Connected cars in cellular network: A measurement study

Carlos E. Andrade, Simon D. Byers, Vijay Gopalakrishnan, Emir Halepovic,
David J. Poole, Lien K. Tran, Christopher T. Volinsky
AT&T Labs - Research
{cea,byers,gvijay,emir,poole,lien,volinsky}@research.att.com

ABSTRACT

Connected cars are a rapidly growing segment of Internet-of-Things (IoT). While they already use cellular networks to support emergency response, in-car WiFi hotspots and infotainment, there is also a push towards updating their firmware Over-The-Air (FOTA). With millions of connected cars expected to be deployed over the next several years, and more importantly persist in the network for a long time, it is important to understand their behavior, usage patterns, and impact — both in terms of their experience, as well as other users. Using one million connected cars on a production cellular network, we conduct network-scale measurements of over one billion radio connections to understand various aspects including their spatial and temporal connectivity patterns, the network conditions they face, use and handovers across various radio frequencies and mobility patterns. Our measurement study reveals that connected cars have distinct sets of characteristics, including those similar to regular smartphones (e.g. overall diurnal pattern), those similar to IoT devices (e.g. mostly short network sessions), but also some that belong to neither type (e.g. high mobility). These insights are invaluable in understanding and modeling connected cars in a cellular network and in designing strategies to manage their data demand.

1. INTRODUCTION

The rapid adoption of connected devices is fueling a growth in Internet of Things (IoT). Since the communication characteristics of typical IoT devices are different from traditional cellular devices, there is a widespread expectation that they will have an impact on cellular networks. In particular, the common wisdom is that the signaling load of IoT devices differs significantly and will motivate a different approach to managing IoT device connectivity.

In this paper, we focus on connected cars, a rapidly growing segment of IoT that defy this wisdom. Connected cars today use cellular networks to support emergency response, in-car WiFi hotspot, infotainment, and other forms of vehicular communication. They also use the network to convey telemetry information and manufacturers are seriously considering cellular networks to push updates to firmware over-the-air (FOTA) [10]. Forecasts predict that 90% of cars will be connected by the year 2020 [16].

The different types of use cases supported by connected cars result in unique communication patterns. Connected cars differ significantly from traditional IoT in terms of data volumes they generate given the infotainment capabilities they support and because of the large volume FOTA downloads (updates ranging from Megabytes to even Gigabytes are not unheard off [7]). At the same time, the fact that they show up in the network periodically, and are almost always mobile (and traveling at high-speed) makes them different from mobile phones. Next, given that the average life of modern cars is over 11 years [6] and rising, connected cars that come out today need to be supported for a lot longer than typical cellular phones (life expectancy of 4.7 years [3]). Finally, many of the updates to cars tend to be time critical given the types of features that are controlled by software as well as the safety and regulatory implications of these updates [13]. Managing large volume downloads, at high speeds, and supporting devices that are typically considered legacy is going to require innovative network planning and management strategies. It may necessitate the use of smart policies and network control mechanisms for the management of network demand, especially at peak hours of demand.

To design the right kinds of policies and mechanisms, it is necessary to understand connected car behavior and the potential impact that they may have on the cellular network. In this paper, we conduct a large-scale measurement study of connected cars in a major cellular network to provide insights and basis for studying and modeling connected car impact in such environments. Using anonymized call detail records for a random set of one million connected cars from one manufacturer over the 90-day period, we seek to understand various aspects of connected cars, including the spatial and temporal distribution of cars' connections to the network, the mobility patterns of connected cars, the typical network conditions that connected cars face, and how are cars' connections and handovers distributed over various radio frequencies.

Based on our analysis, we make several important observations. Connected cars have a very different connectivity pattern than regular smartphones. In particular, cars' sessions are shorter and in many cases concentrated in network busy hours. This is not entirely surprising given that they can connect only when their engines are running, and commute happens during busy hours. However, this has implications when it comes to delivering software updates in a timely manner. We also observe that connections to each radio cell are generally short, which is a combination of mobility (handovers) and small data transfers. Furthermore, cars can concentrate in large numbers in cells triggering the potential for congestion. Most importantly, we observe that cars can be clustered according to predictability in their behavior. This indicates a potential for intelligent capacity and network management in terms of connectivity and content delivery for connected cars.

2. RELATED WORK

Adoption of IoT is imminent and projected connected "things" far outnumber smartphones and computers, with connected cars becoming a significant segment [16]. With widespread geographic distribution of various devices, cellular networks are expected to bear majority of IoT traffic demand. In fact, the 3GPP report projects on the order of 200,000 IoT devices per cell site, as opposed to the current several hundred or thousand [1].

Early large study finds that IoT devices over all exhibit different characteristics from the mainstream cellular devices, such as smartphones, in particular that the ratio of their uplink to downlink traffic is much larger, their mobility is much lower, and the diurnal traffic pattern is different [14]. However, early indications with the increasing, but currently limited numbers of connected cars, is that their characteristics are significantly different. Hence they deserve a dedicated study like this one to understand their impact on the network.

IoT devices are expected to primarily introduce high signaling load into cellular networks, but a subset will add a large data volume. When connected cars are considered, from a smaller sample of 2,100 connected cars, we know that signaling intensity they generate can be 4-7 times higher than regular LTE devices [2], while the average flow sizes in both uplink and downlink are similar. In the future, connected cars are expected to introduce two types of high data demand: user traffic, such as web browsing and multimedia via in-car WiFi hotspot, and FOTA updates [5, 10]. Some work already exists in the area of managing FOTA, mostly related to efficient compression of updates [9, 15]. Specifications and protocols for remote device management are defined by the Open Mobile Alliance Device Management (OMA DM) [12]. While FOTA OEM can use OMA DM to manage devices, it is limited to protocols for data exchange, data formats, security and fault management. Understanding of how cars behave, what is the impact on the network and when to deliver their particular downloads is still heavily dependent on the network context.

General car mobility insights and traffic patterns have been inferred from cellular phone connections in the past, indicating commute patterns and network trajectories [4]. We aim to expand knowledge by directly studying connected cars as opposed to inferring their behavior from user devices like phones.

3. DATA SET AND METHODOLOGY

At the high level, we use network measurements from a large cellular network in the United States to understand the network behavior of connected cars. User devices, such as smartphones, tablets or modem cards connect to a radio cell (or simply a "radio" or a "cell") over a certain radio frequency or a carrier. Each cell covers a geographic area with a directional antenna and it is common to find 3 such cells covering a full circle, approximately 120 degrees each, but there can be more or fewer cells with different coverage areas. For coverage and capacity, there are multiple carriers used on a single base station, organized in the same geographic coverage. Multiple cells covering the same area can be called a sector. There could be anywhere from 3 to 12, sometimes even more cells per base station. There can be hundreds of thousands of cells in the networks and the cars connect to a subset of these cells.

Our data, based on Call Detail Records (CDRs) provides information about radio-level connections made by cars to the cellular network, such as times and durations of connections, as well as radio cells that they connect to. These records are anonymized and aggregated and do not contain sensitive personal or identifiable information about owners of devices or connected cars. The data set consists of over 1.1 billion connections from a random sample of 1 million cars equipped with cellular 3G/4G modems. These provide connectivity to support emergency services, telemetry, FOTA updates and in-car WiFi hotspot. We constrain the data set to a single car maker, also known as Original Equipment Manufacturer (OEM). A single OEM with a large car population allows us to reason about potential FOTA management, since this is managed by each OEM independently.

Our study spans a 90-day period in 2017. We believe that this period is long enough to be representative as a predictor, and to account for variability in daily and hourly load as well as any trend. There can be a vast range of connection durations at radio level due to the normal timeout of 10 to 12 seconds after no data is left to transmit in either direction [8]. We concatenate all

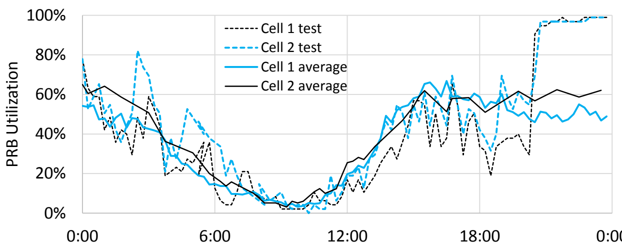


Figure 1: Large downloads start at 20:45 UTC in two cells and last for 4 hours, using all available resources.

connections that are up to 5 seconds apart into sessions where appropriate. Normally, the cars from this OEM can connect to the network only when the engine is running, so connections correlate to car usage and driving. We then derive patterns of connections across time and space (in terms of network location/cell) and analyze them against general car and network usage patterns.

We pre-process the logs to remove erroneous records, such as the ones where connections appear to have lasted exactly 1 hour. These are presumably caused by an automatic periodic reporting feature of the network, where disconnections at the radio level were not recorded correctly. Then, during the data analysis, we also truncate long connections to a single cell to 600 seconds, to mitigate some modems tendency to improperly disconnect.

4. CONNECTED CAR ANALYSIS

To further motivate the study for the reason of potential impact on the network and other users, we show that a single device can saturate the radio cell resources. In LTE, radio resources are finite and measured using Physical Resource Block (PRB) utilization, U_{PRB}. Even a single device can fully saturate a cell with a greedy long download, which is a normal behavior, as shown in Figure 1. This is typically not a problem, but it could be if many such downloads occur concurrently, as could happen during FOTA updates or multiple video users when cars are concentrated in one cell.

We start with high-level data set analysis showing the percentage of unique cars that appear on the network and the percentage of cells that cars connected to, for each day in Figure 2. Due to some data loss during 3 days in the second half of the study period, the number of cars appears smaller, but this does not affect the overall results. We clearly see a weekly pattern in both plots, with fewer cars and cells connecting on the weekend, but with most variability occuring on Friday and Saturday. We also show the trend lines, which indicate a slow increase over time. We can model the average numbers of cars and connected cells per Table 1.

4.1 Macro-level temporal behavior

We analyze the temporal behavior of cars by considering how long and when the cars were connected to the network during the study period. We plot two distributions, one with full session time as indicated by the CDRs, and another where sessions are truncated to 600 seconds in Figure 3. The CDF shows the fraction of cars vs. the percentage of study period that they were connected to the network.

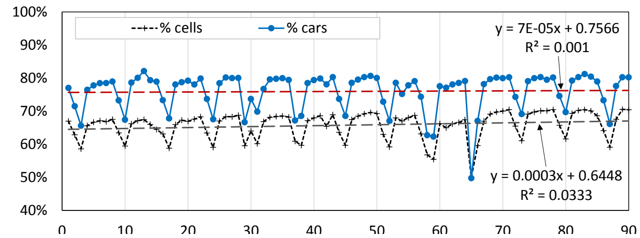


Figure 2: Number of cars that appear on network is relatively consistent over the days throughout the study.

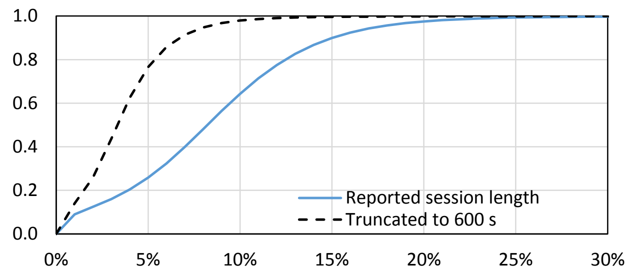


Figure 3: Cars' total time on network is very short.

Averages are about 8% for full and 4% for truncated sessions, respectively. This is about 173 and 86 hours total, or 1.9 and 1 hours per day, respectively. Without truncation the 99.5th percentile of connected time is 27% (6.5 hours per day), and with truncation it is 15% (3.6 hours per day).

Clearly, cars spend much less time connected than smartphones, meaning that the window of opportunity to deliver large amounts of data is very small.

4.2 Weekly and daily temporal behavior

We next analyze car usage patterns over the known weekly and daily network load and commute time periods. The network load follow the known diurnal pattern and peak commute time patterns can be extracted either from CDRs or from known data [4]. Hence we encode important time ranges during the weekly cycles in 24×7 matrices, where each hour of the day for 7 days is represented by a shaded box. Figure 4 shows example

Table 1: Percentage of cars and cells per day.

	% cells with cars		% cars on network	
Day	Mean	StDev	Mean	StDev
Monday	67.2%	1.1%	78.1%	0.8%
Tuesday	68.1%	1.6%	79.1%	1.5%
Wednesday	68.5%	1.4%	79.8%	1.2%
Thursday	68.2%	1.7%	79.3%	0.9%
Friday	67.2%	3.1%	78.0%	3.8%
Saturday	62.0%	4.3%	70.3%	7.0%
Sunday	59.3%	1.5%	67.4%	2.0%
Overall	65.8%	4.1%	76.0%	5.6%

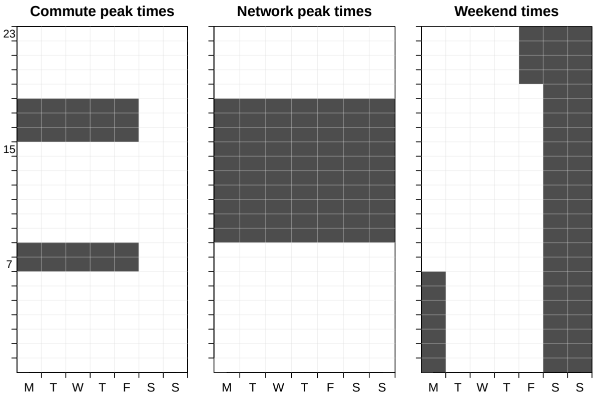


Figure 4: Significant time ranges in the week.

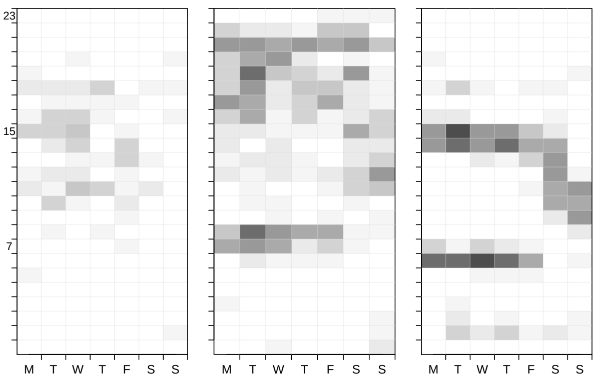


Figure 5: Usage patterns from 3 sample cars

times of interest visualized in local time.

Similarly, we can encode car usage patterns in the same format. Figure 5 shows the 24×7 car usage frequency matrices rendered in respective local times for 3 real cars. Darker colors represent a higher number of car's connections to the cellular network. A white box means that the car has not connected to the network during that hour. Frequency of connections and their regularity over certain hours allow us to see strong patterns in the very dark colors. By aggregating data from multiple weeks onto a 24×7 matrix we can take this hourly and daily pattern into account and find the consistent patterns in the noise.

For these 3 cars, the weekly matrices tell us the following. Car on the left rarely connects to the network, mostly during Monday to Friday network busy hours (14-24 h). Car in the middle exhibits more heavy usage and consistent Monday to Friday commute, stretching into the evenings and in addition has moderate weekend usage. Usage on Tuesdays is much more frequent and spread over multiple hours than on other days. Car on the right shows a very strong and consistent commute, though before start of peak commute in its timezone. Further it shows quite predictable weekend usage during peak network on Saturday and early Sunday morning.

Predictable appearance of cars during busy or non-busy network hours allows for more intelligent management of large data demand.

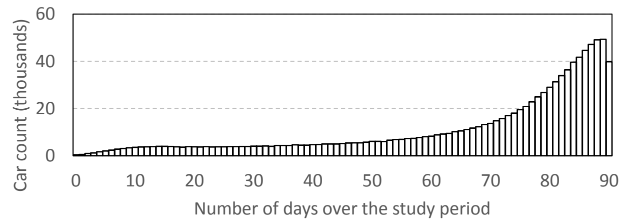


Figure 6: Number of days cars were on the network.

Table 2: Car segmentation.

Segment	Busy	Non-Busy	Both	Total
Rare (≤ 10 days)	0.4%	0.9%	0.9%	2.2%
Common (10+ days)	1.3%	59.0%	37.5%	97.8%
Rare (≤ 30 days)	0.7%	5.0%	4.2%	9.9%
Common (30+ days)	1.0%	54.9%	34.2%	90.1%

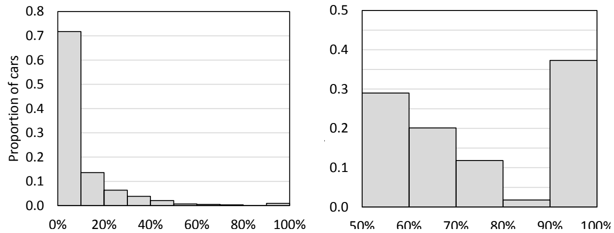
4.3 Combining car and network data

An example of combining the matrices for network busy times and car usage follows. We would like to know what proportion of cars are commonly seen or rare on the network and do they typically appear during network busy hours or not. This *segmentation* can guide various traffic management solutions.

To be able to define *common* and *rare* we need either some intuitive definition, a specific use case definition, or we can derive it from data. To derive from data, we can use the number of days over the study period that cars connected, as shown by the histogram in Figure 6. It appears that 10 days indicates the point under which a sharp drop off exists, and past 30 days is where increasing trend begins. If we use simple definitions that rare means 10 days or less in one use case and 30 in another, we can segment the car population as in Table 2. We consider a car to typically connect in busy or non-busy hour if 65% or more of its time on network is to cells whose average $U_{PRB} > 80\%$ for that 15 minute bin. Otherwise, cars appear more balanced in both busy and non-busy hours.

An example use of this type of segmentation is in the context of FOTA updates. In some managed FOTA scenario, rare cars would be prioritized over the limited FOTA campaign window, and common cars would be perhaps randomized or scheduled depending on the typical time they connect. In particular, cars that typically appear during busy hours will likely need special treatment to avoid impacting the network and other uses during large downloads.

While general temporal pattern of cars' connections offers important insights, we further seek to assess the potential impact by considering what are the *typical network conditions* that cars encounter on their journeys. In particular, understanding how much cars typically connect to busy cells could further refine the management decisions. For example, allowing a large FOTA download in an already loaded cell (e.g. $U_{PRB} > 80\%$) might be considered pouring oil onto the fire.



(a) Time spent in busy cells. (b) At least 50% of time spent in busy cells.

Figure 7: Network conditions that cars encounter

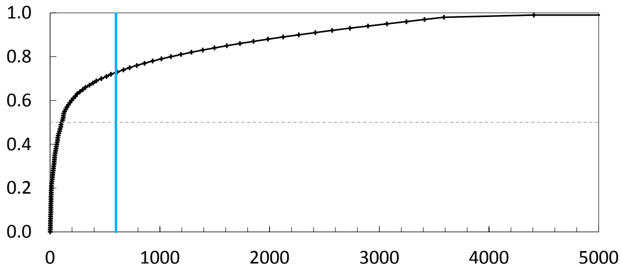


Figure 8: Duration of cars' connections per radio cell.

To asses this type of impact, we plot the deciles of time that each car spends connected to the busy radio cells in Figure 7. It turns out that cars do not spend most of their connected time in highly loaded cells. However, a small number of cars, about 2.4%, spend more than 50% of their connected time on busy radios, with about 1% or 10,000 cars spending all their time on busy radios.

Combining car behavior with network data give us direction towards types of policies to use for different segments of cars, depending whether they need FOTA update, regular user traffic, or infotainment services.

4.4 Temporal behavior at micro level

We now shift from the macro level of looking at the whole network to a micro level, where we can study the connected car behavior per radio cell and per car.

Since cars are expected to drive distances spanning multiple cells and base stations, we are interested in how long does each connection to a cell last. This will provide insights into the length of impact per cell. Figure 8 shows that cell sessions are generally short, with the median of 105 seconds and 73rd percentile at 600 seconds. The mean connection duration is 625 seconds for the full, and 238 seconds for the truncated sessions, respectively. However, a significant number of cell connections are very short.

Even with relatively short time spent on each cell, it is still possible to encounter high concentration of cars under the same cell. Intuitively, this would occur at highway traffic during commute times, at shopping malls, or event parking lots.

We show two examples of concurrent cars over one week in Figure 9. We declare cars concurrent if their connections straddle a 15-minute time bin of the day.

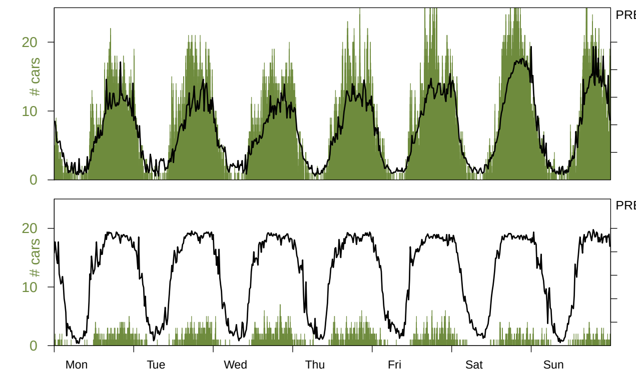


Figure 9: Concurrent cars on radios.

We select a longer time bin than a typical connection because we expect that the highest impact will come from large downloads that may both extend the cars' own connection time and other users connections, simply due to bandwidth sharing.

In these examples, number of concurrent cars follows the same diurnal pattern of the cell load, which is represented by the average U_{PRB} for each 15-minute time bin. In the top plot, we see a moderately loaded cell (solid line) that becomes more busy over the weekend, but consistently sees between 10 and 25 connected cars (impulses) during its busy hours. The bottom plot shows a different combination, a cell that is fully loaded for the most of the day, but sees a few connected cars. It is important to take into account that both scenarios can lead to undesirable consequences. For example, any number of large downloads added to the loaded cell may deteriorate experience for everyone, same as having 20 or more cars attempt overlapping downloads.

Figure 10 visualizes unique cars' sessions in more detail. There were 377 cars connected to this cell over 24 hours and their connections are shown as horizontal lines, one for each car. We can see several typical car behaviors: (i) connections are short, (ii) connections are rare overnight, (iii) significant concurrency exist regardless of each connection being short. The 15-minute time bin with most concurrent cars, 16, is marked.

Short connections suggest a potentially small amount of data that can be transferred through each cell prior to handover and point to judicious use of policies, e.g. on seamless vs. lossless handover, to mitigate data loss [11].

4.5 Spatial behavior

We next consider mobility of cars across the network. While cars' connections are expected to hand over across base stations, it turns out that the radio-level logs do not support such precise analysis. Since connected cars do not constantly send or receive, their connections timeout often. Therefore, cars often do not connect to every cell they traverse, unless there is an immediate request to transfer data. To asses a lower bound on number of cells and handovers, we account for

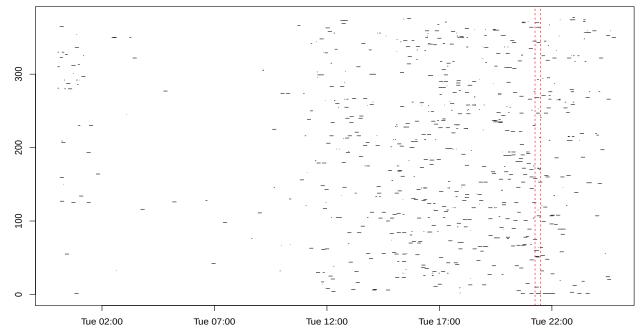


Figure 10: Concurrent cars in one cell over 24 hours.

handovers within sessions on the network during which the longest connection gap is 10 minutes.

We find that the most common handover is across base stations, which is the expected behavior. The median number of handovers is 2, 70th percentile is 4 and 90th percentile is 9. This suggests that for most large downloads by a connected car, the impact will span between 3 to 10 base stations. Other types of handovers are observed in negligible numbers, namely between radio technologies (3G/4G), between carriers of the same sector and between sectors of the same base station. This behavior might change once massive FOTA campaigns start. Similar implications apply as discussed for short connections per cell.

4.6 Frequency band usage

As cellular networks evolve, it is important to understand the capabilities of legacy devices. Connected cars stay "on the roads", meaning continue to be used, far longer than typical smartphones, often decades vs. years, respectively. As celular technology evolves, some connected cars will not be able to catch up using legacy hardware. We study the current capabilities by breaking down cars' network usage per frequency band or carrier. Higher frequency bands allow for wider bandwidth in carriers, which translates to higher data throughput.

The cars under study connect to the network using 5 observed carriers, which we name Ci, where $i = 1, 2, ..., 5$ and larger i indicates the higher frequency band. We first consider how many cars in total connected at least once to each carrier (Table 3). While this breakdown can be affected by availability of each carrier at particular base station that cars connect to, we actually confirm the expected behavior. Connected car modems of this OEM predominantly have the capability to use carriers C1-C4, and only a few C5 connections are registered.

We next asses the current use of carriers as it indicates the maximum achievable performance of the connected car population as a whole. Table 3 also shows the breakdown of total connection time spent on each carrier. The key finding is that higher frequency carriers C3 and C4 are used nearly 75% of the time.

Table 3: Carrier use of connected cars

Carrier	C1	C2	C3	C4	C5
Cars (%)	98.7%	89.2%	98.7%	80.8%	0.006%
Time(%)	18.6%	7.4%	51.9%	22.1%	0.000%

While higher data rates are in general available to cars most of the time, lack of support for the highest frequency clearly confirms the expectation that connected cars will need the legacy technology and carrier support.

4.7 Discussion

After analyzing spatial, temporal, and radio frequency usage of connected cars, we find that connected cars are a very complex type of device that is becoming mainstream in cellular networks. Specifically, we observe that cars actually have three sets of characteristics.

Similarities to smartphones include weekly and diurnal patterns of connecting to the network, high concurrency of connections across multiple cars, predictability in behavior. These are in addition to known ability to generate traffic similar to smartphone using a WiFi hotspot. These findings suggest that both the same treatment of cars as smartphones would work for some data services, but also that different management approaches may be needed for FOTA updates.

Similarities to IoT devices include limited carrier use capability, connecting to a subset of the network cells (most IoT devices are not mobile and connect to the same cell or base station), spend short time on the network overall and in each session. These imply the need for legacy support, that the overall impact of the population of cars may not extend to the whole network on a day-to-day basis, and that handover policies are important for overall efficient use of resources.

Connected car-specific traits include connecting to different cells on different days, having commute-time pattern or no pattern, not connecting to the network for extended periods of time, and inherent mobility. These characteristics call for possible per-car prediction models for efficient content delivery, and mobility management that will ensure efficiency and correct routing, while providing quality of experience in the face of frequent handoffs and high speed.

5. CONCLUSION

We conduct a measurement study of a large population of connected cars in a production cellular network. Using radio-level connection data we derive usage and mobility behavior of cars and obtain insights that enable modeling and analysis of their impact in cellular networks. We find that cars share characteristics of both smartphones and IoT devices, but also exhibit some specific traits. Most importantly, we find that it is possible to classify cars by how often they appear on the network and whether that would occur during busy or non-busy hours.

6. REFERENCES

[1] 3GPP. Cellular system support for ultra-low complexity and low throughput internet of things (CIoT), 3GPP TR 45.820. `http://www.3gpp.org/DynaReport/45-series.htm`, 2016.

[2] R. Archibald, D. Gupta, R. Jana, V. Gopalakrishnan, A. S. Rajan, K. B. Ramia, D. Dahle, J. Cooper, G. Kennedy, N. Rao, S. Sonnads, and M. M. Donald. An IoT control plane model and its impact analysis on a virtualized MME for connected cars. In *2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, pages 1–6, June 2016.

[3] C. T. Association. The Life Expectancy of Electronics. `https://www.cta.tech/News/Blog/Articles/2014/September/The-Life-Expectancy-of-Electronics.aspx`, September 2014.

[4] R. Caceres, R. A. Becker, J. M. Loh, C. Volinsky, S. Urbanek, K. Hanson, and A. Varshavsky. A Tale of One City: Using Cellular Network Data for Urban Planning. *IEEE Pervasive Computing*, 10:18–26, 2011.

[5] R. N. Charette. This Car Runs on Code, 2009.

[6] D. DeMuro. Buying a Car: How long can you expect a car to last? `http://www.autotrader.com/car-shopping/buying-a-car-how-long-can-you-expect-a-car-to-last-240725`, 2015 June.

[7] Ford. How to download SYNC software updates to a USB drive, accessed 2016.

[8] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. A Close Examination of Performance and Power Characteristics of 4G LTE Networks. In *Proc. of ACM MobiSys*, 2012.

[9] R. Kiyohara, S. Mii, M. Matsumoto, M. Numao, and S. Kurihara. A new method of fast compression of program code for ota updates in consumer devices. *IEEE Transactions on Consumer Electronics*, 55(2):812–817, May 2009.

[10] G. Nelson. Over-the-air updates on varied paths: Companies scramble to satisfy customers, capture savings, January 2016.

[11] B. Nguyen, A. Banerjee, V. Gopalakrishnan, S. Kasera, S. Lee, A. Shaikh, and J. Van der Merwe. Towards understanding tcp performance on lte/epc mobile networks. In *Proceedings of the 4th Workshop on All Things Cellular: Operations, Applications, & Challenges*, AllThingsCellular '14, pages 41–46, New York, NY, USA, 2014. ACM.

[12] Open Mobile Alliance Device Management. `http://openmobilealliance.org/about-oma/work-program/device-management/`, 2017.

[13] Popular Science. Software Now To Blame For 15 Percent Of Car Recalls, 2016.

[14] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. Large-scale measurement and characterization of cellular machine-to-machine traffic. *IEEE/ACM Transactions on Networking*, 21(6):1960–1973, Dec 2013.

[15] M. Steger, M. Karner, J. Hillebrand, W. Rom, C. Boano, and K. Rmer. Generic framework enabling secure and efficient automotive wireless sw updates. In *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8, Sept 2016.

[16] Telefonica. Connected Car Industry Report 2014. `https://iot.telefonica.com/multimedia-resources/connected-car-industry-report-2014-english`, Jul 2014.