Evolving Landscape of Cellular Network Traffic

Han Liu*, Chen-Nee Chuah*, Hui Zang†, and Sara Gatmir-motahari†
*Department of Electrical & Computer Engineering, University of California, Davis, CA, USA
†Sprint Labs, Burlingame, CA, USA
{bhgliu,chuah}@ucdavis.edu, {hui.zang,sara.gatmir-motahari}@sprint.com

***Abstract*—Recent technological advances have resulted in a dramatic change in the market shares of cellular mobile devices. However, little is known about the impact of these changes on the landscape of cellular network traffic. Using anonymized traces from one million cellular subscribers, we conduct a comparative study of the usage characteristics of three different types of mobile devices: feature phones, air cards, and smart phones. Our study covers three aspects: traffic volume in terms of data, voice, and short messages and corresponding temporal fluctuations, applications breakdown in data access, and the presence of malicious traffic. Our study reveals some similarities as well as distinct differences among the three device types. These insights into the modern cellular network traffic could influence how cellular carriers manage and provision their networks.**

***Index Terms*—cellular network traffic, smart phone, air card, feature phone, cellular device usage**

I. Introduction

The cellular network has seen a rapid change in the market shares of mobile devices in recent years. In 2009, smart phone penetration was only 25% in the US and 14% in the world. In the second quarter of 2011, it has risen to 55% in the US and 28% in the world [1]. Simultaneously, the market of air card products is also expanding very rapidly [2], while the usage of feature phones (also known as ordinary or dumb phones) declines every year [1]. There is currently very limited understanding of how these changes impact the landscape of cellular network traffic, which in turn have implications on how cellular carriers manage and provision their networks. Most previous studies on cellular networks are unaware of device type differences [3] [4] or dedicated only on smart phones [5], and hence can provide little insight on the differences and similarities among different mobile devices.

In this paper, we attempt to fill this knowledge void by conducting a detailed comparative study of the usage characteristics of three different types of mobile devices: feature phone, air card, and smart phone. Our study is based on two anonymized data sets from the CDMA2000/EVDO network of a major commercial cellular carrier in the USA. The first data set contains session level cellular traffic records of around 800,000 subscribers over one-month period. The second data set is a one-week long trace of IP packets generated by over 500,000 cellular data service users.

Our analysis covers three aspects: volume characteristics of cellular network traffic (including voice calls, short messages, and data access), application breakdown for IP (or data) traffic, and the presence of malicious IP traffic. We find immense diversity among different types of devices. For example, the ratios between the average IP traffic generated by an air card, a smart phone, and a feature phone are 4.5: 2.5: 1. The volume distributions of IP traffic with respect to mobile applications are quite diverse among different device types. Only smart phones and air cards generate port-scan traffic, each exhibit distinct scanning patterns.

Despite the differences mentioned above, similarities are also found among the different types of mobile devices. We find that HTTP and HTTPS packets account for the dominant share of IP traffic volume on all types of devices. In addition, the temporal volume fluctuations of IP traffic generated by different devices follow almost the same diurnal patterns. The distributions of IP traffic volume with respect to individual device types all follow the Zipf model. Furthermore, we observe that the volume patterns of voice calls and short messages are strongly related to the usage of data services and almost independent of the device types.

Our results provide important insights into the evolving landscape of cellular network traffic. A better understanding of the usage characteristics of different types of devices would allow cellular carriers to perform a better traffic demand forecast and resource provisioning as the relative popularity of these device types changes in the future.

The rest of the paper proceeds as following. We discuss related work in Section II, and describe data sets used in our study in Section III. Section IV presents the overall characteristics of cellular traffic for different types of devices. Section V provides a detailed analysis of the application breakdowns for IP traffic generated by different device types, while Section VI focuses on malicious traffic generated by cellular terminals. Finally we conclude the paper in Section VII.

II. Related Work

There are several recent measurement studies that cover different aspects of cellular traffic, including characterization of IP traffic (i.e., data access) [6], [5], [7], [8], [3], [9], [10] and voice call patterns [4], [11], [12]. However, these studies either consider overall cellular traffic (agnostic to device types), or focus specifically on smart phone usage. None of them have considered the differences between smart phones, feature phone, and air cards. For instance, Falaki et al. [5] presented a comprehensive study of smart phone usage from four perspectives: 1) user interactions with the phone, 2) application usage, 3) network traffic, and 4) energy drain. This study only considers IP traffic generated by smart phones and ignores other device types such as feature phones or air cards.

978-1-4673-1544-9/12/$31.00 ©2012 IEEE

In addition, it analyzed a small data set containing only 255 users. Our current work utilizes a much larger data set and covers a wider scope.

The only previous work that studies the differences among various types of cellular devices is done by Shafiq et al.[8]. Their study showed that the accuracy of the models describing cellular Internet traffic volume can be improved by customization based on device differences. In particular, cellular devices were grouped according to their trademarks and version numbers. On the other hand, our study is conducted at a different granularity, where we consider different device types: smart phones, feature phones, and air cards. Moreover, we consider overall usage of the cellular device, including not only data traffic, but also voice and short messages.

To the best of our knowledge, our paper presents the first measurement-based characterization of malicious traffic in cellular networks. However, the methods we use to identify malicious traffic are based on existing works. We quantify two types of malicious traffic: (i) port scan activity by detecting TCP SYN packets that get no response[13], [14], and (ii) packets sent to unused IP address space [15].

III. Overview of Data Sets

Our work is based on two sets of anonymized data collected from the CDMA2000/EVDO network of a commercial cellular carrier in a major state of the USA.

Dataset1: *Dataset1* is collected from four geographically-adjacent switches covering 955 cell towers, and contains the session records for over 800,000 cellular terminals from the month of October 2011. The records in this data set include 1) starting time and session length of every voice call, 2) starting time of every short message, 3) starting time and transmitted bytes of every data session, and 4) (hashed) identifiers and device type of the terminal for every session. Note that this dataset does not include any packet-level information.

Dataset2: *Dataset2* is a week-long IP packet trace collected in October 2011 from a gateway that connects the cellular EDVO network to the Internet. More than 500,000 cellular terminals appear in the dataset. The records include the arrival time stamp and complete application layer header for all up-link (sent from cellular terminals to the network) HTTP packets, and the five-tuple {Timestamp, Source Port, Destination Port, Source IP, Destination IP} for all other data packets. In addition, the "User-Agent" field of HTTP headers uniquely identifies each terminal and we obtained the device type for 85% terminals from an independent source. In our study, we only include the terminals with identified device type. Among the identified devices, around 76% are smart phones (the majority are Android phones), 14.5% are feature phones, and 9.5% are air cards. Note that this trace does not contain voice calls or short messages and only subscribers who use data service (referred as data users) appear in this dataset.

IV. Characteristics of Cellular Traffic Volume

In this section, we analyze the characteristics of overall cellular network traffic (including data traffic, voice calls, and short messages) generated by different types of devices. We are primarily interested in traffic volume, in terms of number of bytes, calls, and short messages, and call duration. We base this part of our study on *Dataset1*.

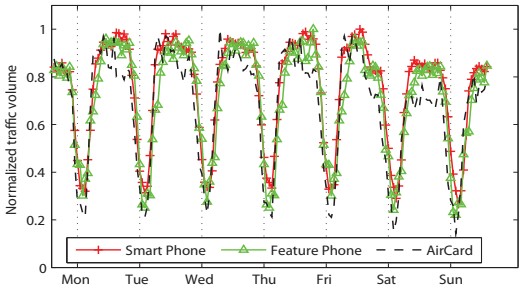


Fig. 1. Diurnal patterns of aggregate IP traffic volume for three device types: smart phone, feature phone, and air card

A. IP Traffic

1) Aggregate Traffic Volume Characteristics and Diurnal Patterns: Fig. 1 shows the normalized aggregate data traffic volume (measured in bytes) generated by each type of devices (smart phone, feature phone, and air card) at per-hour granularity for one representative week drawn from *Dataset1*. Results from other weeks demonstrate very similar characteristics, but are not shown here due to space limitation. There are clearly strong diurnal fluctuations in the traffic volume across all types of devices, and the observed fluctuations follow very similar patterns. Each day, the busy period are observed to be from 7:00 am to 9:00 pm. The lowest hourly traffic volume occurs between 2:00 am and 3:00 am. This diurnal pattern coincides with the typical work schedule of cellular users. In addition, the daily peaks observed on weekends are slightly lower than those observed on weekdays. Furthermore, lower peaks can also be seen on holidays, e.g. the Columbus Day on October 12. Overall, our findings show that whether it is a working or non-working day has a visible, but not significant, impact on the IP traffic volume generated by cellular terminals.

Fig. 2 plots the mean and standard deviation of the average per-hour IP traffic volume per individual device for different time the day for each device type. The means and standard deviations are calculated from the observed average traffic volumes observed in the same hour of each working day throughout the month of October 2011. Fig. 2 shows the same diurnal pattern of traffic volume fluctuations as observed in Fig. 1. Moreover, the standard deviations in Fig. 2 are very small, which indicate that the average per-hour IP traffic volume are relatively stable across all 31 days, and this stability is prevalent among all types of devices. Fig. 2 also reveals the difference in bandwidth/resource demand of a single device depending on its type. For example, the average IP traffic volume (hence bandwidth consumption) of a single air card is around 2.5 times compared to that of a smart phone, and around 4.5 times of what a typical feature phone generates.

Note that only session records of *data service users* are

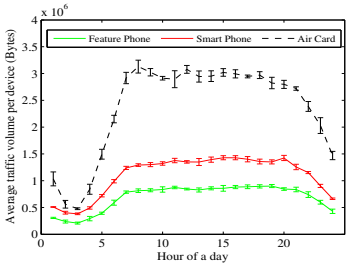


Fig. 2. Average per-hour IP traffic volume per device

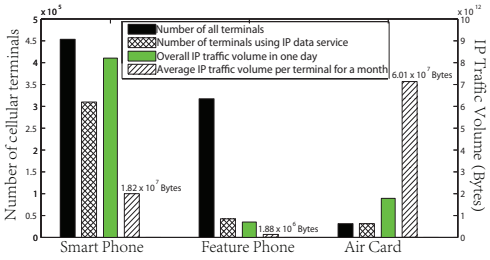


Fig. 3. CDF of the number of application categories used by individual cellular terminals in one day

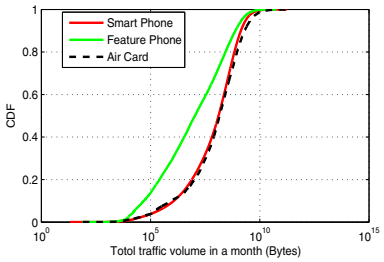


Fig. 4. CDF of IP traffic volume of individual terminals across one month

included in Fig. 2. Fig. 3 summarizes the statistics of the number of data-service subcribers and the total number of devices for each category (smart phone vs. air card vs. feature phone), along with the associated IP traffic volume. We find that 69.4% of smart phone users are data service users, and the remaining 30.6% only make phone calls or send short messages. On the other hand, only 12.37% feature phones users generate data sessions. If we take all the terminals into account, the actual ratios between the average IP traffic volume generated by a typical air card, smart phone, and feature phone are around 36.3 : 9.8 : 1. The difference is much larger than what we observe in Fig. 2. This finding has impact on how cellular carriers forecast data traffic growth (and the correspondence bandwidth demand) as the market share of different device types changes in the future.

2) Data Usage Characteristics of Individual Cellular Terminals: In Fig. 5(a), (b), and (c), we plot the ranked distribution of traffic volume with respect to individual cellular terminals for smart phone, feature phone, and air card, respectively. As reported in previous studies [16] [3], the distribution of traffic volume in cellular network roughly follows the Zipf model. In our study, we also perform a fit for the traffic volume distribution using the Zipf model, and the results are shown Fig. 5. The goodness of fit (R-Square) for smart phones, feature phones, and air cards are 0.8764, 0.8799, and 0.9056, respectively. These values indicate that the traffic volume distributions with respect to individual cellular terminals indeed possess some of the fundamental features of Zipf model, i.e. the IP traffic volume generated by a cellular terminal device is inversely proportional to its rank, and most traffic is generated by a small fraction of terminals.

A closer observation on Fig. 5, however, reveals that only the traffic volume distribution of the top 10% most active devices/users strictly follow the Zipf model. For all three device types, the distribution curve of the remaining users of lower ranks gradually deviates from the ideal Zipf model. The exponent b we obtained in the Zipf fitting for smart phones, feature phones, and air cards are -0.51, -0.56, and -0.58, which indicate that traffic volume diversity among the top 10% air cards is the most significant, followed by feature phones and smart phones, respectively.

Extending the comparison to include the other 90% less active data users, we plot the cumulative distribution function (CDF) for the IP traffic volume distribution of all cellular terminals for each device type in Fig. 4. We can see that the distributions for smart phones and air cards are similar with each other, while the CDF of feature phones shows a larger spread (diversity) of traffic volume among individual terminals. For both smart phones and air cards, the monthly traffic volume correspond to almost 80% of all terminals are within the range of 5 MBytes to 1 GBytes, while in the case of feature phones the same percentage covers a much larger range from 100 KBytes to 1 GBytes, and more than 30% of feature phones are within the range of 100 KBytes to 5MBytes. A wider range of diversity in terms of data traffic usage makes it more challenging for cellular carriers to project overall data traffic growth for resource provisioning purposes.

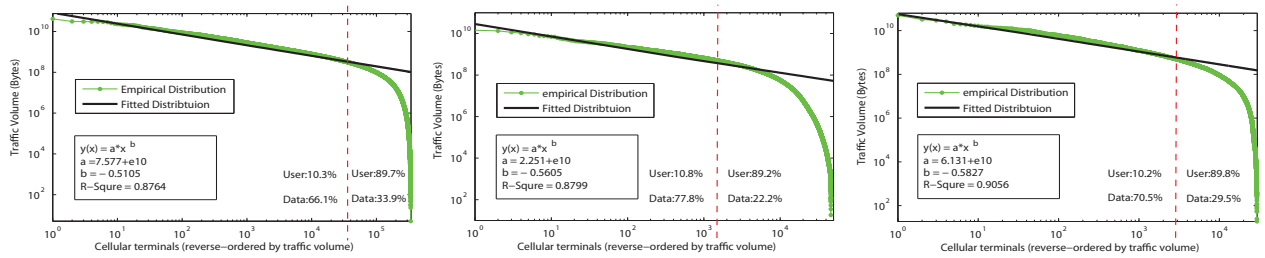
TABLE I
p-VALUES OF KRUSKAL-WALLIS TEST FOR CALL COUNT DISTRIBUTIONS

	feature phone non-data	feature phone data
smart phone non-data	0.0703	0
smart phone data	0	0.6909

B. Voice Call

We now investigate the characteristics of voice calls made by smart phone and feature phone users. Air cards are not included because they can only be used for data communication. In order to understand the potential relationship between data traffic usage and voice call behaviors, we further divide both smart phone and feature phone users into two subgroups: data users (i.e., whose data session records appeared in *Dataset1*) and non-data users. We analyze two aspects of voice calls: duration and frequency of calls. Our analysis show that the call durations follow very similar distributions for all four categories of users: smart phone users with and without data services, feature Phone users with and without data services.

We then analyze the frequency of voice calls made by these four user categories during the entire month, as shown in Fig 6. Quite interestingly, we find that the numbers of calls made by data-service subscribers, independent of the types of devices they use (smart or feature phones), follow very similar distributions. Similarly, the distributions of call frequency for non-data users are almost identical for the two device types as well. However, there is a significant difference between



(a) Smart Phone (b) Feature Phone (c) Air Card

Fig. 5. Ordered distribution of IP traffic volume of individual cellular terminals (largest to smallest) and fitted Zipf distribution

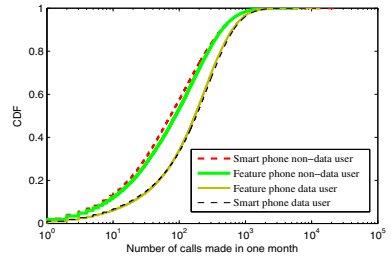


Fig. 6. The distribution of the number of voice calls made by individual cellular terminals in one month

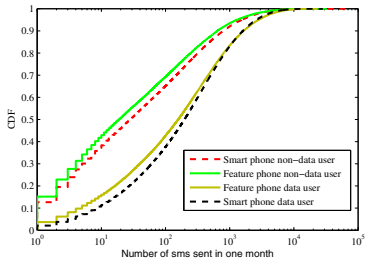


Fig. 7. The distribution of the number of short messages sent by individual cellular terminals in one month

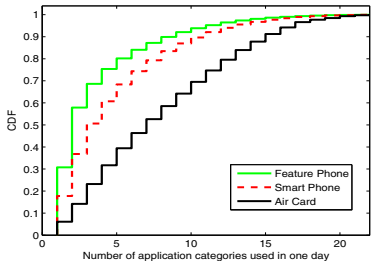


Fig. 8. CDF of the number of application categories used by individual cellular terminals in one day

data-users and non-data users, i.e., data users (who actively accessed data services over the observed month) tend to make more calls than non-data users.

To quantify our findings, we randomly select five thousand terminals from each of the four user groups, and run pairwise Kruskal-Wallis test [17] among the selected terminals. The Kruskal-Wallis test is designed to test the null hypothesis that samples in all tested groups are drawn from populations with the same distribution. We list the results of Kruskal-Wallis test (the *p-values*) in Table I. The *p-value* between non-data users using smart phones and non-data users using feature phones is 0.0703, which is not significant enough (i.e. < 0.05 by default, or < 0.01 in strict cases) to reject the null hypothesis. In other words, the call count distributions of those two user groups are statistically similar to each other. The same conclusion can also be drawn for data users using smart phones and data users using feature phones. On the other hand, we see that the tests between non-data users and data users all return 0, which means that their call count distributions can be considered as different. Our analysis results show that, the difference in observed voice call patterns is associated to whether the subscribers use data services or not, independent of what type of devices they have.

C. Short Message

Fig. 7 shows the number of short messages sent by subscribers using different types of mobile phones. Except for the slight differences between the curves of smart phones and feature phones, most short message characteristics resemble the observations we have on voice call patterns, i.e., the distributions of short message numbers have a stronger dependence on the data-service usage than on the type of devices, and data users tend to send more short messages than non-data users.

Our results in Section IV-B and Section IV-C can help cellular carriers to predict how frequently subscribers are going to make phone calls or send short messages. However, how voice call and/or short message pattern will change after non-data users switch to data-users is still unclear due to the lack of long term records, which is left for future work.

V. IP Traffic Application Breakdown

In this section, we perform a more detailed analysis of the type of data applications running on smart phones, feature phones, and air cards. Unlike previous attempts to understand cellular device usage that focus on individual data applications [16], we group applications into several categories according to the services they provide and study the functional-level properties of mobile data traffic generated by different device types. This part of our study is based on *Dataset2*.

We use a three-step method to identify the category of data applications that generate the IP traffic observed in *Dataset2*. First, we classify all data packets according to their TCP/UDP port numbers because most mobile applications use port numbers registered at the Internet Assigned Numbers Authority (IANA) [18], and applications adopting random port use, such as P2P, are not popular in cellular networks. We found that more than 70% of the cellular data communications are HTTP or HTTPS traffic. As a result, in the second step we apply a simple keyword mining [19] [20] on the destination URL of every HTTP packet for further application identification (If a URL cannot be matched with any keyword, we categorize it as “browsing”, which is the default application that HTTP

protocol is designed for). In the final step, after we identify the application category of a URL, we group all packets transmitted to or from the IP address of this URL into the same category. Using this method, we manage to identify more than 85% of all the data traffic in *Dataset2*.

A. Application Diversity

Fig. 8 shows the distribution of the number of data application categories used by an individual terminal within a day for different device types. It is obvious that air card users access a wider range of IP/data-services than smart phone and feature phone users. More than 50% of air card users access more than ten categories of data applications in a single day. A smart phone user, on average, accesses more categories of data applications than a feature phone user. These observations indicate a trend that cellular terminals will have more various data usages as the number of smart phones and air cards grows.

B. IP Traffic Breakdown by Application Categories

Fig. 9 illustrates the distributions of traffic volume with respect to different categories of data applications. Note that we list "facebook" as an independent application category because it contributes to a significant portion of traffic volume on all types of devices. We also studied how these volume distributions change throughout a day, and found that for all device types, the bandwidth shares of different application categories are roughly stable over time.

The results show that Facebook, browsing, and e-mail are very popular on all types of mobile devices. This is quite different from wired networks [21], especially for e-mail, which only accounts for a very small portion of wired traffic. Furthermore, we find that the proportions of IP traffic generated by some categories of applications, such as video, music, e-commerce and content distribution network (CDN), vary dramatically among smart phones, feature phones, and air cards due to differences in devices' capabilities. For instance, video applications generate only 5% of IP traffic on feature phones, but around 15% and 20% on smart phones and air cards, respectively. This phenomenon is possibly due to the insufficient resources and capabilities to display multi-media content on most feature phones. Meanwhile, the share of music traffic on feature phones is the highest among all devices. It might be because most feature phone users turn to music for entertainment due to the limited video function of their handsets. Moreover, the proportion of e-commence traffic on air cards is higher than on mobile phones. A possible explanation is that it is more convenient to do on-line payment on a desktop or laptop computer.

Therefore, as the smart phones and air cards become more popular, we may expect to see a larger percentage of video traffic in cellular networks. The increase of this kind of bandwidth-consuming applications may further intensify the need for larger backhaul capacity. On the other hand, the observed results also show that the bandwidth shares of several major application categories, namely facebook, e-mail, and browsing, are unlikely to change much in the future.

VI. Malicious IP Traffic in Cellular Network

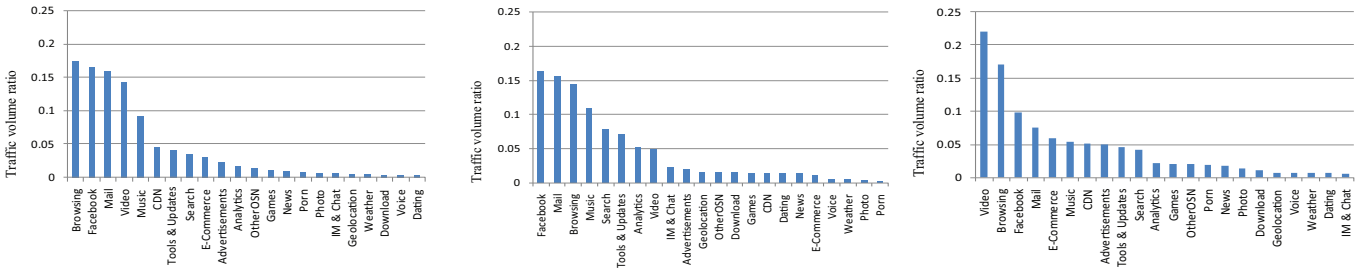
With a boost in the functionality and programmability of modern devices, there is an increasing possibility that cellular terminals can be leveraged for malicious activities targeting other hosts in the cellular network or the wide-area Internet. By identifying and analyzing port scan traffic in cellular networks (using *Dataset2*), we characterize several fundamental features of malicious traffic generated by different types of cellular devices.

We apply two methods to detect port scan traffic in our dataset. Our first method focuses on TCP SYN packets and UDP packets sent out by cellular terminals that receive no responses. These TCP SYN packets and UDP packets indicates failed connection attempts, which usually reveal malicious activities [13] [14]. We mark all the sources of non-responded packets as potential attackers, and use two thresholds to further classify potential attackers into normal users and real attackers. The first threshold is the number of failed connections initiated by a device in half an hour. This threshold can detect active attackers that generate a burst of port scan traffic in a short time. The second threshold is on the total number of failed connections a cellular device initiates in a whole day. This threshold is designed to detect the port scan traffic, such as the scans initiated by worms, that is generated at a low rate but will accumulate to a large volume over a longer period. A cellular device that meets either thresholds will be identified as an attacker. We ran a series of experiments to determine the value of both thresholds by the following two principles. First the thresholds should be as low as possible to ensure a small false negative rate. On the other hand, in order to also keep the false positive rate low, we stop further decreasing the thresholds after reaching a point where reducing the threshold even a little bit would cause a sharp increase in the number of detected attackers. According to the experiment results, we set the two thresholds to be 500 failed connections in half an hour, and 2000 failed connections in a day.

Our second method for malicious traffic detection is to monitor the packets destined for unused IP addresses. This approach has been proven to be useful for measuring a variety of abnormal behaviors [15]. In our study we identify five class-A IP blocks that are not in common use: 1.0.0.0/8, 5.0.0.0/8, 7.0.0.0.8, 37.0.0.0/8, 45.0.0.0/8, and classify all the packets sent to those IP blocks as potentially malicious traffic. We find that more than 90% of packets sent to the five unused IP blocks are also detected as port scan traffic by our first method. Other potential causes of malicious traffic, such as DoS attacks, IP address bytes-order misconfiguration, improper distribution of routing information, and abuse of P2P software, are rarely observed in our cellular network data trace.

A. Temporal Analysis

Fig. 10 shows the number of active port scanners in each day through the week during which *Dataset2* is captured. We can see that only two feature phones are detected (on Tuesday and Friday respectively), and most scanners reside on smart phones and air cards. This observation implies that the increase in the



(a) Smart Phone (b) Feature Phone (c) Air Card

Fig. 9. Traffic volume breakdown per application category for different types of devices

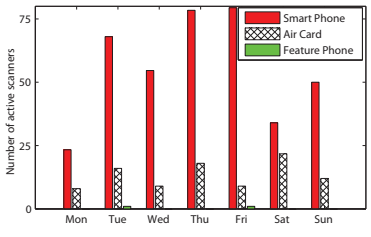


Fig. 10. Number of active scanners in each day through a week

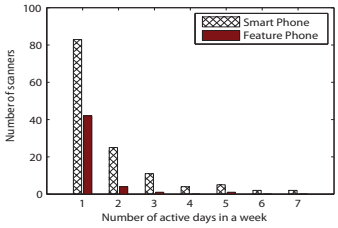


Fig. 11. Active days of scanners in a week

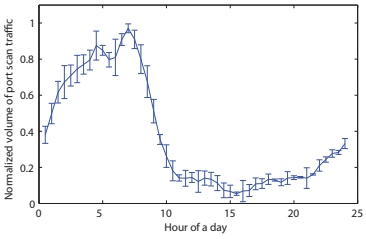
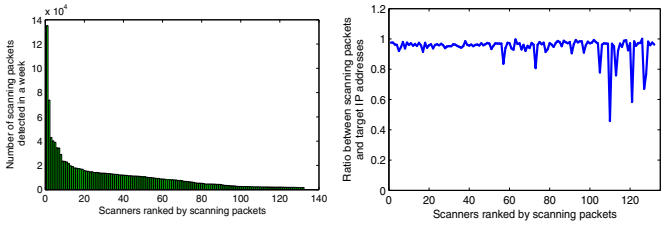


Fig. 12. The diurnal pattern of port scan traffic volume

population sizes of smart phones and air cards might bring new security threats to cellular networks. The remaining analysis in this section mainly focuses on those two device types. Fig. 10 also reveals the presence of port scans initiated by smart phones and air cards through the whole week. On average, the smart phone scanners are around 3.5 times as many as air card scanners. However, considering that the total number of air cards in *Dataset2* is only around 1/8 of the number of smart phones, the actual percentage of air cards with port scan activity (around 0.12%) is larger than the percentage of smart phone scanners (around 0.04%).

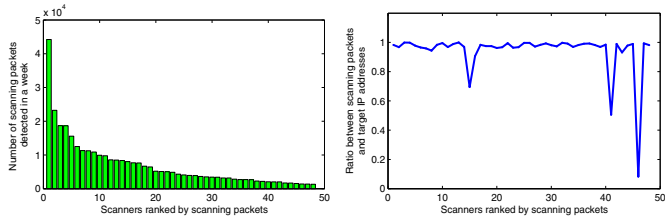
Fig. 11 illustrates the histogram of the number of active days for all port scanners. For both smart phone and air card scanners, the majority are only detected in one day throughout the whole week. However, compared with air cards, a larger portion of smart phone scanners are active for more than one day in the week. Several smart phone scanners are even detected in six or seven days while the most active air card scanner only appears in five days through the week. It seems that smart phones are more preferred platforms for long term scanning while air cards, although more likely to be used as scanners, are usually leveraged for short scanning missions.

Fig. 12 illustrates the mean and standard deviation of normalized port scan traffic volume as a function of the hour of a day. Traffic volume in every hour is normalized by the maximum hourly volume through the whole day. None of the standard deviations shown in Fig. 12 is substantial. This indicates that the diurnal pattern of port scan traffic volume is stable through the week. Additionally, port scan traffic volume changes dramatically between day and night, and the peak volume takes place at night. This feature is different from



(a) Port scan traffic volume generated by individual terminals in a week (b) Ratio between the number of scanning traffic and target IP addresses

Fig. 13. Spatial characteristics for port scan traffic initiated by smart phones



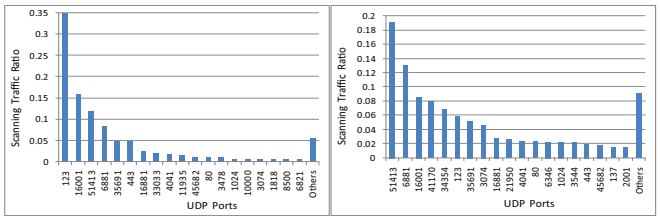
(a) Port scan traffic volume generated by individual terminals in a week (b) Ratio between the number of scanning traffic and target IP addresses

Fig. 14. Spatial characteristics for port scan traffic initiated by air cards

what we have observed on the diurnal pattern of aggregated IP traffic volume, which reaches the peak during day time. One explanation for this phenomenon is that most port scans are done at night to take advantage of the low bandwidth usage.

B. Spatial Analysis

The histograms of scanning traffic volumes generated by smart phones and air cards are illustrated in Fig. 13(a) and Fig. 14(a), respectively. For both device types, there exist a



(a) Smart Phone (b) Air Card

Fig. 15. Destination UDP ports for port scan traffic

small number of scanners responsible for a significant amount of overall port scan packets. This characteristic coincides with our previous result that a small group of scanners active in more days of a week than others. After a cross check between the terminal identities in Fig. 11, Fig. 13(a), and Fig. 14(a), we find that the scanners generating most port scan traffic are also active for most days during the week.

Fig. 13(b) and Fig. 14(b) illustrate the ratio between the number of port scan packets sent out by each scanner and the number of distinct IP addresses that are destined for. For most scanners, the total number of scan packets is very close to the number of destinations reached by those packets. This reveals that most port scanners in cellular networks are doing horizontal scans, i.e. scanners usually send only one packet to each target IP address probing a single port on that host.

Horizontal scans are usually conducted to search for specific services, backdoors, worms, trojans, or security leaks in the network. To better understand the objectives of those detected horizontal scans, we study the volume distribution of scan traffic among destination ports. More than 80% of TCP port scan packets generated by both smart phones and air cards are sent to either port 80 or port 443. This is not surprising since searching for the existence of HTTP and HTTPS services is the first step for many different kinds of attacks. Moreover, there is also a number of trojans and worms, such as "slapper" or "711 trojan", working on TCP port 80 and TCP port 443.

On the other hand, the results on UDP port scans, as illustrated in Fig. 15, seem to imply that the smart phone scanners may cause more serious security threat than the air card scanners. The top UDP port destined by smart phone scanners is the port 123, which is used by Network Time Protocol (NTP) and may provide both rich system information and possible avenue of attack for intruders. Scans on this port usually mean on-going preparation for further intrusions. Moreover, UDP port 16001, which is the second popular destination port for smart phone scanners, is also frequently used by malicious softwares and worms. On the other hand, the most popular UDP destination ports for air card canners, such as 51413 and 6881, are usually used by Bit Torrent clients to search for new peers. This kind of port scans are not very likely to cause significant security threat to the network.

VII. Conclusions

In this paper, we conduct a comparative study of cellular network traffic generated by the three types of mobile devices: feature phones, air cards, and smart phones. Our results reveal some similarities as well as differences among the traffic volume characteristics of those devices. In addition, we perform a detailed application breakdown for IP traffic in cellular networks. Lastly we study port scan activities on different types of cellular devices. Most scanners we detected reside on smart phones and air cards, suggesting that the increase in the population sizes of those devices might bring new security threats to cellular networks.

References

[1] C. S. C. (Chetan Sharma Consulting), "Us wireless market update q2 2011." http://www.chetansharma.com/USmarketupdate2011.htm.

[2] "Sierra wireless reports third quarter 2011 results." http://www.sierrawireless.com/AboutUs/investorinformation/quarterlyresults.aspx.

[3] C. Williamson, E. Halepovic, H. Sun, and Y. Wu, "Characterization of CDMA2000 cellular data network traffic," in *Conference on Local Computer Networks*, pp. 711–719, 2005.

[4] S. Mukund, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskovec, "Mobile call graphs: Beyond power-law and lognormal distributions," in *International conference on Knowledge discovery and data mining*, pp. 596–604, 2008.

[5] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *International conference on Mobile systems, applications, and services*, pp. 179–194, 2010.

[6] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones," in *Conference on Internet measurement*, pp. 281–287, 2010.

[7] G. Maier, F. Schneider, and A. Feldmann, "A first look at mobile hand-held device traffic," in *Passive and Active Measurement Conference*, pp. 161–170, 2010.

[8] M. Shafiq, L. Ji, A. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," in *International conference on Measurement and modeling of computer systems*, pp. 305–316, 2011.

[9] Y. Lee, "Measured TCP performance in CDMA 1xEV-DO network," in *Passive and Active Measurement Conference*, pp. 208–221, 2006.

[10] Y. Won, B. Park, S. Hong, K. Jung, H. Ju, and J. Hong, "Measurement analysis of mobile data networks," in *Passive and Active Network Measurement*, pp. 223–227, 2007.

[11] S. Pattaramalai, V. Aalo, and G. Efthymoglou, "Evaluation of call performance in cellular networks with generalized cell dwell time and call-holding time distributions in the presence of channel fading," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 3002–3013, 2009.

[12] Z. Xu, N. Hu, and Z. He, "Call dropping and blocking probability of the integrated cellular ad hoc relaying system," in *Global Telecommunications Conference*, pp. 1–6, 2008.

[13] J. Mikians, P. Barlet-Ros, J. Sanjuas-Cuxart, and J. Solé-Pareta, "A practical approach to portscan detection in very high-speed links," in *Passive and Active Measurement Conference*, pp. 112–121, 2011.

[14] S. Staniford, J. Hoagland, and J. McAlerney, "Practical automated detection of stealthy portscans," *Journal of Computer Security*, vol. 10, pp. 105–136, 2002.

[15] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, "Internet background radiation revisited," in *Conference on Internet measurement*, pp. 62–74, 2010.

[16] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Characterizing radio resource allocation for 3G networks," in *Conference on Internet measurement*, pp. 137–150, 2010.

[17] M. Hollander and D. Wolfe, "Nonparametric statistical methods," 1999.

[18] I. A. N. A. (the Internet Assigned Numbers Authority), "Port numbers." http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xml, 2011.

[19] R. Keralapura, A. Nucci, Z. Zhang, and L. Gao, "Profiling users in a 3G network using hourglass co-clustering," in *International conference on Mobile computing and networking*, pp. 341–352, 2010.

[20] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: connecting people, locations and interests in a mobile 3G network," in *Internet measurement conference*, pp. 267–279, 2009.

[21] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," in *Computer Communication Review*, pp. 75–86, 2010.