CHARACTERIZATION

Characterizing M2M traffic is tantamount to defining a stochastic process that models the traffic generated by communicating machines. M2M traffic models can be grouped in:

- *source traffic*, in which every traffic stream is separately analyzed: this model is precise and flexible but has a high complexity;
- *aggregated traffic*, in which machines of the same kind are treated as a single source: this approach is less precise and flexible but has a lower complexity, independent of the number of devices. Aggregated analysis is a convenient approach to perform large scale analysis; however, it fails to catch real features of M2M traffic, such as mutual dependencies between machines.

Aggregated traffic models were proposed in [10, 13], where the authors dealt with homogeneous traffic, and in [1, 9], where a non-homogeneous traffic was studied. [13] proposed a contention based access mechanism for M2M communications in the LTE network. The authors focused on the uplink nature of M2M traffic and studied innovative scheduling techniques, taking advantage of dedicated uplink signalling channels. [10] also studied the performance of machine-type-communications over the LTE network but mainly focusing on the physical layers: the authors provided a detailed classification of M2M applications with average transmission times, average message sizes and data rates. It was observed that LTE is well-suited to M2M communications since it provides extensive coverage with large capacities. On the other hand, in [9], a performance analysis of M2M communications over GSM was carried out. A Poisson-distribution was used to model asynchronous traffic, whereas a Beta-distribution was used for synchronous traffic. The main contribution of the work was the study of the combination of synchronous and asynchronous traffic. The main inconvenient is the use of the old GSM network but this is a minor drawback because the analysis can be easily extended to the LTE contention mechanism and also because GSM is still used by a large number of machines. A similar model was proposed by the 3GPP group in [1] where traffic was divided into two categories: *Model 1*, dealing with uncoordinated traffic and *Model 2*, treating synchronous devices. As in [9], uniform and Beta distributions were respectively used in the two aforementioned models.

The authors studied the performance of the two different traffics over LTE and UMTS. The main limits of [1] are that: (i) the transmission of simultaneous packets (burst) was not considered, (ii) the two models were studied in a standalone fashion, without evaluating their mutual combination, and (iii) a large amount of machines could not be considered due to the complexity of the model.

An alternative approach was proposed in [3], where a source traffic is adopted and a Markov Modulated Poisson Process (MMPP) is used to model the traffic generated by each machine. The solution proposed by [3] consists in a basic modelling of a single machine and in a background *master process* that modulates all the MMPPs: this allows to perform analyses with a large number of nodes with low computational times. The problem with the use of a MMPP for a M2M system is that it cannot represent the coupling among devices.

A completely different approach to model M2M traffic was followed in [11], in which the authors identified the Type Allocation Code (TAC) associated to 150 particular machines and used it to retrieve M2M traffic within a large volume of mobile traffic, provided by an anonymous US mobile operator. The importance of this work is related to the use of real M2M traffic data. However, only few types of machines were identified, restricting the set of M2M applications included in the study. Moreover, it is not representative of future M2M applications, which are not implemented yet.

Throughout the literature, we noticed some gaps that need to be filled in order to keep pace with the expected growth of M2M communications.

Differently from the literature, in this paper we propose to (i) employ real data on the position of the machines, (ii) use several probability distributions in order to better represents realistic M2M applications, (iii) combine the analysis of multiple M2M applications using the same communication network. Moreover, the source-traffic approach is adopted: each machine is considered as a single source and can produce traffic according to the types of applications it is associated to.

3 DATA COLLECTION: MACHINES IN MONTREAL

We decided to build a case study for our analysis in Montreal, the second largest city in Canada, recently recognized as one of the *top intelligent communities*[1] in the world by the Intelligent Community Forum (ICF). In Figure 1 a map of Montreal is reported highlighting the geographical area we used. The large datasets we were able to build using data on

Figure 1: A map of the area which is in scope for the M2M traffic characterization project.

several machine types are summarized in Table 1. In what follows, each machine type is separately described:

- **Houses:** machines are located in residential premises; suffice it to think of smart meters and all the smart grid applications that use this type of devices.
- **Bus stops:** public transportation is one of the sectors that is most involved in M2M communications. A large number of applications rely on transit data: as a consequence, a new standard GTFS[2] was implemented by Google. Public transportation data were found in the website of STM, the local public transportation company. Further details can be found in [12].
- **Traffic lights:** they can be used in some M2M applications, mainly concerning traffic monitoring. Data about the position of traffic lights in the island of Montreal were found in the Portail Données Ouvertes and are available at [6].
- **Traffic cameras:** they are installed for traffic monitoring and public safety. Their position was retrieved from [5].
- **Parking spots:** information on these is used for smart parking, which is one of the most relevant applications in the domain of IoT, particularly of interest within the smart city context. A data-set of available parking spots in Montreal was retrieved in [7].

All the geographic data were structured using *GeoJson*[3], a geospatial data interchange format proposed by Internet Engineering Task Force (IETF) and based on JavaScript Object Notation (JSON) file extension.

[1] http://www.intelligentcommunity.org/intelligent_community_forum_names_montreal_quebec_canada_the_2016_intelligent_community_of_the_year

[2] Additional information at https://developers.google.com/transit/gtfs/.

[3] For more information, visit https://tools.ietf.org/html/rfc7946.

Table 1: The built database of the geographical position of machines in Montreal.

Type	Number	Type of Applications	Source
Houses	335158	Smart metering	[4]
Bus stop	8965	Smart public transportation	[12]
Traffic lights	1804	Congestion monitoring	[6]
Traffic cameras	430	Public safety	[5]
Traffic signs	313731	Advertising	[8]
Parking spots	18769	Smart parking	[7]

4 TRAFFIC GENERATION

A M2M traffic generator is implemented to represent the packet generation according to (i) the desired set of M2M applications, (ii) the set of machines, and (iii) the geographical area under study (including the adopted area subdivision, as will be shown in Section 5). The traffic generator, implemented in Java, uses the three aforementioned set of parameters (generated through the web application described in Section 5) to determine the traffic (i.e., number of packets, data volume) produced during the total simulation time. The traffic generation is composed by a set of probability distributions (described in Section 4) and a set of possible M2M applications (presented in Section 4).

Probability distributions

In order to represent the packet generation for different types of M2M applications, three types of probability distributions are used:

- **Poisson** This memory-less distribution is particularly used to describe events for which the packet generation of one machine has no impact on the others. λ_i represents the mean packet generation rate of machine i and B_i the packet size.
- **Beta** This well known distribution is widely used in 3GPP specs (e.g., in [1]), to account for synchronized M2M traffic (e.g., the reaction of a set of machines to an alarm or to a request to transmit data). A parameter W is used to represent the maximum time after which all the nodes have transmitted.
- **Deterministic** This is used when packet transmission takes place according to scheduled events (e.g., the transmission of packets from a bus stop whenever a bus arrives in it). An uncertainty parameter U is used so that a packet is transmitted within an interval centered on the scheduled time T_i, i.e. $(T_i - U, T_i + U)$. The default value is $U = 30$ s.

Although results with Beta-distribution are not presented in this paper, its implementation is already available in the tool.

M2M applications

In our analysis, we included a set of realistic M2M applications, each of them with different traffic characterization properties. This is important to gain meaningful insights into the peculiarities of M2M traffic.

In what follows, we describe and define the M2M applications included in this study.

- **Smart metering** Smart metering applications consists in the transmission of very small packets from the smart meters to a metering data management system. The packet generation time may vary from 15 minutes to several hours. In our case study we assumed $\lambda_i = 5.5610^{-4}\ s^{-1}$ (i.e., one packet each 30 minutes) and $B_i = 150$ Bytes.
- **Traffic monitoring** Traffic monitoring is achieved by collecting data from sensors installed in the traffic lights throughout the city. A Poisson-distributed packet generation is assumed with packet generation rate $\lambda_i = 0.01667$ (i.e., one packet each minute) and average packet size $B_i = 500$ Bytes.
- **Public transportation** The objective of this application is to provide users waiting at a bus stop with real-time forecasts on bus arrival times. In order to do that, each bus stop is expected to send a packet when a bus arrives. The information is centrally processed by the public transportation provider who spread the updated time of arrivals throughout its network. Packet generation rate is deterministic[4]. Two different profiles are used: one for working days and one for weekends and holidays.
- **Smart parking** In the smart parking applications, sensors are expected to be installed on parking lots. They can be used to manage payments, to check special permits for residents, or to issue fines for inattentive drivers. The absence of real data on car arrivals at

[4]Data on the time of arrivals was estimated with the nominal one, retrieved from http://www.stm.info/en/about/developers

Table 2: Characteristics of the M2M applications considered in this work.

M2M application	Type of nodes	Distribution	$1/\lambda_i$ (minutes)	Packet size (Bytes)
Smart metering	Houses	Poisson	30	150
Traffic monitoring	Traffic lights	Poisson	1	500
Public transportation	Bus stops	Deterministic	n.a.	200
Smart parking	Parking spots	Poisson	30	200
Public safety	Traffic cameras	Poisson	5	2000

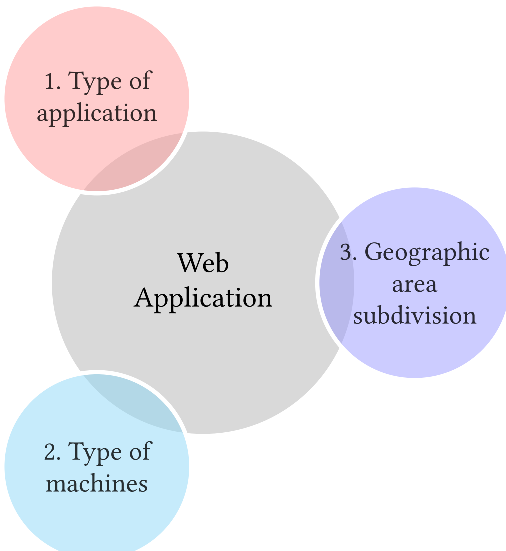


Figure 2: A diagram of the different components of the web application.

parking lots entailed the use of a Poisson distribution. A mean packet generation time of 30 minutes is used with 200 Bytes as average packet size.

Public safety Traffic cameras are used to periodically transmit images to a central management server. This images can be used, in combination with a facial recognition tool, by police to ensure public safety. A Poisson distribution is assumed with $\lambda_i = 0.0667$ (i.e., one packet each 5 minutes) and $B_i = 2000$ Bytes. Table 2 collects information on M2M traffic for these applications. It is important to notice that each and every machine in the simulation produces its own particular traffic at each time step. For example, distinct traffic lights produce traffic with distinct Poisson distribution, with the same average transmission rate and packet size. This is what makes our traffic model source-based.

5 THE WEB APPLICATION

A web application was implemented to allow a number of different studies on M2M communications. It allows to access through a GUI our database on the geographic position of nodes (described in Section 3) and to simulate traffic generated by different M2M applications (described in Section 4). Depending on the type of analysis sought by the user, different geographic area subdivisions are also available and are the subject of this section. The three components of the web application are displayed in Figure 2.

Four different modes of operation were implemented: (i) neighborhoods, (ii) circles, (iii) hexagonal cells, and (iv) LTE network cell.

In the first mode of operation the area is subdivided into neighborhoods: information about neighborhoods (e.g., number of machines, boundaries, names, IDs) are precomputed and included in a GeoJson file. This mode permits to identify the total number of machines in each neighborhood and can result particularly useful for administrative purposes.

When the second mode of operation is selected, the analysis is restricted to a circle in the map. The user can decide where to place the center of the circle by simply clicking on the map and inserting the desired radius in a dialogue box. The user can select the desired set of machines and click on the UPDATE button: a pop-up window will show the number of those machines.

In the third mode of operation, the geographic area at hand is covered with a grid of hexagonal cells, as displayed in Figure 3. The use of hexagonal cells is driven by the fact that a cellular network is planned to be used as transmission support of M2M packets. Each hexagon represents the area covered by a base station. The user can select a grid among a set of previously computed grids (through a drop-down menu on the left) or create a new hexagonal grid.

The fourth mode of operation can be used to evaluate the performance of M2M traffic in a realistic LTE infrastructure, which was designed according to the exact location of NodeB and eNodeB in Canada. The complete list is publicly available in [2]; we selected the base stations located in Montreal and relative to the LTE technology. We have also subdivided the base stations according to mobile operators, which can

Figure 3: A screenshot of the *hexagonal cell* mode of operation. The displayed grid is composed of hexagons of 2000 meter size.

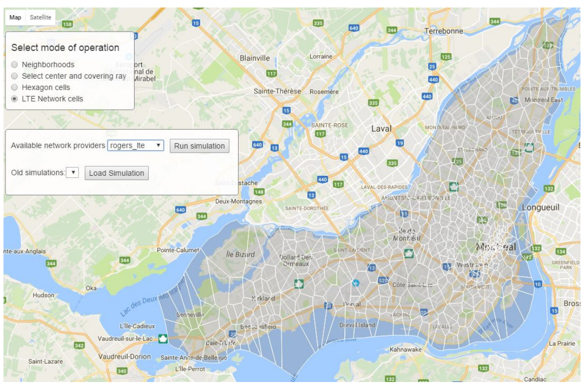


Figure 4: A sample partitioning of Montreal according to the location of the 544 eNodeB owned by the operator Rogers.

be selected using a drop-down menu on the left. Then, the Montreal area is partitioned according to a Voronoi diagram so that each zone contains one base station. All geographic points are included in the zone of their closest base stations. The result of this partitioning for one of the operators is shown in Figure 4.

6 PRELIMINARY RESULTS

In what follows, the LTE network mode of operation is adopted. The LTE infrastructure shown in Figure 4 in used in our case study. Random traffic was generated using the M2M applications defined in Table 2 over a 24 hour time horizon.

Two performance indexes are computed for the whole island and for each LTE cell and are discussed in Sections 6

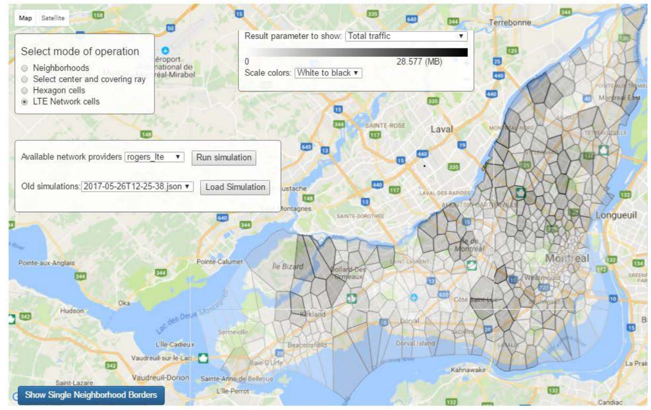


Figure 5: Heat-map of the total traffic generated using the scenario defined in Table 2.

and 6: (i) the data volume and (ii) the total number of packets.

Data volume

Data volume is important to compute the aggregated throughput and to study congestion in the backbone of the network (e.g., in the Evolved Packet Core (EPC) in LTE). In figure 5 we reported a heat-map of the total traffic generated in the LTE network cells. The total data generated over the 24 hour simulation time is 3.996 GB, whereas the maximum traffic produced in an LTE network cell is 28.577 MB.

In Figure 7a, a pie chart with the percentage of data volume generated by each application is reported. As one can see, the largest portion of data volume is produced by smart metering, followed by traffic monitoring with 30.61%. The three remaining applications only accounts for slightly more than 12% of the whole volume of data generated in the simulation.

Number of packets

The number of transmitted packets represents the number of attempts of machine on the cellular network. This index is very important because it is expected to primarily affect the cellular access network. M2M communications will most probably consist in a high number of devices transmitting small packets; the access segment is the one that will mostly suffer by their massive introduction.

In Figure 6a, a heat-map of the total number of packets generated in 24 hours. The association between the colour of the LTE network cells and the number of transmitted packets is expressed through the colour-bar on the top of the aforementioned figure. In this figure, we can see one of the main differences between human traffic and machine traffic. In the downtown area, LTE cells are very small in order to serve to large amount of people connected to the network. However,

(a) Global

(b) Smart metering

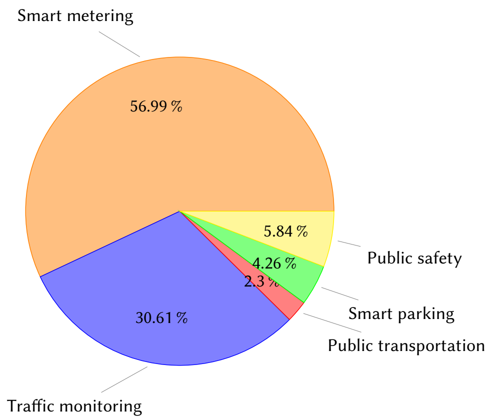
(c) Traffic monitoring

(d) Public Transportation

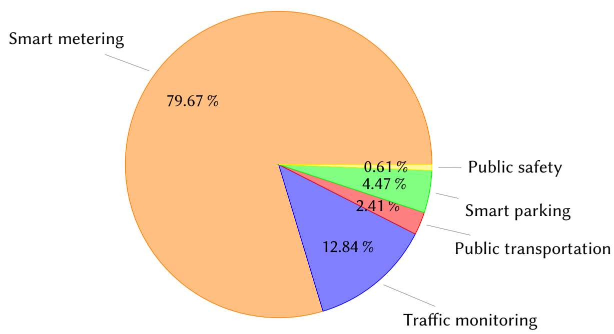
(e) Smart Parking

(f) Public safety

Figure 6: Total number of generated packets during the 24-hour simulation.



(a) Data volume



(b) Number of packets

Figure 7: Pie charts of the percentage of (a) number of packets and (b) data volume of the considered M2M applications.

the figure shows how these cells are only minimally affected by machine traffic since they are less extended, therefore containing a lower number of communicating machines. A total number of 20209615 packets were generated over 24 hours. The maximum number of packets generated in any LTE network cell is 167732.

In Figures 6b-6f, separate heat-maps are reported for the number of generated packets per each M2M application. As one can see, the heat-map related to smart metering (i.e., Figure 6b) is very similar to the global heat-map, reported in Figure 6a. It is possible to notice it by looking at the disposition of colours throughout the two figures, and at the maximum number of packets generated in a LTE cell (i.e., 167570 in the global figure and 162547 in the one related to smart metering). Traffic monitoring, reported in Figure 6c, and public transportation, reported in Figure 6d, generates number of packets that are similarly distributed throughout the island of Montreal, implying a correlation between the position of traffic lights and the position of bus stops. However, traffic monitoring produces an higher amount of packets, provided the higher transmission rate indicated in Table 2. Smart parking, displayed in Figure 6e, and public safety, reported in Figure 6f, appears to generate more packets in the downtown area, identified by smaller LTE cells.

The difference between the amount of packets generated by each application can also be observed in the pie chart reported in Figure 7b. As already pointed out, smart metering is the M2M application that produce the highest number of packets (79.67%), followed by traffic monitoring with 12.84%. All the other applications accounts for the 7.59% of the total number of packets generated in the simulation.

7 CONCLUSIONS AND FUTURE WORKS

This work presented a thorough traffic characterization for M2M communications. A web-based framework was implemented to represent the traffic generated by a wide variety of realistic M2M applications.

Real geographic data about the position of machines were adopted in order to have meaningful results and to better shape the M2M traffic.

Three probability distributions were employed to model the packet generation of different M2M applications. The Poisson and Beta distributions have already been used in 3GPP specifications; additionally, a deterministic distribution was adopted in order to represent M2M applications whose packet generation follows specific time patterns.

The presented web application framework can be used to study the suitability of different communication infrastructures to support M2M traffic.

A case study was presented using publicly available data on Montreal, one of the top intelligent communities according to the ICF. The analysis presented was focused on LTE, which is one of the best available solutions, but can be easily extended to other infrastructures (e.g., LoRa, 3G, WiFi). LTE infrastructure was represented with fidelity through the use of the exact position of eNodeB stations in the area under study. Numerical results on the data volume and on the total number of packets were presented and they shows some important differences between M2M and human traffic and their impact on the LTE infrastructure.

ACKNOWLEDGMENTS

This project was funded by Ericsson within the "2015 Request For Proposal". The authors would also like to thank Konstantin Fedorov, Dominik Courcelles, Laurent Olivier Chiquette, and Louis-Philippe Lafontaine-Bédard for their valuable support in the development of the web application presented in this paper, and Eleonora Botta for helping in revising the paper.

REFERENCES

[1] 3GPP. 2011. *Study on RAN Improvements for Machine-type Communications (Release 11) - TR 37.868.* Report.

[2] Science Innovation and Economic Development Canada. 2016. List of base stations installed and active in Canada. http://www.ic.gc.ca/engineering/SMS_TAFL_Files/Site_Data_Extract_2017-05-01.zip. (2016).

[3] Markus Laner, Philipp Svoboda, Navid Nikaein, and Markus Rupp. 2013. Traffic Models for Machine Type Communications. In *Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium on.* 1–5.

[4] Portail Donnï£¡es Ouvertes Montreal. 2016. List of residential addresses. http://donnees.ville.montreal.qc.ca/dataset/adresses-ponctuelles. (2016).

[5] Portail Donnï£¡es Ouvertes Montreal. 2016. List of traffic cameras. http://donnees.ville.montreal.qc.ca/dataset/cameras-observation-routiere. (2016).

[6] Portail Donnï£¡es Ouvertes Montreal. 2016. List of traffic lights. http://donnees.ville.montreal.qc.ca/dataset/feux-tous. (2016).

[7] Portail Donnï£¡es Ouvertes Montreal. 2017. List of public parking spots. http://donnees.ville.montreal.qc.ca/dataset/stationnements-municipaux-tarifes-sur-rue-et-hors-rue. (2017).

[8] Portail Donnï£¡es Ouvertes Montreal. 2017. List of traffic signs. http://donnees.ville.montreal.qc.ca/dataset/stationnement-sur-rue-signalisation-courant. (2017).

[9] Rafael CD Paiva, Robson D Vieira, and M Saily. 2011. Random access capacity evaluation with synchronized MTC users over wireless networks. In *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd.* IEEE, 1–5.

[10] R. Ratasuk, J. Tan, and A. Ghosh. 2012. Coverage and Capacity Analysis for Machine Type Communications in LTE. In *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th.* 1–5. https://doi.org/10.1109/VETECS.2012.6240186

[11] Muhammad Zubair Shafiq, Lusheng Ji, Alex X Liu, Jeffrey Pang, and Jia Wang. 2012. A first look at cellular machine-to-machine traffic: large scale measurement and characterization. In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 40. ACM, 65–76.

[12] STM. 2017. Montreal public transportation data. http://www.stm.info/en/about/developers. (2017).

[13] Kaijie Zhou, Navid Nikaein, Raymond Knopp, and Christian Bonnet. 2012. Contention based access for machine-type communications over LTE. In *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th.* IEEE, 1–5.