Large Scale Measurement and Characterization of Cellular Machine-to-Machine Traffic

M. Zubair Shafiq, Lusheng Ji, Alex X. Liu*, Jeffrey Pang, and Jia Wang

***Abstract*—Cellular network based Machine-to-Machine (M2M) communication is fast becoming a market-changing force for a wide spectrum of businesses and applications such as telematics, smart metering, point-of-sale terminals, and home security and automation systems. In this paper, we aim to answer the following important question: Does traffic generated by M2M devices impose new requirements and challenges for cellular network design and management? To answer this question, we take a first look at the characteristics of M2M traffic and compare it with traditional smartphone traffic. We have conducted our measurement analysis using a week-long traffic trace collected from a tier-1 cellular network in the United States. We characterize M2M traffic from a wide range of perspectives, including temporal dynamics, device mobility, application usage, and network performance.**

Our experimental results show that M2M traffic exhibits significantly different patterns than smartphone traffic in multiple aspects. For instance, M2M devices have a much larger ratio of uplink to downlink traffic volume, their traffic typically exhibits different diurnal patterns, they are more likely to generate synchronized traffic resulting in bursty aggregate traffic volumes, and are less mobile compared to smartphones. On the other hand, we also find that M2M devices are generally competing with smartphones for network resources in co-located geographical regions. These and other findings suggest that better protocol design, more careful spectrum allocation, and modified pricing schemes may be needed to accommodate the rise of M2M devices.

I. Introduction

Smart devices that function without direct human intervention are rapidly becoming an integral part of our lives. Such devices are increasingly used in applications such as telehealth, shipping and logistics, utility and environmental monitoring, industrial automation, and asset tracking. Compared to traditional automation technologies, one major difference for this new generation of smart devices is how tightly they are coupled into larger scale service infrastructures. For example, in logistic operations, the locations of fleet vehicles can be tracked with Automatic Vehicle Location (AVL) devices such as the CalAmp LMU-2600 [6] and uploaded into back-end automatic dispatching and planning systems for real-time global fleet management. More and more emerging technologies also heavily depend on these smart devices. For instance, a corner stone for the Smart Grid Initiative is the capability of receiving and controlling individual customer's power usage on a real-time and wide-area basis through devices such as the electric meters equipped with Trilliant CellReader [24] modules.

This kind of leap in technology would not be possible without the support of wide area wireless communication infrastructure, in particular cellular data networks. It is estimated that there are already tens of millions of such smart devices connected to cellular networks world wide and within the next 3-5 years this number will grow to hundreds of millions [2], [3]. This represents a substantial growth opportunity for cellular operators as the increase in mobile phone penetration rate is flattening in the developed world [11], [25].

M2M devices and smartphones share the same network infrastructure, but current cellular data networks are primarily designed, engineered, and managed for smartphone usage. Given that the population of cellular M2M devices may soon eclipse that of smartphones, a logical question to ask is: What are the challenges that cellular network operators may face in trying to accommodate traffic from both smartphones and M2M devices? Existing configurations may not be optimized to support M2M devices. In addition, M2M devices may compete with smartphones and impose new demand on shared resources. Hence, to answer this question, it is crucial to understand M2M traffic patterns and how they are different from traditional smartphone traffic. The knowledge of traffic patterns can reveal insights for better management of shared network resources and ensuring best service quality for both types of devices.

In this paper, we take a first look at M2M traffic on a commercial cellular network. Our goal is to understand the characteristics of M2M traffic, in particular, whether and how they differ from those of smartphones. To the best of our knowledge, our study is the first to investigate the characteristics of traffic generated by M2M devices. We summarize our key contributions below.

• **Large Scale Measurement:** We conduct the first large scale measurement study of cellular M2M traffic. For our study, we have collected anonymized IP-level traffic traces from the core network of a tier-1 cellular network in the United States. This trace covers all states in the United States during one week in August 2010. This trace contains M2M traffic from millions of devices belonging to more than 150 hardware models. In addition, we have also collected anonymized traffic traces from millions of smartphones from the same cellular network. Overall, we find that M2M devices generate significantly less traffic compared to smartphones. Furthermore, in our trace, we observe that the number of M2M devices is also significantly smaller than the number of

The preliminary version of this paper titled "A First Look at Cellular Machine-to-Machine Traffic – Large Scale Measurement and Characterization" was published in the proceedings of the ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/Performance), London, UK, June, 2012.

∗ Alex X. Liu is the corresponding author of this paper.

M. Zubair Shafiq and Alex X. Liu are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA. e-mails:{shafiqmu,alexliu}@cse.msu.edu

Lusheng Ji, Jeffrey Pang, and Jia Wang are with AT&T Labs – Research, Florham Park, NJ, USA. e-mails:{lji,jeffpang,jiawang}@research.att.com

smartphones. However, the number of new M2M devices and their total traffic volume is increasing at a very rapid pace. In fact, a longitudinal comparison of M2M traffic in this cellular network showed that total M2M traffic volume has increased more than 250% in 2011 since previous year. In comparison, Cisco reported that mobile data traffic grew "only" 132% in 2011, which is almost half of the increase observed for M2M traffic [4]. Consequently, it is important to understand the peculiarities of M2M traffic, especially its contrast to the traditional smartphone traffic, for future network engineering. In this study, we compare M2M and smartphone traffic in the following aspects: aggregate volume, volume time series, sessions, mobility, applications, and network performance.

• **Aggregate Traffic Volume:** We jointly study the distributions of aggregate uplink and downlink traffic volume. Our major finding is that, though M2M devices do not generate as much traffic as smartphones, they have a much larger ratio of uplink to downlink traffic volume compared to smartphones. Since existing cellular data protocols support higher capacity in the downlink than the uplink, *our finding suggests that network operators need careful spectrum allocation and management to avoid contention between low volume, uplink-heavy M2M traffic and high volume, downlink-heavy smartphone traffic*.

• **Traffic Volume Time Series:** We analyze the traffic volume time series of M2M devices and smartphones. Our analysis shows that different M2M device models exhibit different diurnal behaviors than smartphones. However, some M2M device models do share similar peak hours as smartphones. Hence, *M2M traffic imposes new requirements on the shared network resources that need to be considered in capacity planning, where network is usually provisioned according to peak usage*. Another finding from time series analysis is that some M2M device models generate traffic in a synchronized fashion (like a botnet [23]), which can result in denial of service due to limited radio spectrum. Therefore, *M2M protocols should randomize such network usage to avoid congesting the radio network*.

• **Traffic Sessions:** To understand the usage behavior of individual devices, we conduct session-level traffic analysis in terms of active time, session length, and session inter-arrival time. We find that high traffic volume does not always correlate with more active time. *This finding calls for new billing schemes, which go beyond per-byte charging models*. We also find that M2M devices have different session length and inter-arrival time characteristics compared to smartphones. *This finding can be utilized by device manufacturers to improve battery management and by network operators to optimize radio network parameters for M2M devices*.

• **Device Mobility:** We compare the mobility characteristics of M2M devices and smartphones from both device and network perspectives. We find that M2M devices, with a few exceptions, are less mobile than smartphones. We also find that M2M and smartphone traffic competes for network resources in co-located geographical regions. *This finding indicates that careful network resource allocation is required to avoid contention between low-volume M2M traffic and high-volume smartphone traffic*.

• **Application Usage:** We also study the contribution of different applications to the aggregate traffic volume of M2M devices and smartphones. *We find that M2M traffic mostly uses custom application protocols for specific needs, which is undesirable because it is difficult for network operators to understand and mitigate adverse effects from these protocols compared to standard protocols*.

• **Network Performance:** The network performance results of M2M traffic, in terms of packet loss ratio and round trip time, show strong dependency on device radio technology (2G or 3G) and expected device environment (*e.g.* indoors vs. outdoors). *This implies that network operators will need to be cognizant of a large population of M2M devices on legacy networks even as they re-provision spectrum for 4G technologies to support newer smartphones*.

The rest of this paper proceeds as follows. We first provide details of our collected trace in Section II. Sections III–VIII present measurement analysis of M2M and smartphone traffic. Finally, we conclude in Section IX.

II. Data

A. Data Set

The data used in this study is collected from a nation-wide cellular network operator in the United States that provides 2G and 3G cellular data services. It supports GPRS, EDGE, UMTS, and HSPA technologies. Architecturally, the portion of its network that supports cellular data service is organized in two tiers. The lower tier, the radio access network, provides wireless connectivity to user devices, and the upper tier, the core network, interfaces the cellular data network with the Internet. More details about cellular data network architecture can be found in [21].

The data collection apparatus that produced the trace used in our study is deployed at all links between Serving Gateway Support Nodes (SGSN) and Gateway GRPS Support Nodes (GGSN) in the core network. This apparatus is capable of anonymously logging session level traffic information at 5 minute intervals for all IP data traffic between cellular devices and the Internet. In other words, each record in the trace is a 5-minute traffic volume (*i.e.*, TCP payload size in bytes) summary aggregated by unique device identifier and application category. Each record also contains the cell location of the device at the start of the session. Each record is originally timestamped according to the standard coordinated universal time (UTC), which is then converted to the local time at the device for our analysis. This trace was collected during one complete week in August 2010. Geographically, the trace covers the whole United States.

Applications are identified using a combination of port information, HTTP host and user-agent information, and other heuristics. Overall, traffic is classified into the following 17 categories: (1) `appstore`, (2) `jabber`, (3) `mms`, (4) `navigation`, (5) `email`, (6) `ftp`, (7) `gaming`, (8) `im`, (9) `miscellaneous`, (10) `optimization`, (11) `p2p`, (12) `apps`, (13) `streaming`, (14) `unknown`, (15) `voip`, (16) `vpn`, and (17) `web`. POP3 and IMAP traffic is classified as `email`. Additional control channel information is used to identify `voip` traffic. Most HTTP traffic is classified as

`streaming` or `web` based on mime type. Some heuristics are employed to identify non-HTTP `p2p` traffic. Gnutella and BitTorrent tracker-based HTTP traffic is also labeled `p2p`. User-agent information is used to identify specific mobile app traffic such as `appstore`. Port number information is used to identify other traffic classes. Many low volume applications are jointly labeled as `miscellaneous`. The remaining unclassified traffic is labeled as `unknown`. More information about application classification can be found in [10], [20].

B. M2M Device Categorization

The data set contains traffic records for all cellular devices, so we first need to separate M2M devices from the rest. Furthermore, because M2M devices are usually developed for specific applications, significant behavioral differences are expected between M2M devices for different target applications. Thus, it is reasonable to sub-divide M2M devices into categories based on their intended application to better understand the unique traffic characteristics of different M2M categories. We start this process by identifying the hardware model of each cellular device using the device's Type Allocation Code (TAC), which is part of the unique identifier of each cellular device. Although the records in our data set are anonymized, the TAC portion of the unique identifier is retained. Thus, the hardware model of each cellular device is obtained by consulting the TAC database of the GSM Association.

Because there is no rigorous definition for M2M devices or standard ways for determining their application categories, and many devices have multiple uses, knowing the device model is not sufficient for identifying a device with certainty as M2M device nor for identifying its M2M category. Towards this end, we adopt the device classification scheme of a major cellular service provider as a base template for categorizing M2M devices [1]. To supplement and verify this template, we also use public information such as production brochures and specification sheets. In total, we have classified more than 150 device models as M2M devices, and further divide them into the following 6 categories.

1) Asset Tracking: These M2M devices are used to remotely track objects like cargo containers and other shipments. These devices are often coupled with other sensors for tasks like temperature and pressure measurement. In our trace, about 18% devices belong to this category.

2) Building Security: These M2M devices are typically used to manage door access and security cameras. In our trace, about 14% devices belong to this category.

3) Fleet: These M2M devices are used to monitor vehicle locations, arrivals, and departures and provide real-time access to critical operational data for logistic service providers. In our trace, about 51% devices belong to this category.

4) Miscellaneous (Misc.): These M2M devices are used as generic cellular communication modems with embedded system data input and output ports such as serial, I2C, analog, and digital. They provide network connectivity for customized solutions. In our trace, about 9% devices belong to this category.

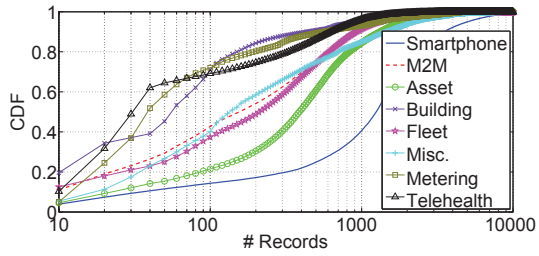
5) Metering: These M2M devices are mostly used for remote measurement and monitoring in agricultural, environmental, and energy applications. In our trace, about 6% devices belong to this category.

6) Telehealth: These M2M devices are mostly used for remote measurement and monitoring in healthcare applications. In our trace, about 2% devices belong to this category.

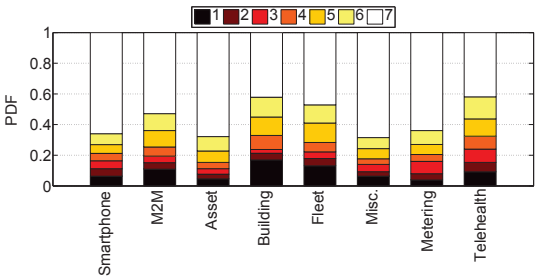
We acknowledge that due to lack of more detailed usage information and ambiguity in device registry databases, our classification may contain some errors. To limit such errors, we try to be as conservative as possible when deciding whether to include a M2M device model in our study. For example, cellular routers are generally excluded from this study because the actual end devices behind these routers cannot be identified. For cellular modems and modules, we exclude models with data interfaces likely used by modern day computers such as USB, PCI Express, and miniPCI but keep those with UART, SPI, and I2C interfaces. Note that we may miss some M2M devices in our analysis that are not active and hence they do not not appear in our trace. For the sake of comparing M2M and typical human-generated traffic characteristics, we have also included in our study traffic records from a uniformly sampled set of smartphone models, covering millions of smartphone devices.

C. Data Set Characteristics

Given the device categorization, we now investigate the following two basic characteristics of devices in our data set. First, we plot the cumulative distribution functions (CDFs) of record counts for smartphone and M2M devices in Figure 1(a). Note that "M2M" is the weighted average of the aforementioned six M2M device categories. We observe that M2M devices have lesser number of records as compared to smartphones. We note that about 40% of M2M devices have less than 100 records in our trace, while this number is up to 1000 for smartphones. We also observe diversity across different M2M device categories. Asset tracking, fleet, and



(a) # Records

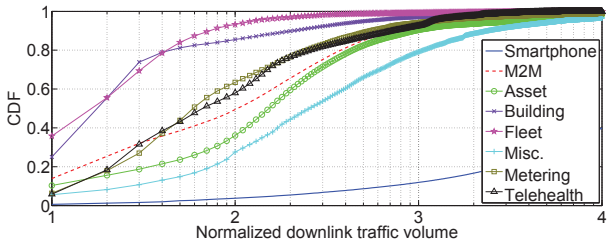


(b) # Unique Days

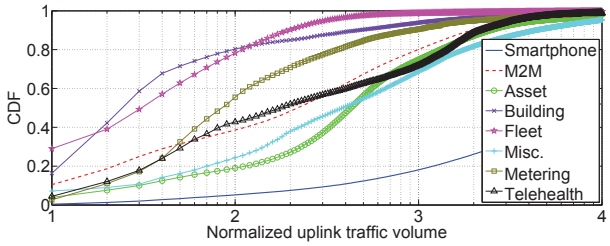
Fig. 1. Distributions of record count and number of unique days that devices appear in our trace.

misc. devices have significantly more records as compared to building, metering, and telehealth devices. M2M devices may have lesser number of records as compared to smartphones because they often have one-off appearance in our trace. Second, to rule out the one-off appearance hypothesis, we plot the probability distribution functions (PDFs) of the number of unique days smartphone and M2M devices appear in our trace in Figure 1(b). We observe some differences across M2M device categories; however, overall M2M and smartphones both have the most fraction of devices that appear on all days of the week in our trace. Therefore, we can conclude that differences observed in Figure 1(a) are not simply because M2M devices appear sporadically in our trace. The observations from these two plots provide us a first evidence of difference in network activity of smartphone and M2M devices, and across M2M device categories.

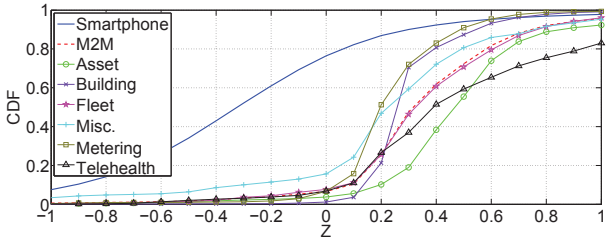
In the following sections, we conduct a detailed analysis and comparison of M2M and smartphone traffic characteristics in our data set. The traffic characteristics analyzed in this paper include aggregate data volume, volume time series, session analysis, mobility, application usage, and network performance. Note that some results presented in this paper are normalized by dividing with an arbitrary constant for proprietary reasons. However, normalization does not change the range of the metrics used in this study. Furthermore, the missing information due to normalization does not affect the understanding of our analysis. These characteristics are discussed below in separate sections.



(a) Downlink



(b) Uplink



(c) Ratio: log(Uplink/Downlink)

Fig. 2. CDFs of aggregate downlink and uplink traffic volume.

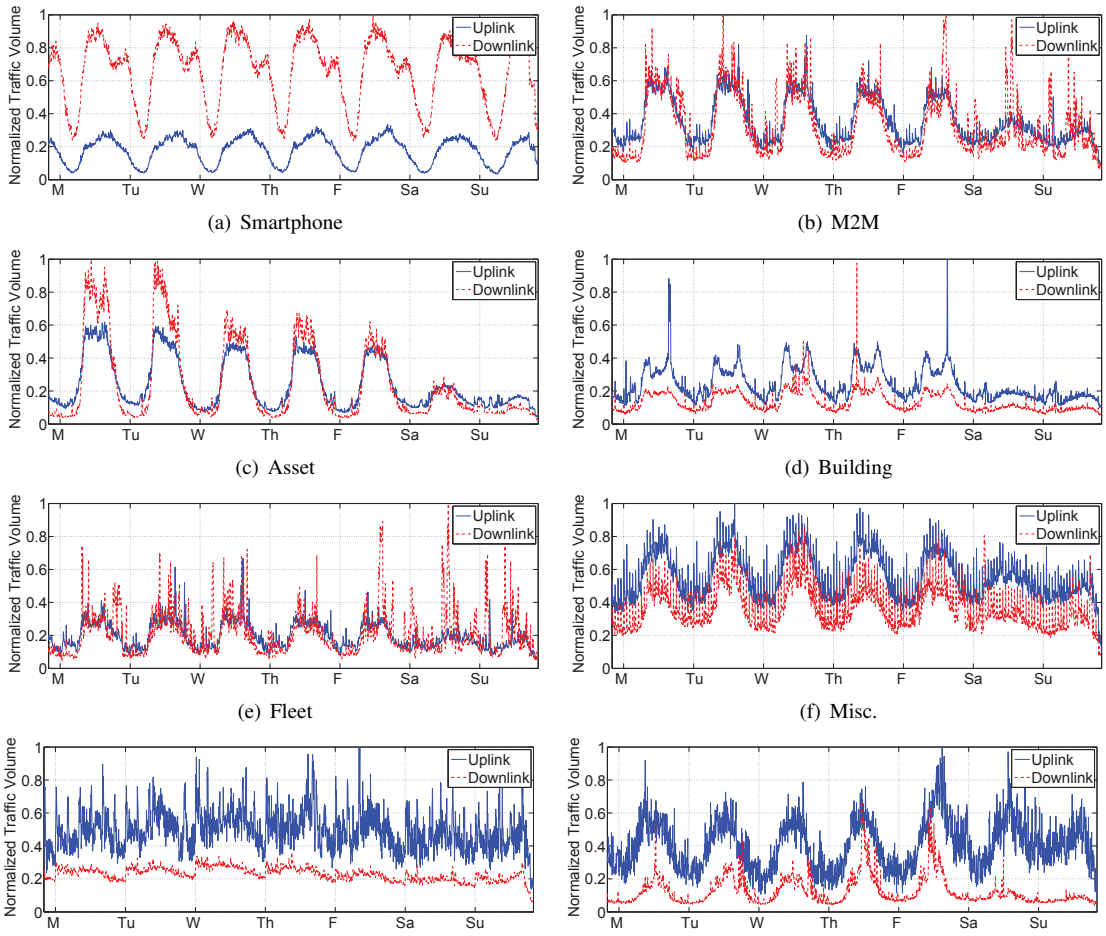
III. Aggregate Traffic Volume

When a new technology emerges and it has to share resources with existing parties, a natural first question is the level of competition and how different parties can better coexist. This is why we first study and compare the distribution of aggregate traffic volume for M2M devices and smartphones. Moreover, we also investigate whether the long established perception of traffic volume being downlink heavy remains true for M2M devices [16].

Figure 2 shows the CDFs of downlink and uplink normalized traffic volume for M2M devices and smartphones separately. The normalized traffic volume ranges between 1 (about 1KB) and 4 (maximum volume for M2M devices). For M2M, we show both the distributions for all M2M devices together and for each M2M category. We first notice that different device categories exhibit strong diversity in aggregate downlink and uplink traffic volume distributions. However, we do observe a consistent relative ordering of CDFs for different device categories. We note that the average downlink and uplink traffic volume for smartphones is about two orders of magnitude larger compared to all M2M device categories. Within M2M device categories, misc. category has the largest downlink traffic volume, followed by asset category; whereas, building security and fleet categories have the smallest downlink traffic volume. A similar ordering is also observed for uplink traffic volume.

We now study the distribution of ratios of uplink traffic volume to downlink traffic volume. For the sake of clarity, we plot the ratios after taking their logarithm, denoted by Z. The positive values of Z represent more uplink traffic volume than downlink traffic volume and its negative values represent more downlink traffic volume than uplink traffic volume. It is not surprising that approximately 80% of smartphone devices have $Z \leq 0$; thereby, indicating larger downlink traffic volumes. However, this trend is reversed by large margin for all M2M device categories, which all have $Z > 0$ for more than 80% of devices indicating larger uplink traffic volumes. This finding provides another evidence that M2M traffic has significantly different characteristics compared to traditional smartphone traffic. Comparing different M2M device categories, we observe that building and metering categories have the lowest average Z values; whereas, asset and telehealth have the highest average Z values. Such differences provide insight into the functionality of M2M device categories.

Summary: Overall, the average per-device traffic volume of M2M devices is much smaller than that of smartphones. However, the strength of M2M devices is really in the size of their population. As M2M population continues to increase, how network operators efficiently support a large number of low volume devices will become an important issue. Our finding that M2M traffic has more uplink volume than downlink volume shows that M2M devices act more as "content producers" than "content consumers". Interestingly, this difference coincides with the paradigm shift in web and mobile computing towards user-centric content generation. The momentum of such a shift may eventually question the assumptions for optimization approaches exploiting downlink asymmetry of network traffic [14].



(a) Smartphone (b) M2M

(c) Asset (d) Building

(e) Fleet (f) Misc.

(g) Metering (h) Telehealth

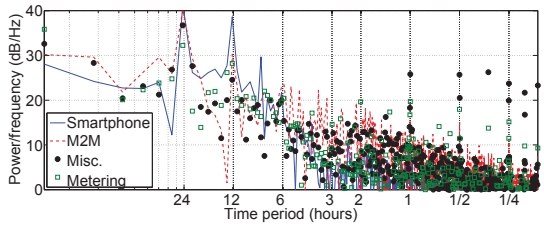
Fig. 3. Downlink and uplink traffic volume time series.

IV. Traffic Volume Time Series

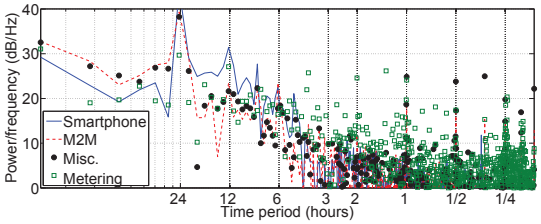
Having gained an understanding of aggregated M2M traffic volume, the next step is to study the temporal dynamics of M2M traffic volume. It would be interesting to know whether M2M devices exhibit similar daily diurnal pattern as smartphones. One particular use of such information is to evaluate the potential benefits of incentive programs such as billing discounts encouraging non-peak time usage. This information can also be utilized to group devices into separate clusters with different billing schemes. Time series analysis is also helpful for gaining insights into the operations of M2M devices.

As mentioned in Section II, the logged traffic records contain timestamps at 5-minute time resolution. Therefore, we can separately construct averaged traffic volume time series for smartphones and all M2M device categories. We plot these averaged uplink and downlink traffic volume time series in Figure 3. While the daily diurnal pattern is evident for both M2M and smartphone traffic, the comparison of Figures 3(a) and (b) reveals the following two interesting differences. First, the volume of downlink traffic dominates that of uplink traffic for smartphones, whereas these are relatively same in M2M traffic time series. This finding follows our earlier observations in Section III. Second, we also observe that peaks in smartphone traffic time series are wider, starting in the morning and prolonging up to mid-night, whereas peaks in M2M traffic time series are narrower, ending by the evening time; and M2M traffic volume exhibits significant reduction during weekend compared to weekdays while smartphone traffic volume remains virtually unchanged. It appears that smartphone traffic time series is coupled with human "waking" hours while M2M traffic time series is coupled with human "working" hours. This is a strong indication that currently a majority of M2M devices are employed for business use. They are not yet in the mainstream for residential users, or as tightly integrated into people's daily life as smartphones.

We have also separately plotted averaged uplink and downlink traffic volume time series for all M2M device categories in Figures 3(c)–(h). We observe strong diurnal variations for all M2M device categories. However, the weekday-weekend pattern comparison reveals different results for most M2M categories, illustrating that M2M categories indeed behave vastly differently from each other due to the different applications they serve. The previously mentioned association of M2M traffic time series with daily business activity cycle is highlighted the most by Figure 3(d) (Building), where each working day pattern displays not only elevated volume during working hours but also two peaks which in time coincide with the beginning and the end of typical business hours. Contrastingly, we see that there is virtually no difference in traffic volume for the metering category for different days.



(a) Downlink



(b) Uplink

Fig. 4. Periodograms containing power spectral density estimate of traffic volume time series.

Frequency Analysis: On a finer scale, we observe repetitive spikes in time series of most M2M device categories. To investigate these high frequency spikes in more detail, we plot the periodograms of some downlink and uplink traffic volume time series in Figures 4(a) and (b), respectively. The periodogram is an estimate of power spectral density, or frequency spectrum, of a given signal and is defined as:

$$S(f) = \frac{1}{F_s N} \left| \sum_{k=1}^{N} x_k e^{-j(2\pi f/F_s)k} \right|^2,$$

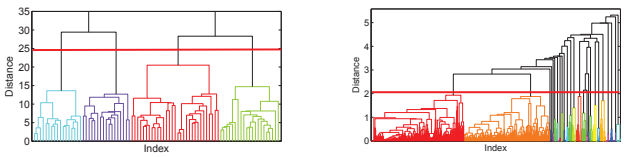
where f is the frequency in Hertz, N is the total number of signal samples, and F_s is the sampling frequency [22]. In this study, we have $F_s = 5$ minutes and $N = 2,016$. The x-axis in Figure 4 represents time period on logarithm scale and y-axis represents power in decibels (dB) for each frequency. We observe distinct spikes in the periodograms corresponding to multiple time periods, *e.g.* 1 hour, 30 minutes, and 15 minutes, strongly suggesting the timer-driven nature of many M2M operations. The perfect alignment of the spikes in Figure 3(f) and (g) to one, half, and quarter hour marks in time also suggests that these timers are highly synchronized.[1] Such synchronized communication by large number of devices is highly undesirable both for the M2M application service providers and cellular network operators because it may create disruptive congestion at various locations in the infrastructure [23]. It is noteworthy that such sub-hour frequency components are absent for smartphone traffic time series, highlighting peculiar nature of M2M traffic.

Time Series Clustering: Until now we have only examined the averaged time series for M2M device categories. To gain more fine-grained insights, we construct more than 150 M2M device model traffic time series from our trace, each representing averaged time series of individual devices of respective device models at 5-minute time resolution. Likewise, we construct individual device traffic time series at the same time resolution. However, the time series of individual devices at 5-minute time resolution over the duration of one week are less useful because they are highly sparse. To reduce their sparsity, we change the time resolution to 1 hour and also average them across all days. Therefore, the time series of individual devices each contain 24 data points representing hourly time series averaged over all days of the week.

With these two sets of time series (device models and individual devices) at hand, we now aim to find some structure across them by clustering together similar traffic time series. In this paper, we utilize discrete wavelet transform to analyze and compute similarity score between time series at multiple time scales [7]. Wavelet transform is a generalized form of Fourier transform, which resolves it as a series of sines and cosines of different frequencies. Using discrete wavelet transform, a traffic time series is decomposed into multiple time series, each containing information at different scales that range from coarse to fine. There are several well-known wavelet families whose qualities vary according to several criteria. We need to select an appropriate wavelet type for our given problem. We explored a wide range of wavelet types. However, we focus on the results based on the well-known Daubechies-1 wavelet type, which is computationally and memory-wise efficient and is known to appropriately handle discontinuities [7]. Note that the traffic time series in our data, especially at finer time resolutions, often contain discontinuities. We rely on Daubechies-1 to smooth out these discontinuities. The wavelet transform can be applied for varying decomposition levels to capture varying levels of detail (or scales). In this paper, we have used Coifman and Wickerhauser's well-known method to select the optimal number of decomposition levels [8]. The basic idea of this method is to select the decomposition level for which the joint information entropy of approximation and detail is minimized. We applied this method separately for device model and subscriber traffic time series and then selected the optimal decomposition level at the 95th percentile. Using the aforementioned criterion, we chose the optimal decomposition level to be 5 and 3 for device model and subscriber time series respectively.

Given the wavelet decompositions of device model and subscriber time series, we aim to group time series into distinct clusters. Towards this end, we need to select appropriate similarity metric and clustering mechanism to group them. We first note that the length of all traffic time series is the same. Therefore, we can compute one-to-one difference between any two given time series and compute its l^2 norm to quantify their dissimilarity. This continuous definition of similarity between two traffic time series allows us to apply hierarchical clustering. In this clustering method, we start by considering each traffic time series as a separate cluster and then recursively combine two clusters that have the smallest distance between them. Here we need to define the distance between two clusters each of which may contain more than one traffic time series. A well-known method is called Ward's method, which selects to merge two clusters for whom the increase in the sum of squared distances is minimum [13]. We use the Davies-Bouldin index to select the optimal number of

[1]For a causal analysis of the spikes in Figure 3, we have manually analyzed the traffic logs for potential patterns. Our analysis showed that all spikes are caused by coordinated activities from thousands of devices belonging to the same device models, not by a small number of "outliers". For instance, spikes for misc. category are caused by traffic belonging to thousands of devices of a particular model exactly at hour marks.



(a) Device model (b) Individual device

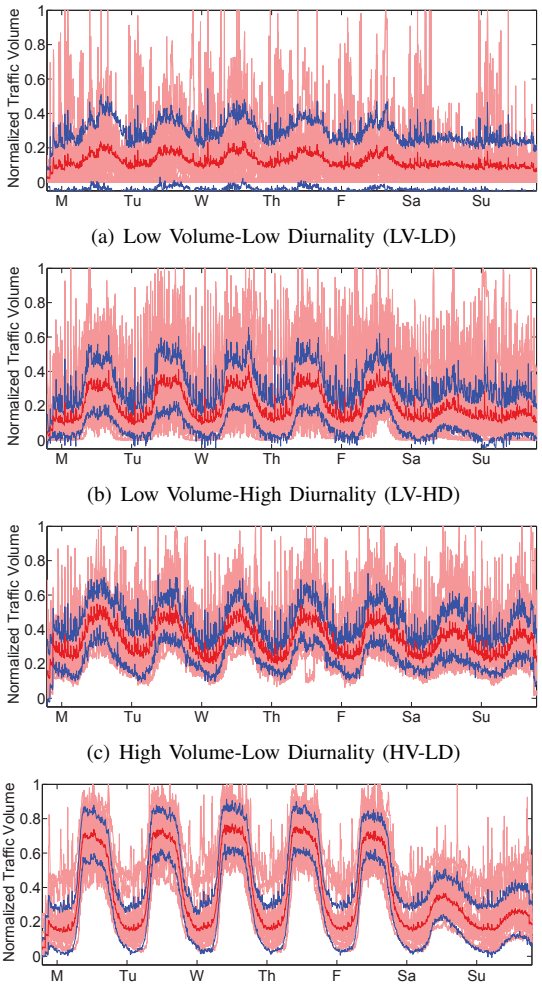
Fig. 5. Dendrograms for hierarchical clustering.

clusters from dendrogram, which is known to result in compact and well-separated clusters [13]. Figure 5(a) shows the dendrogram for hierarchical clustering of M2M device model traffic volume time series. The x-axis represents the indices of time series and y-axis represents the l^2 norm distance metric. In the dendrogram, we visually observe an obvious grouping of device models into well-separated clusters. Using the Davies-Bouldin index, the optimal number of clusters is selected to be four for M2M device model dendrogram. Similarly, Figure 5(b) shows the dendrogram for hierarchical clustering of individual device traffic volume time series. We observe a different structure in this dendrogram as compared to the one in Figure 5(a). Here we note that the bottom-right of the tree contains several clusters, each containing one or a small number of individual device time series. Intuitively, these small clusters potentially represent outliers whose distance to other clusters is fairly large. We visually observe two clusters on the bottom-left of Figure 5(b), each containing a major chunk of time series. After separating out the sparse outlying clusters, these two clusters are selected as optimal by the Davies-Bouldin index. To further study the clusters identified using the above-mentioned methodology, we plot their centroids with point-wise standard deviations in Figures 6 and 7. In these figures, the dark red lines represent the centroids, the blue lines represent point-wise standard deviations, and the light red lines in the background represent the member time series for each cluster.

Figure 6 shows device model clusters where we label the identified centroids based on two of their temporal characteristics: traffic volume and diurnal variations. We label a cluster centroid as high volume if its average normalized daily peak volume for weekdays is more than ≈ 0.5. Otherwise, the cluster centroid is labeled as low volume. Similarly, we label the cluster centroids based on the diurnal variations in the following way. Let DI denote the diurnality coefficient, and $V_{max}(d)$, $V_{min}(d)$, and $V_{avg}(d)$ denote the maximum, minimum, and average traffic volumes, respectively, on day d of a traffic time series spanning $|D|$ days. The diurnality coefficient is quantified as:

$$DI = \frac{1}{|D|} \sum_{\forall d \in D} \frac{V_{max}(d) - V_{min}(d)}{V_{avg}(d)}.$$

If the diurnality coefficient of a cluster centroid is more than 1.0 then it is labeled as high diurnality. Otherwise, it is labeled as low diurnality. Using this labeling methodology, we label the identified clusters as low volume-low diurnality (LV-LD), low volume-high diurnality (LV-HD), high volume-low diurnality (HV-LD), and high volume-high diurnality (HV-HD). We now study the composition of these labeled clusters with respect to the categories defined in Section II. Table I



(a) Low Volume-Low Diurnality (LV-LD)

(b) Low Volume-High Diurnality (LV-HD)

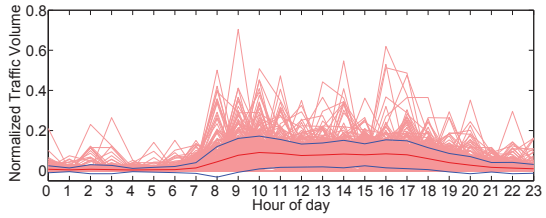
(c) High Volume-Low Diurnality (HV-LD)

(d) High Volume-High Diurnality (HV-HD)

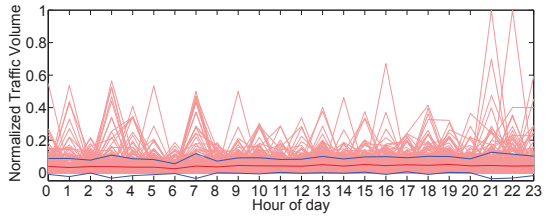
Fig. 6. Cluster centroids identified using device models traffic time series.

shows the composition of the identified clusters across all categories and the largest value in every row is marked as bold. We note that asset tracking and fleet device models mostly belong to HV-HD cluster. Furthermore, we note that building security and telehealth device models mostly belong to LV-HD cluster. These observations follow our intuition that the activity of these device models is tightly coupled with human activities. They also indicate that building security and telehealth device models tend to generate low traffic volume. Similarly, we observe that metering device models mostly belong to LV-LD cluster. This observation follows our earlier finding from Figure 4 that metering devices tend to download or upload data after periodic time intervals throughout the day. Finally, we observe that misc. device models mostly belong to HV-LD clusters. Similar to metering device models, misc. device models also tend to generate traffic after periodic time intervals throughout the day resulting in low diurnality.

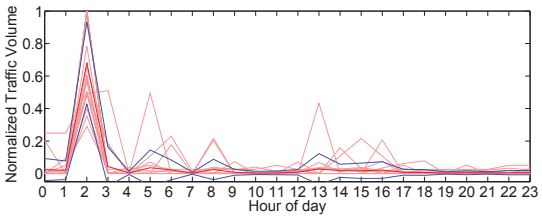
For individual device traffic volume time series clustering, we identified a handful number of outlier clusters and two main clusters containing a majority of devices. We plot the centroids of two main clusters and one of the outlier clusters in Figure 7. The cluster centroid in Figure 7(a) shows strong diurnal behavior with higher traffic volume during day time as compared to night time; therefore, we label this cluster as



(a) Diurnal



(b) Flat



(c) An outlier cluster

Fig. 7. Cluster centroids identified using individual device traffic time series.

diurnal. On the other hand, the cluster centroid in Figure 7(b) does not show any diurnal characteristics and is labeled as flat. We also show an outlier cluster in Figure 7(c), which consists of devices generating traffic volume spikes at late night. To gain insights from the clustering results of individual device traffic volume time series, we study their composition across various M2M device categories. Table II shows the cluster composition results with the largest value in every row marked as bold. We have similar observations for device level clustering as we previously had for device model traffic volume time series clustering. For instance, asset tracking and fleet devices mostly belong to the diurnal cluster. Misc., metering, and telehealth devices, with spiky traffic volume time series, mostly belong to the outlier cluster. Finally, building security devices mostly belong to the flat cluster. The individual device level clustering results further improve our understanding about M2M traffic behavior.

Summary: In this section, we have presented time series analysis for M2M traffic volume. Just like that of smartphones, M2M traffic volume also exhibits strong daily diurnal pattern. However, M2M traffic volume peaks correspond to people's working hours while smartphone traffic volume peaks correspond to waking hours, which indicates that a majority of M2M devices are employed for business use. The overlap between M2M peaks and smartphone peaks suggests that incentive based leverage mechanism such as off-peak time pricing for encouraging better sharing of network capacity can be beneficial. Towards this end, we presented a clustering algorithm which classifies M2M device models into four primary cluster categories, which can serve as a guideline to network operators in determining how to differentiate pricing for different device models. Specifically, high volume and high diurnality traffic results in higher peak load and wasted resources during non-peak hours. Therefore, devices belonging to high volume and high diurnality cluster should be charged relatively more than those belonging to low volume and low diurnality cluster. We have also investigated fine-grained features in traffic volume time series for different categories and uncovered the differences in behaviors among different M2M categories. For example unlike other M2M categories, metering devices show only weak diurnal pattern, suggesting that the traditional approach of scheduling service down time in early morning hours may not be the best for them. Finally, the surprising discovery of synchronized communication among M2M devices highlights the importance of developing and imposing standard traffic protocols and randomization methods. Such synchronized communication can be discouraged by employing 95th percentile pricing [9].

V. Session Analysis

We now analyze and compare session-level traffic characteristics of M2M devices and smartphones. Understanding session duration and inter-arrival distribution is a time honored tradition for the telecommunication industry because they are important inputs for network resource planning and management. Such information is valuable for cellular network operators too because device active time corresponds more closely to radio resource usage than aggregate traffic volume [18]. Moreover, being able to accurately estimate session timing parameters not only improves radio resource use efficiency for cellular operators, it also helps M2M service providers to better design their devices and protocols for better battery management.

Towards this end, we first formally define a session and then study different metrics based on session-level information. A flow consists of all packets in a given transport layer connection, including TCP and UDP. To study characteristics of flows at a given time resolution, we need to define equally-spaced time bins denoted by Δ_i where $i = 1, 2, ...$ and $|\Delta|$ denotes the magnitude of time bin and i is the index

TABLE I
Composition of device model clusters.

	LV-LD %	LV-HD %	HV-LD %	HV-HD %
Asset	13.1	16.1	23.5	**33.3**
Building	13.1	**19.3**	11.7	0.0
Fleet	26.2	29.2	5.9	**61.2**
Misc.	34.5	25.8	**47.1**	5.5
Metering	**13.1**	3.2	5.9	0.0
Telehealth	0.0	**6.4**	5.9	0.0

TABLE II
Composition of individual device time series clusters.

	Diurnal %	Flat %	Outlier %
Asset	**16.9**	15.7	4.8
Building	11.8	**19.1**	15.7
Fleet	**57.1**	47.5	44.6
Misc.	8.3	11.1	**12.0**
Metering	2.3	3.8	**9.6**
Telehealth	3.6	2.8	**13.3**

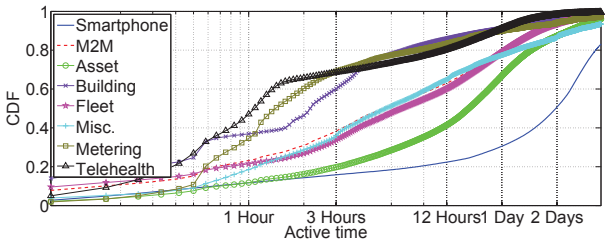
variable. Recall from Section II that the smallest available time resolution in our traffic trace is 5 minutes; therefore, we use $|\Delta| = 5$ minutes in this analysis. We look at flow arrivals in 5 minute time bins as a binary random process, which is denoted by $\{F_t : t \in T, F \in \{0,1\}\}$ and where 0 and 1 respectively denote absence or presence of flow arrival, respectively. We now define a session as a run of flow arrivals in consecutive time bins, where a flow spanning multiple time bins is marked for all time bins during its span. A session is denoted by $S_{\{t_x(i),t_y(i)\}}$, where $t_x(i)$ and $t_y(i)$ are the times corresponding to the first flow arrival and the last flow arrival of i-th session. In the following text, we separately investigate several metrics that capture diverse characteristics of the session arrival process.

Active Time: The first metric that we study is device active time, denoted by T_{active}, which is the total amount of time in our week-long trace when a device is sending or receiving traffic. In our study, it is calculated by multiplying number of unique time bins in which we have at least one flow arrival by the bin duration. Using this metric, we are primarily interested in studying the impact of devices on the network in terms of radio channel occupation. Note that a given time bin may have multiple flow arrivals but they are all mapped to 1. Mathematically, active time is defined as:

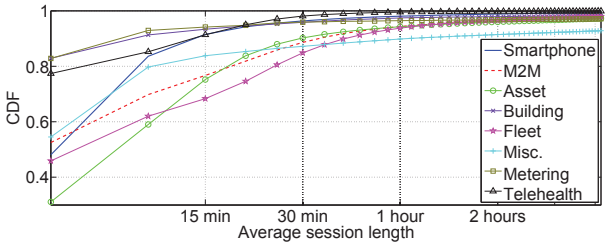
$$T_{active} = \sum_{\forall t \in T} F_t \text{ (counts)} = \sum_{\forall t \in T} F_t * |\Delta| \text{ (time units)}.$$

In Figure 8(a), we plot the CDFs of active time for smartphones and all M2M categories defined in Section II. The x-axis represents active time, which ranges from a minimum of $|\Delta| = 5$ minutes to a maximum of one week (*i.e.* the duration of trace collection). We first observe significant diversity in active time of devices of smartphone and all M2M categories. Our second observation is that smartphones tend to have significantly more active time compared to all M2M device categories. The median active time for smartphones is approximately 2 days, which is approximately 30% of the total trace time duration. It is important to note that active time cannot be accurately related to the interaction time of users with smartphones because of the following two reasons. First, users can interact with smartphone without actually generating network traffic, *e.g.* playing offline games. Second, some applications may generate background traffic when the user may not be actually interacting with the smartphone. We also observe diversity in the distributions of active time across M2M device categories. Misc. and asset tracking categories, with high aggregate traffic volume per device, have the largest active time values among all categories. It is noteworthy that the fleet category, despite small aggregate traffic volume per device, have above average active time values. This observation suggests that fleet devices tend to generate well spread out traffic across different time bins. We have also verified this conjecture from the data. Finally, telehealth and metering devices have the smallest active time among all M2M device categories.

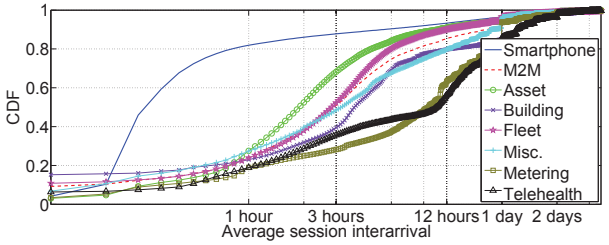
Average Session Length: Another metric that we study is average session length L_{avg}, which is defined as the average count of consecutive time bins with flow arrivals. Mathematically, for a device with n sessions L_{avg} is defined as:



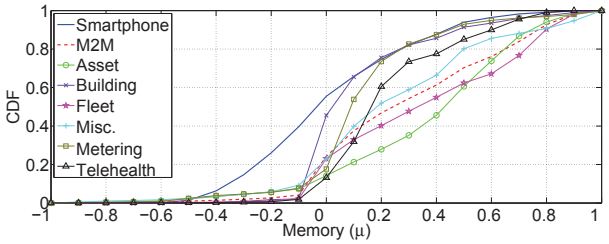
(a) Active time



(b) Average session length



(c) Average session inter-arrival time



(d) Memory

Fig. 8. CDFs of active time, average session length, session inter-arrival, and memory.

$$L_{avg} = \sum_{i=1}^{n} \frac{t_y(i) - t_x(i) + 1}{n}.$$

Note that session lengths are potentially inflated because session timeouts may be missed due to the coarse measurement granularity. Figure 8(b) shows the CDFs of average session length for M2M devices and smartphones. We note that a significant chunk of devices for all categories have average session lengths smaller than or equal to 5 minutes (left most points). However, remaining devices do have average session lengths significantly larger than the minimum value. For instance, 10% devices of misc. category have average session lengths larger than one hour, which reflects the way these devices operate. It is also interesting to note that smartphones typically have significantly smaller average session lengths compared to asset tracking, fleet, and misc. categories. Among M2M categories, telehealth, metering, and building security have the smallest average session lengths.

Average Session Inter-arrival: We also study the average session inter-arrival metric, which is defined as the average

of inter-arrival times between consecutive sessions. Using the earlier notion, we can mathematically define average session inter-arrival T_{avg} as follows:

$$T_{avg} = \sum_{i=1}^{n-1} \frac{t_x(i+1) - t_y(i)}{n-1}.$$

Figure 8(c) shows the CDFs of average session inter-arrival time for smartphone and M2M categories. We observe an approximately opposite trend as compared to active time and average session length for M2M device categories. For instance, metering and telehealth categories, with relatively small active time and average session lengths, have relatively large average session inter-arrival time with median values of approximately 9 hours. On the other hand, asset tracking and fleet categories have relatively relatively small average session inter-arrival time with median values of less than 3 hours. Smartphones tend to have even smaller average session inter-arrival time, where approximately 80% of devices have less than one hour average session inter-arrival time.

Burstiness of Session Arrivals: Another useful metric for the flow arrival process is burstiness. Burstiness jointly takes into account the runs of zeros and ones in a binary random process. As mentioned earlier in this section, we have modeled the flow arrival process as a binary random process, where arrivals are not independent. Given the assumption of conditional independence between consecutive flow arrivals, we can model the burstiness of the discrete flow arrival process using a 1^{st} order and 2 state discrete time Markov chain. This Markov chain is also known as the Gilbert-Elliot model and is shown in Figure 9. The two states of the Markov chain represent the arrival or non-arrival of a session in a given time bin; for instance, state 0 refers to non-arrival and state 1 refers to arrival of a session. A suitable metric to model the burstiness of the Gilbert-Elliot model is its memory, which is denoted by μ and is defined as: $\mu = 1 - P_{0|1} - P_{1_0}$, where $-1 \leq \mu \leq 1$. Furthermore, $\mu = 0$ corresponds to zero memory, $\mu \geq 0$ corresponds to persistent memory, and $\mu \leq 0$ corresponds to oscillatory memory. When $\mu = 0$, the probability of a session arrival at any time instance is independent of whether or not there was a session arrival in the previous time bin, *i.e.* the process is memory-less. Figure 8(d) shows the CDFs of memory for smartphone and M2M categories. We again observe significant differences across smartphones and M2M devices. Specifically, we note that more than 50% smartphones have oscillatory memory, whereas, more than 80% M2M devices have persistent memory. This indicates that most M2M devices, on average, tend to show persistence in network activity, *i.e.* a time bin with no flow arrival is likely to be followed by another without flow arrival and a time bin with flow arrival is likely to be followed by another with flow arrival. Among M2M device categories, building security category has the largest percentage of subscribers with negative memory values, indicating the presence of oscillatory memory. These subscribers are more likely to follow an active time bin with an inactive time bin and an inactive time bin with an active time bin. The rest of the M2M device categories only have a small fraction of subscribers with negative memory values.

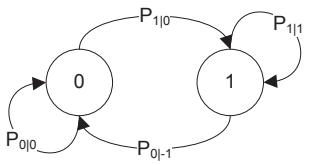


Fig. 9. Gilbert-Elliot Markov chain to model burstiness of session arrivals.

Summary: Once again, M2M traffic sessions exhibit rather different characteristics from smartphone traffic sessions. Overall M2M devices are active for traffic for much less time than smartphones. M2M traffic sessions occur much less frequently; however, M2M traffic sessions are more bursty. Consequently, the values of Radio Resource Control (RRC) timeouts of M2M devices can be decreased to avoid excessive radio channel occupation. Likewise, the values of RRC timers of smartphones can be increased to avoid excessive state transitions that result in degraded network performance [18]. It is also worth noting that 3 out of 6 M2M categories have about 80% of the devices with average session time lasting less than 5 minutes. This indicates that byte volume of data traffic for these devices is likely not an accurate reflection of their network resource use due to disproportional amount of control plane overhead for establishing and tearing down short sessions. The large differences between different M2M categories also advocate for differentiated RRC configurations for different categories.

VI. Mobility

In this section, we study and compare the mobility characteristics and geographical distribution of M2M devices and smartphones. Mobility patterns for different devices, constructed from our nation-wide trace, helps establishing an understanding for how much they move. Understanding mobility patterns for different devices has a direct impact on network resource planning. More importantly, we are interested in investigating how the locations of M2M device population are distributed relative to those of smartphones. Previously in Section IV, we have discovered that M2M traffic volume peaks overlap with those of smartphones in time. Here we investigate whether they also overlap in space.

It is important to note that cell identifiers derived from information collected within the core network are not considered an accurate approximation for device location. This is because many low-level radio access network operations such as handoffs of mobile devices between cells are not exposed to the core network. However, we consider such inaccuracy acceptable for three reasons. First, Xu *et al.* reported that although cell-sector information collected from the core network is not exact for the purpose of being used as device location, the median error is < 1 kilometer [26]. Second, we do not use the locations of the cell tower to proximate user device locations. We simply count the number of unique cells a device is involved with. Finally, the scope of our study covers the whole United States, compared to which cell-level errors at kilometer scale are rather minor.

Device Mobility: We quantify mobility in terms of the number of unique cells to which a device connects. It is noteworthy

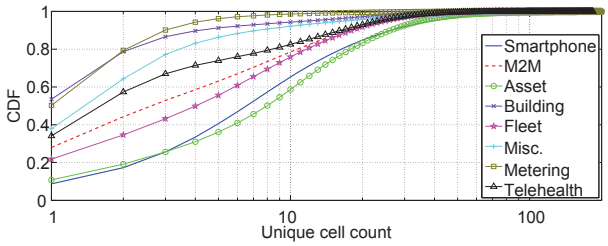


Fig. 10. CDFs of unique cells for Smartphone and M2M categories.

(a) Aggregate M2M

(b) Smartphone

Fig. 11. Geographical distribution of aggregate M2M and Smartphone traffic volume.

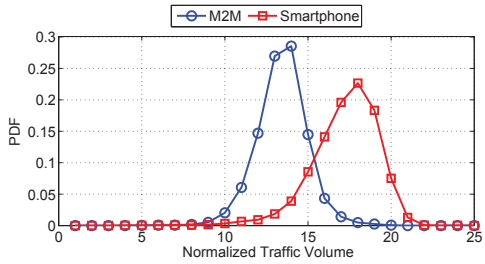
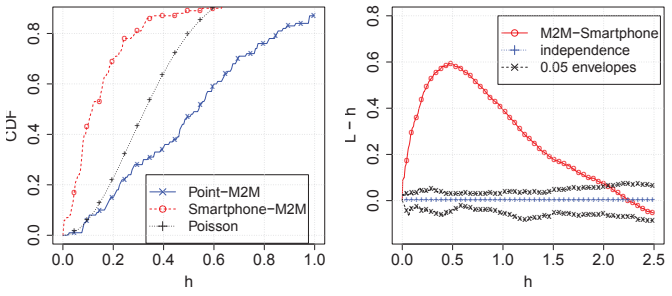


Fig. 12. PDFs of cell-level traffic volume for M2M and Smartphone devices.

that a device can be stationary and still connect to multiple cells at different time instances when it is in the coverage area of multiple cell sectors simultaneously. Figure 10 shows the CDFs of unique cells for different M2M device categories and smartphones. We observe that devices of M2M categories appear across less unique cells compared to smartphone devices, with the exception of asset tracking category. This is expected because asset tracking devices are typically connected to automotive vehicles for transportation of goods and tend to transmit more frequent updates, *e.g.* temperature and check-in information. Interestingly, fleet devices are less mobile than asset tracking devices and even smartphones. This is probably because fleet devices in our trace are mostly used by car rental companies and tend to transmit less frequent updates, *e.g.* milage and accidents. Overall, while we might intuitively believe that asset tracking and fleet devices are more mobile than the average smartphone, our results show that the difference is rather insubstantial. Furthermore, as expected, we observe that building security and metering devices appear across the least number of cells.

Geographical Distribution: We now investigate the geographical distribution of M2M traffic. Towards this end, we first plot the Voronoi diagrams for traffic volume of cell-level aggregated M2M and smartphone traffic in Figure 11. The geographical region shown in this figure covers more than 1 million square kilometers, spanning multiple states in the United States (about 1/9 of its total area). The Voronoi diagrams are generated by partitioning the 2D space, containing points representing base station locations, into polygons such that each polygon contains one base station and every point in a given polygon is closer to its base station than others. Note that the polygons in the Voronoi diagrams represent cells covering varying geographical areas. The clusters of cells covering small geographical areas appear around major population centers. The colors of polygons representing the traffic volume show that cells typically carry more smartphone traffic than M2M traffic. We plot the distributions of cell-level aggregated traffic volume for smartphones and M2M devices in Figure 12, which verify our earlier observation. From the perspective of network operators, we are interested in identifying locations with highest traffic volume for smartphones and M2M devices. Towards this end, in the rest of this section, we focus on the top-10% cells in terms of traffic volume for smartphones and M2M devices. These correspond to the right-hand side tails of the distributions plotted in Figure 12. Furthermore, we want to identify geographical dependencies among their locations for the top-10% cells for smartphones and M2M devices. The three possible types of geographical dependencies between two sets of locations are: attraction, repulsion, and independence. Attraction or repulsion between two sets of location respectively indicate correlation or anti-correlation, whereas independence indicates no correlation at all.

A well-known method to characterize geographical dependency between two sets of points is based on the nearest neighbor statistics [5]. Specifically, for two sets of points i and j, we can define $G_{ij}(h)$ as the probability that the distance from a randomly selected point i to the nearest event j is less then or equal to h. Likewise, we can define $F_j(h)$ as the probability that the nearest point j to a random point is less then or equal to h. If the two sets of points are geographically independent then $G_{ij}(h) = F_j(h)$. Given i and j respectively point to the top-10% locations for M2M and smartphone traffic in terms of traffic volume, Figure 13(a) plots G_{ij} (Smartphone-M2M) and F_j (Point-M2M) for varying values of h. A theoretical Poisson line is also plotted for reference, which indicates the expected pattern if both sets of points are independently distributed as homogeneous Poisson processes. We observe that both G_{ij} and F_j significantly depart from the theoretical Poisson line and they are also not close to each other. This observation indicates that the point sets i and j do not follow homogeneous Poisson distribution and are also not independently distributed of each other. The question remains if the point sets show attraction or repulsion to each other. This question is also addressed by the relative positioning of



(a) Nearest Neighbor (b) Cross-L

Fig. 13. Point pattern interaction analysis for high volume M2M and Smartphone cell locations.

G_{ij} and F_j lines, where G_{ij} line rises above F_j. This pattern shows that we have more than expected high volume M2M locations nearest to high volume smartphone locations. This indicates that these point sets are attracted to each other.

Another well-known method to study the geographical dependency of two point sets i and j, called cross-L, is based on Ripley's cross-K function [5]. It is denoted by $\hat{L}_{ij}(h) - h$ and is defined as:

$$\hat{L}_{ij}(h) - h = \sqrt{\frac{\hat{K}_{ij}(h)}{\pi}} - h,$$

where $\hat{K}_{ij}$ is the empirical Ripley's cross-K function and h denotes the distance. The empirical Ripley's cross-K function is defined as:

$$\hat{K}_{ij} = E(\#\text{ type } j \text{ points} \leq h \text{ from an arbitrary } i \text{ point})/\lambda_j.$$

Here λ_j is the average intensity of point set j and $E(.)$ is the expectation operator. Positive and negative values of $\hat{L}_{ij}(h)-h$ respectively indicate attraction and repulsion between two point sets. The co-independence is indicated if the $\hat{L}_{ij}(h) - h$ remains between the estimated confidence envelope lines for co-independent homogeneous Poisson processes. This method overcomes one limitation of the nearest neighbor analysis that it is not restricted to only considering the closest points. However, it is also limited because it gives us the average impression of all points in the data set and may overlook small-scale local dependencies. Figure 13(b) shows the plot of $\hat{L}_{ij}(h) - h$ for varying values of h. We again observe a significant attraction pattern between high volume M2M and smartphone traffic locations. The above-mentioned two sets of experiments jointly provide a strong evidence that high volume M2M and smartphone traffic locations are attracted to each other.

Summary: Most M2M devices are more likely to remain within a smaller geographical area compared to smartphones, with the exception of asset tracking devices. On the other hand, the geographical distribution of M2M device population, especially those with high traffic volume, exhibits "attraction" to high volume smartphone devices. In other words, the information provided by the analysis of mobility characteristics of M2M devices is mixed for network operators. While M2M devices are less mobile, which suggests that service optimization is easier to conduct because it only involves a small area, the co-location of high volume M2M devices with smartphone devices brings more chance for congestion in such areas.

VII. Application Usage

So far in this study we have treated all data bits equally, simply as "traffic volume". However, the truth is that not all bits are equal. For example, a bit that is part of an 8-bit encoding of a temperature reading obviously has higher information density than a bit in an image of a thermometer that displays the temperature. In this section, we attempt to understand how M2M devices use data traffic by exploring the application-mix of M2M data traffic.

Recall from Section II that traffic records in our data set are tagged with application identifiers. These identifiers cover traffic of 17 different application realms, including HTTP, email (POP, IMAP, *etc.*), and all common video streaming protocols (HTTP streaming, flash, *etc.*), all of which make up the vast majority of smartphone traffic volume. Using this application classification, we can compute application distribution of traffic volume for all M2M device categories. We observe that 95% of all flows in our trace belong to TCP. This observation is in accordance with the findings reported in prior literature [12], [17], [19]. We provide the averaged uplink and downlink application distribution of traffic volume for all M2M device categories in Figure 14. To first average a device category, we take the ratio of the sum of traffic volume of all devices and the total number of devices. We then normalize the averaged traffic application volumes by their maximum value. We note from Figures 14(a) and (b) that the traffic of all device categories mostly belongs to unknown or miscellaneous realms. This indicates that M2M devices typically use custom protocols that are either not identified by our application classification methodology, mentioned in Section II, or they use atypical protocols.

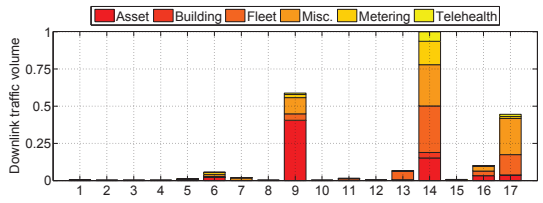
It is interesting to compare application distribution of M2M traffic with that of smartphone traffic shown in Figure 14(c). As expected, smartphone traffic mostly belongs to web browsing, audio and video streaming, and email applications. This is in sharp contrast to what we have observed for M2M traffic.

Summary: M2M devices mostly use custom application protocols. This makes it more difficult for network operators to understand and mitigate adverse effects from these protocols compared to the standard ones such as HTTP. Towards this end, better standardization of M2M protocols would certainly be a mutually beneficial solution for both M2M application service providers and cellular network operators.

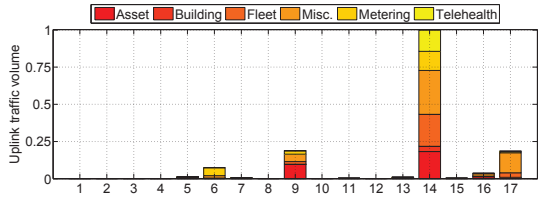
VIII. Network Performance

We now characterize the network performance of M2M traffic. We examine network performance in terms of round trip time (RTT) and packet loss ratio, both of which provide us unique perspectives of network performance.

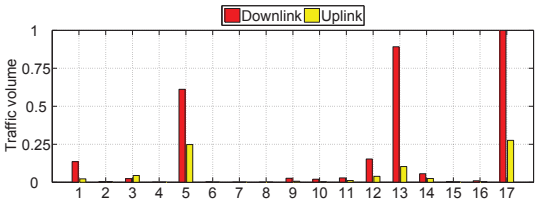
Round Trip Time: RTT is an important metric for network performance evaluation and is a key performance indicator that quantifies delay in cellular networks. The RTT metric is especially important for M2M applications that are real-time critical. It is important to note that RTT measurements can be potentially biased by differences in the paths between different cellular devices and the external servers they communicate with. For this study, we only have RTT measurements for TCP flows, which are estimated by the time duration between



(a) Downlink M2M



(b) Uplink M2M

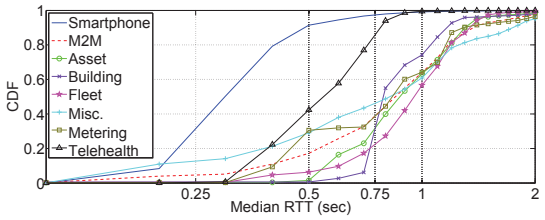


(c) Smartphone

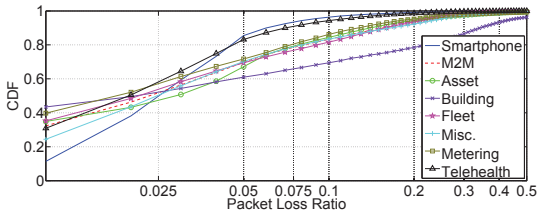
Fig. 14. Traffic application distributions of M2M device categories and smartphone. The application indices along x-axis are: (1) `appstore`, (2) `jabber`, (3) `mms`, (4) `navigation`, (5) `email`, (6) `ftp`, (7) `gaming`, (8) `im`, (9) `miscellaneous`, (10) `optimization`, (11) `p2p`, (12) `apps`, (13) `streaming`, (14) `unknown`, (15) `voip`, (16) `vpn`, and (17) `web`.

the trace collecting apparatus seeing a SYN packet and its corresponding ACK packet in the TCP handshake. Figure 15(a) shows the CDFs of the median RTTs experienced by each device for smartphones and all M2M device categories. We observe that all M2M device categories experience larger RTT compared to smartphones. Furthermore, within M2M device categories, telehealth devices have smaller RTT than all other categories. Our manual investigation of hardware specifications showed that smartphones and telehealth devices are mostly equipped with 3G modems, in contrast to other categories that typically rely on 2G modems. 2G RTTs are larger due to longer delays on the air interface, which explains these observations. In addition, smartphones are generally equipped with more powerful processors than M2M devices. Therefore, faster TCP/IP stack implementations on smartphone processors can also impact RTT.

Packet Loss Ratio: Packet loss ratio is a key performance indicator metric that quantifies reliability in cellular networks. We estimate the packet loss ratio from the fraction of the observed TCP sequence number range to the observed TCP payload bytes, summed over all TCP flows. This ratio is subtracted from 1 to obtain the packet loss ratio. Since most packet loss occurs in the radio access network (RAN) and our measurement point is in between the RAN and the Internet, this metric effectively estimates the downlink packet loss ratio. Figure 15(b) shows the CDFs of packet loss ratio for smartphones and all M2M device categories. Similar to the CDFs of RTT shown in Figure 15(a), we observe differences for packet loss ratio distribution in terms of third and fourth quartile values, where smartphones and telehealth devices



(a) Round trip time



(b) Packet loss ratio

Fig. 15. CDFs of round trip time and packet loss ratio.

experience at least an order of magnitude lower loss ratios than other M2M device categories due to a larger ratio of 3G to 2G modems. We also observe that building security devices have much higher third and fourth quartile loss ratios than other M2M devices, despite using similar technologies. This may be due to the placement of these devices indoors where the signal quality is poorer.

Summary: M2M traffic's network performance also differs from that of smartphones. The RTT of M2M traffic is significantly larger than smartphone traffic. Careful inspection of the hardware specifications of M2M devices reveals that M2M devices generally fall behind smartphones in choice of cellular technology. A majority of M2M devices still use 2G technologies such as GPRS and EDGE. Although 2G technologies are often adequate for M2M communication in terms of data rates, such lagging does present a challenge for cellular operators because they would need to maintain older generation services, instead of repurposing 2G spectrum for newer technologies of higher spectral efficiency. M2M traffic also generally has higher packet loss ratios. This is probably because some M2M devices are placed in locations with poor signal reception. It shows the need for M2M devices to have screens displaying cellular signal strength like cell phones.

IX. Conclusions and Future Directions

This paper presents the first attempt to characterize M2M traffic in cellular data networks. Our study was based on a week long traffic trace collected from a major cellular service provider's core network in the United States. In our analysis, we compared M2M and smartphone traffic in several aspects including temporal traffic patterns, device mobility, application usage, and network performance. We found that although M2M devices have different traffic patterns from smartphones, they are generally competing with smartphones for shared network resources.

Our findings presented in this paper have important implications on cellular network design, management, and optimization. Through better understanding of M2M traffic, cellular service providers can improve resource allocation mechanisms and develop better billing strategies for different categories of

M2M devices. Towards this end, Software Defined Networking (SDN) can be used for device- or subscriber-aware dynamic and flexible resource allocation and management [15]. SDN can also help to isolate or slice cellular network resources via virtualization to avoid contention between smartphone and M2M traffic. Note that this isolation or slicing can be fine-grained for different M2M device categories, or even different M2M applications. Moreover, delay tolerant and non-mission critical M2M traffic can be relayed over white spaces. The aforementioned network design and management techniques can impact the dynamics of M2M traffic and in turn may require introduction of novel pricing models by network operators.

REFERENCES

[1] AT&T specialty vertical devices. http://www.rfwel.com/support/hw-support/ATT_SpecialtyVerticalDevices.pdf.

[2] 3G machine-to-machine (M2M) communications: Cellular 3G, WiMAX, and municipal Wi-Fi for M2M applications. Technical report, ABIresearch, 2007.

[3] The global wireless M2M market. Technical report, Berg Insight, December 2010.

[4] Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016. White Paper, 2012.

[5] R. S. Bivand, E. J. Pebesma, and V. Gomez-Rubio. *Applied Spatial Data Analysis with R*. Springer, 2008.

[6] CalAmp. LMU-2600 GPRS fleet tracking unit. http://www.calamp.com/pdf/LMU-2600.pdf.

[7] P. Chaovalit, A. Gangopadhyay, G. Karabatis, and Z. Chen. Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys*, 43(2), 2011.

[8] R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2 Part 2):713–718, 1992.

[9] X. Dimitropoulos, P. Hurley, A. Kind, and M. P. Stoecklin. On the 95-percentile billing method. In *International Conference on Passive and Active Network Measurement (PAM)*, 2009.

[10] J. Erman, A. Gerber, M. T. Hajiaghayi, D. Pei, and O. Spatscheck. Network-aware forward caching. In *WWW*, 2009.

[11] Z. M. Fadlullah, M. M. Fouda, N. K. A. Takeuchi, N. Iwasaki, and Y. Nozaki. Toward intelligent machine-to-machine communications in smart grid. *IEEE Communications Magazine*, 49(4):60–65, 2011.

[12] A. Gerber, J. Pang, O. Spatscheck, and S. Venkataraman. Speed testing without speed tests: Estimating achievable download speed from passive measurements. In *ACM IMC*, 2010.

[13] N. Grira, M. Crucianu, and N. Boujemaa. Unsupervised and semi-supervised clustering: A brief survey. Report of the MUSCLE European Network of Excellence (FP6), 2004.

[14] L. K. Law, S. V. Krishnamurthy, and M. Faloutsos. Capacity of hybrid cellular-ad hoc data networks. In *INFOCOM*, 2008.

[15] L. E. Li, Z. M. Mao, and J. Rexford. CellSDN: Software-defined cellular networks. Technical report, Princeton University Computer Science Technical Report, 2012.

[16] Z. Moczar and S. Molnar. Comparative traffic analysis study of popular applications. In *International Conference on Energy-aware Communications*, 2011.

[17] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding traffic dynamics in cellular data networks. In *INFOCOM*, 2011.

[18] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. Characterizing radio resource allocation for 3G networks. In *ACM IMC*, 2010.

[19] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. Characterizing geospatial dynamics of application usage in a 3G cellular data network. In *IEEE INFOCOM*, 2012.

[20] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang. Characterizing and modeling Internet traffic dynamics of cellular devices. In *ACM SIGMETRICS*, 2011.

[21] L. Song and J. Shen, editors. *Evolved Cellular Network Planning and Optimization for UMTS and LTE*. CRC, 2010.

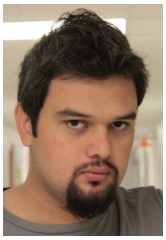
[22] P. Stoica and R. L. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[23] P. Traynor, M. Lin, M. Ongtang, V. Rao, T. Jaeger, P. McDaniel, and T. L. Porta. On cellular botnets: measuring the impact of malicious devices on a cellular network core. In *ACM CCS*, 2009.

[24] Trilliant. CellReader digital cellular meters. http://www.trilliantinc.com/products/cellreader/.

[25] I. T. Union. World telecommunication/ICT indicators database 2011. http://www.itu.int/ITU-D/ict/ publications/world/world.html, 2011.

[26] Q. Xu, A. Gerber, Z. M. Mao, and J. Pang. AccuLoc: Practical localization of peformance measurement in 3G networks. In *ACM MobiSys*, 2011.

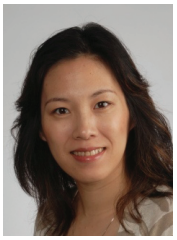


M. Zubair Shafiq received his B.E. degree in Electrical Engineering from National University of Sciences and Technology, Islamabad, Pakistan in 2008. He is currently a Ph.D. candidate in the Department of Computer Science and Engineering at Michigan State University. He was co-recipient of the IEEE ICNP 2012 Best Paper Award. He also received the 2012 Fitch-Beach Outstanding Graduate Research Award by College of Engineering, Michigan State University. His research interests are in the areas of cellular networks, online social networks, and content delivery networks with emphasis on measurement, performance analysis, and mathematical modeling.

Lusheng Ji is a Principal Member of Technical Staff - Research at the AT&T Shannon Laboratory, Florham Park, New Jersey. He received his Ph.D. in Computer Science from the University of Maryland, College Park in 2001. His research interests include wireless networking, mobile computing, wireless sensor networks, and networking security. He is a Senior Member of the IEEE.

Alex X. Liu received his Ph.D. degree in computer science from the University of Texas at Austin in 2006. He is currently an associate professor in the Department of Computer Science and Engineering at Michigan State University. He received the IEEE & IFIP William C. Carter Award in 2004 and an NSF CAREER award in 2009. He received the Withrow Distinguished Scholar Award in 2011 at Michigan State University. His research interests focus on networking, security, and dependable systems.

Jeffrey Pang is a researcher at AT&T Labs - Research. He received his Ph.D. in Computer Science from Carnegie Mellon University in 2009. He currently builds systems to measure and optimize cellular networks. His research interests include networking, mobile systems, distributed systems, and privacy.



Jia Wang received her Ph.D. in Computer Science from Cornell University in January 2001. She joined AT&T Labs - Research since then and is now a Principal Technical Staff Member in the Network Measurement and Engineering Research Department. Her research interests focus on network measurement and management, network security, performance analysis and troubleshooting, IPTV, social networks, and cellular networks. She was co-recipient of the ACM SIGMETRICS - IFIP WG 7.3 Performance 2004 Best Student Paper Award and the ACM CoNext 2011 Best Paper Award. She is a senior member of IEEE and a member of ACM. She serves as the associate editor of IEEE/ACM Transaction of Networking (ToN) and the area editor of ACM SIGCOMM Computer Communication Review (CCR). She also served on program committees and/or organizing committees of many top technical conferences including ACM SIGCOMM, ACM SIGMETRICS, ACM IMC, ACM CoNext, IEEE INFOCOM, IEEE ICNP, IEEE ICDCS. She is the general co-chair of SIGCOMM 2013.