Title

Uncovering the Footprints of Malicious Traffic in Cellular Data Networks

Permalink
https://escholarship.org/uc/item/10k6c6hb

ISBN
978-3-319-15508-1

Authors
Raghuramu, Arun
Zang, Hui
Chuah, Chen-Nee

Publication Date
2015

DOI
10.1007/978-3-319-15509-8_6

Uncovering the Footprints of Malicious Traffic in Cellular Data Networks

Arun Raghuramu[1], Hui Zang[2], and Chen-Nee Chuah[3]

[1] Department of Computer Science, University of California, Davis,
araghuramu@ucdavis.edu
[2] Guavus Inc., San Mateo, CA
hui.zang@guavus.com
[3] Department of Electrical & Computer Engineering, University of California, Davis,
chuah@ece.ucdavis.edu

Abstract. In this paper, we present a comprehensive characterization of malicious traffic generated by mobile devices using Deep Packet Inspection (DPI) records and security event logs from a large US based cellular provider network. Our analysis reveals that 0.17% of mobile devices in the cellular network are affected by security threats. This proportion, while small, is orders of magnitude higher than the last reported (in 2013) infection rate of 0.0009%. We also perform a detailed comparison of infection rates of various mobile platforms and show that platforms deemed to be more secure by common opinion such as BlackBerry and iOS are not as safe as we think. However, Android still remains the most affected platform with an infection rate of 0.39%. We present a detailed discussion of the top threat families targeting mobile devices observed in our dataset. Lastly, we characterize the aggregate network footprint of malicious and benign traffic in the cellular network and show that statistical network features can be used to distinguish between these traffic classes.

1 Introduction

The pervasive use of mobile devices such as smartphones to access an array of personal and financial information makes them rich targets for malware writers and attackers. Studies have revealed threats and attacks unique to mobile platforms, such as SMS and phone call interception malware [1]. The claims about prevalence of mobile malware were recently disputed when Lever et. al [2] showed that mobile malware appears only in a tiny fraction (0.0009%) of devices in their dataset, indicating that mobile application markets are providing adequate security for mobile device users. However, the work in [2] failed to provide a comprehensive view of malicious network traffic since the analysis was limited to threats which issue DNS requests to known malicious domains. Also, [2] did not quantify the prevalence of specific types of threats affecting the network in their characterization study.

In this paper, we performed a detailed characterization of malicious traffic generated by mobile devices using deep-packet flow records and security event logs from a major US-based cellular network. Our analysis revealed that 0.17% of over 2 million devices in the cellular network triggered security alerts. This fraction, while still small, is *orders of magnitude higher* than the previous infection rate reported in [2] and is in agreement with recent direct infection rate measurements focusing on the Android platform [3]. This alarming infection rate calls for a more careful and thorough study of malicious traffic in the mobile ecosystems.

A second area of our focus deals with the problem of 'detecting' malicious hosts/URLs. Previous studies such as [4, 5] treat this as a supervised learning problem where a classifier learns on a combination of DNS, WHOIS, lexical, and other features associated with a given host to decide whether it is malicious or benign with high accuracy. Other studies such as [6, 7] exclusively utilize lexical features to achieve similar goals. A different approach, Nazca [8], was proposed recently to detect malware distribution networks by tracking web requests associated with malware downloads and installations.

Instead of focusing on features associated with the malware or hosts (e.g., URL content, WHOIS,etc.), we examined the network-level statistical features of traffic associated with malicious domains. We observed that there are distinctive network access patterns that can be leveraged to distinguish between benign and malicious sites. To the best of our knowledge, this is the first study that applies such network-level features to the malicious host identification problem.

The contributions of our work are two-fold:

a) We provide a large-scale characterization of malicious traffic by analyzing DPI records and security alerts of over 2 million devices. Apart from revealing higher infection rate, we show that four classes of threats: privacy-leakage, adware, SIP attacks and trojans - are most prevalent in mobile devices. Also, we find that 0.39% of Android devices are infected, while the infection rates of BlackBerry and iOS devices which are commonly considered more secure are observed to be comparatively high (0.32% and 0.22% respectively).

b) We analyze the aggregate network-level features of user traffic for both malicious and benign domains, and demonstrate that they are sufficiently distinct. This allows us to build a machine learning classifier that identifies malicious domains utilizing statistical properties of network traffic. We believe that this opens up an interesting direction for detection of unknown malicious domains.

The remainder of the paper is organized as follows: Section 2 provides an overview of our datasets and methodology. In Section 3 we present the findings of our characterization study of mobile threats. Section 4 describes the nature of network footprints of malicious traffic. Section 5 concludes the paper.

Data source	**Alert triggering event(s)**
IDS-1	DNS requests seen to known malicious domains
IDS-2	(a) The HTTP request header contains a known malicious user agent string or URI (b) Leakage of IMEI, IMSI, phone number or location information through a HTTP header or URI. (c) Attempts to connect to a known C&C server. (d) DNS request to a known malicious domain (Utilizes a different set of malicious domains from IDS-1). (e) Known malicious behavior. Eg. Attempt to trigger a DDoS, replay attack, etc.
AV-1	Known malware detected on a device through a signature.

Table 1: Security data sources and their alert triggering mechanism.

2 Data Summary & Methodology

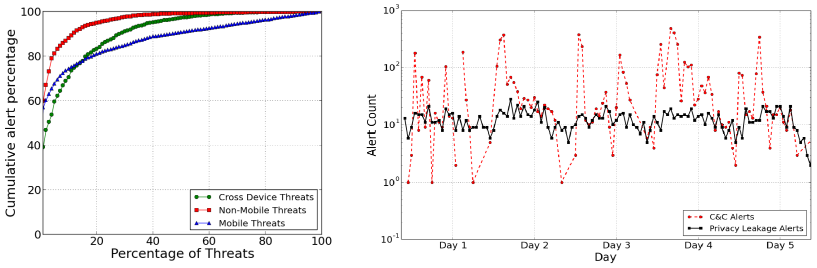
Our dataset, collected at a distribution site operated by a US cellular service provider, is multiple terabytes in size and logs HTTP activities of over two million subscribers for a week-long period in summer 2013. What makes the dataset more interesting is the associated security alert logs generated by commercial systems deployed in the network.

Specifically, the following traces are contained in our dataset:

- Deep Packet Inspection (DPI) Records: These records log HTTP activity of subscribers in the network and contain flow level information associated with each HTTP request, such as, the timestamp, duration, bytes transmitted in each direction, source IP address, URL, and User Agent of the flow.
- Intrusion Detection System (IDS) and Anti-Virus (AV) Alert Logs: These logs contain threatname (usually vendor specific), subscriber IP address, timestamp, destination HTTP domain, and destination port of the alerted activity.
- IP Assignment Records: These records map dynamically assigned IP addresses to anonymized subscriber device IDs.
- VirusTotal, McAfee scan results: We performed additional scans on certain domains and IP's in the IDS and AV logs to obtain additional information about the threats and number of malware detection engines flagging it as positive (malicious).

We perform two processing tasks to help characterize malicious events in the carrier's network. We describe each of these tasks in greater detail below.

a) *Building Ground Truth for Malicious Traffic:* As mentioned earlier, the carrier deploys two separate commercial IDS's in its premises. Each IDS passively monitors different characteristics of traffic and flags security events without initiating any 'block' actions. We utilize logs produced by these appliances in our characterization study. We also use records logged at AV scanners deployed at select end-client devices as an additional auxiliary source of security evidence. Table 1 describes the alert triggering mechanism of these IDS and AV systems. We collect IP's and URL's associated with the alert events and submit them to



(a) Threat Alerting Behavior

(b) Timeseries of Privacy Leakage and Botnet Communication Alerts

Fig. 1: Macroscopic characterization of alert data

commercial URL scanners such as VirusTotal [9] to eliminate false positives and to gather detailed information about the threats associated with these alerts. In addition, we manually group the most prominent threats in the network into four general categories or "Threat classes" as: Trojans, Privacy leakage threats, Potentially Unwanted Applications(PUA) and SIP threats based on the common characteristics and infecting behavior of the threats.

b) Identifying Devices and Platforms: The events in our malicious traffic ground truth database could have been triggered by either mobile devices such as smartphones and tablets or laptops and desktops that connect to the cellular network via hotspots/modem devices. We were provided with the registered make, model and operating system information for about half of the anonymized subscribers in the trace. For the other subscribers, we infer the device type, make, and OS type using the User-Agent fields from their DPI records with the help of an in-house tool[4]. The devices in our alert datasets are then classified manually as one of the four general categories: phones, tablets, hotspots/modems and other devices.

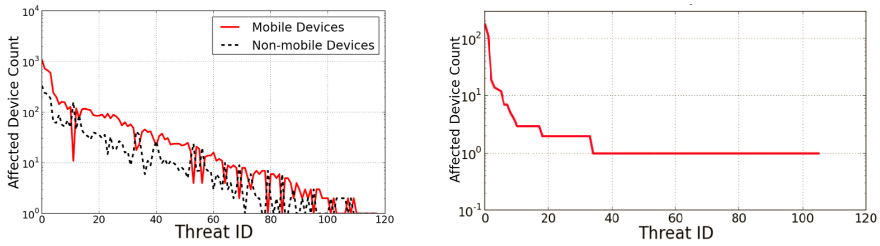
3 Characterizing Mobile Threats

3.1 Prevalence of Malicious Traffic

As described earlier, we do not limit our characterization to web traffic generating DNS requests to malicious domains. Instead, we include non-HTTP malicious traffic such as VoIP security events occurring on ports 5060 and 5061 and a number of security events on non-standard ports such as 8080/8090 in our study. Thus, we capture a more complete view of malicious traffic in the cellular network.

In the dataset, 0.23% of devices were observed to trigger security alerts and 73.2% of these events originated from mobile devices such as smartphones and tablets while the rest are triggered by devices behind wireless hotspots or modems,

[4] This utility analyzes every User-Agent string in the DPI trace associated with the unknown device to make an estimate of its make, model and platform.



(a) Cross-Device Threats

(b) Mobile Only Threats

Fig. 2: Infection Effectiveness of Threats

and hence cannot be uniquely identified as being mobile or non-mobile. This puts the lower bound of the overall infection rate of mobile devices at 0.17%, which is orders of magnitude higher than those reported in the most recent work by Lever et al [2]. Also, our observed infection rate is in agreement with the reported rate in a recent study focusing on direct measurement of Android malware infection rates [3]. We note that our notion of 'infection' is similar to that of [2]. We mark a device as infected when we observe a security alert originating from it. This is reasonable since *a*) The IDS systems in the network are passively monitoring threats and do not engage in 'blocking' malicious traffic, *b*) We only use alerts which are known true positives and *c*) This allows us to do a one-to-one comparison of infection rates with previous work.

Further, we rank the individual infected devices based on the total number of security alerts generated by them over the course of the week, and found that the top 20% of the devices account for more than 80% of the security alerts. Interestingly, the top 20% of the infected devices primarily consisted of Android and iOS based phones/tablets.

Based on the methodology described in Section 2, we extracted detailed information about the threat associated with each security event by leveraging commercial virus-scanning tools, and through manual inspections. We found 327 unique threats in our *malicious traffic groundtruth* dataset that spans over the course of one week. After performing device classification, we further categorized these 327 threats into three classes with 75% confidence intervals as follows: (a) mobile-only threats that infect mobile devices (b) non-mobile threats that infect non-mobile devices, and (c) cross-device threats that infect both types of devices. Figure 1a characterizes the macroscopic alerting behavior of the three classes of threats in the network. The x-axis in this graph represents the top n% of threats in terms of the total number of alerts generated. In general, a small fraction of threats (5-15%) are responsible for a major proportion (over 80%) of the observed alert traffic. However, we note that mobile threats in general tend to generate less number of alerts than their non-mobile counterparts. This might indicate that attackers have adapted mobile malware to be stealthier and harder to detect on the network. Moreover, some mobile-specific threats (e.g., privacy leakage) generate less network footprints and hence trigger less number of alerts.

Threat Class	Threat Description	Unique Threats	# Mobile	# Non-mobile & Unknown	# Associated IPs	Associated Ports
Trojans	Malware which utilizes techniques of social engineering, drive-by download & advanced rootkits to affect user devices	8	1669	470	159	53
Privacy leakage	Leakage of sensitive information such as IMEI number & user location	2	1277	418	77	8080, 80
Adware & PUA	HTTP Requests to known adware domains & Requests with known malicious UA strings	3	1179	368	45	80
SIP threats	Illegal session information modification & Replay attacks on SIP protocol	2	161	98	21	5060, 5061

Table 2: Top categories of prevalent mobile malware

Exploring this further, we see that the number of alerts observed to be generated per threat is a function of the threat family (e.g. botnet, data leakage, etc.) and the number of devices affected by the threat. Privacy leakage threats such as threats responsible for leaking IMEI or location information from a device generally do not generate as many alerts as devices affected by a botnet threat (as shown in Figure 1b). A 'zombie' bot device makes regular call-backs to command and control servers for downloading instructions, data exfiltration and so on, hence generating a much larger footprint in the security alert logs. This implies that mining alert logs generated by network access activities could be effective in early detection and prevention of botnet-like threats. However, similar methodology will be ineffective for other threats, such as data leakage, that leave very little footprints.

3.2 Top Mobile Threats

Next, we examine the threats that infected the most number of *mobile* devices. Malware writers often aim to infect as many devices as possible in order to maximize their financial or other gains. Therefore we use the number of devices affected by a threat to quantify its success in the cellular network.

Figure 2 plots the infection effectiveness of two categories of threats: cross-device threats and mobile-only threats, respectively. The x-axis plots threat id in decreasing order of rank based on the total number of devices affected (i.e., the first threat id affects the most number of devices). Notice from this graph that only a few threats are able to successfully affect a large number of devices (either non-mobile or mobile). To better analyze the nature of these prominent threats, we further classify the top 15 threats (either mobile-only or cross-device threats) affecting the most number of mobile devices in the network into four different classes based on unique characteristics exhibited by each threat as shown in Table 2. We now describe the characteristics of each of these malware categories and how they affect end users:

Trojan Threats: These programs deliberately cause harm to a user device while posing to be a benign application such as a free anti-virus solution. The harm can be either in terms of allowing unauthorized remote access to the device,

Type of data	Affected Devices
IMEI number	757
Device Location	603
Phone number	14
Call Logs	5
SMS Logs	1

Table 3: Types of Privacy Leakage

hijacking device resources, turning the device into a bot/proxy, stealing user information etc. This class of malware is observed to be the most effective form of threat currently affecting mobile devices. Interestingly, through the course of our analysis, we detected instances of the Zeus trojan affecting 82 distinct iOS based mobile devices in the network. Although mobile variants of this threat affecting other platforms such as Windows Mobile and Android have been seen in the wild, to the best of our knowledge, this is the first time a variant of this threat was identified affecting iOS devices [10]. Unfortunately, we were not able to explore characteristics of this malware variant further due to limitations in the dataset.

Privacy Leakage Threats: Threats which maliciously leak IMEI (International Mobile Equipment Identity) number or device location information in the HTTP headers or URI affect over 1200 unique mobile devices, making this one of the most prevailing attacks targeting mobile devices in our dataset. Although traditional desktop malware which leak sensitive user data exist, this problem is more pronounced in the mobile ecosystem. This may be due to the sensitive nature of data stored on mobile devices which attackers deem valuable, issues of application over-privilege in some mobile platforms, and the availability of third party app stores which facilitates deploying such malicious applications. Table-3 categorizes the types of privacy leakage issues revealed in our ground truth data. Clearly, information such as those presented above would potentially allow an attacker to uniquely observe a targeted user and his activities, making this a serious violation user privacy.

Adware & Potentially Unwanted Applications (PUA): This class of applications sneak into a device deceptively and get installed in such a way that it can be difficult to detect and remove. The primary motive of these programs is to display unwanted advertisements to users, often in the form of pop-up ads. While some of these apps may just be a minor irritant to the user, they may, in some cases, also act as dangerous spyware that monitor user behavior and collect data without consent.

SIP Threats: Session Initiation Protocol (SIP) is widely used for controlling multimedia communication sessions such as VoIP calls over the internet. Our results indicate that vulnerabilities in this protocol is seen to be a popular target for attackers seeking to exploit mobile devices. These are alarming trends since such vulnerabilities can potentially give attackers the ability to listen-in on confidential voice communications or launch denial of service attacks as reported in previous studies [11, 12].

Device Platform	**% Total devices**	**% Infection Rate**	**% Mobile Alerts**
iOS	40.57%	0.22%	53.12%
Android	20.09%	0.39%	45.74%
Windows	0.2%	0.12%	0.76%
RIM OS	0.08%	0.32%	0.15%
Custom Feature Phone OS & Others	39.06%	0.0009%	0.21%

Table 4: Affected mobile platforms

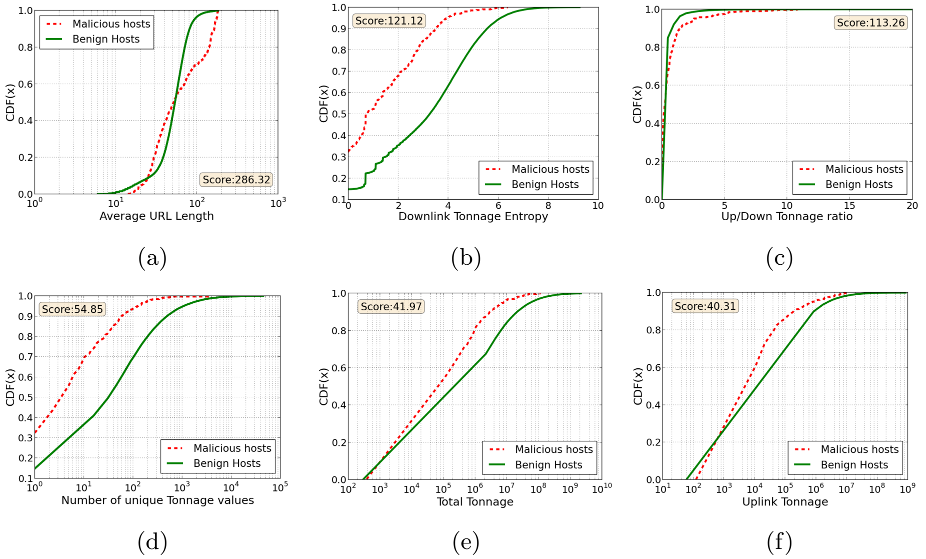
3.3 Infection rates of popular mobile platforms

The question of which mobile platforms are most vulnerable to security threats has been a hot topic of debate for several years. We attempt to answer this question by utilizing ground truth data obtained from the operational cellular network. Table [4] presents the following data points: *a*) The proportion of devices belonging to each identifiable mobile platform in our dataset; *b*) The proportion of devices of a given platform that are infected, or the infection rate; and, *c*) The proportion of alerts observed in the ground truth originating from a given platform.

We observe from the second column of the table that Android is the most vulnerable platform with a 0.39% infection rate. This infection rate is slightly higher than those claimed by the most recent independent study of malware infection rates in Android by Truong et al [3] who measure it to be in the range of 0.26-0.28% and three times the rate reported by Google [13]. Android is followed closely by Blackberry with an infection rate of 0.32% and iOS with 0.22%. These figures show that the walled garden approach / security through obscurity as employed by these platforms are failing to ensure against malware spread. Blackberry devices are often used for business purposes due to their security capabilities. However, the nature of data stored on these devices may induce attackers specifically target this platform which can explain its high infection rate. Attackers are however failing to affect a large proportion of users with devices running Windows based mobile platforms as noted by recent industry reports [14].

4 Network Footprints of Cellular Threats

In this section, we investigate if network access patterns associated with malicious domains/hosts contacted by infected user devices exhibit distinct statistical features when compared to accesses to their benign counterparts. There are many existing studies that target accurate detection of malicious domains/URL's by using different methodologies. Some of these studies utilize a combination of DNS and WHOIS features, host based features, content of the webpage, etc in order to achieve their goals [4, 5] while some other studies such as [6] and [7] exclusively use lexical features. The motivation of our study however is to investigate



(a) (b) (c)

(d) (e) (f)

Fig. 3: Network Footprint of Malicious Domains

if any of the statistical network features can complement the existing detection rules. This can be helpful in situations where other data such as DNS, WHOIS, etc. which are useful for the malicious domain classification task is infeasible to obtain or is otherwise unavailable.

4.1 Feature Extraction and Selection

In order to perform our classification experiment, we first build a set of known malicious domains using information from the ground truth alert database. We then create a set of benign domains by randomly choosing domains visited by subscriber devices which are otherwise not listed in the ground truth database. We further verify they are benign by running the domains through commercial URL scanners. For these set of known malicious and benign domains, we extract lexical and statistical network features as follows:

a) *Lexical features*: Each target domain/host name is broken into multiple ‘tags’ or ‘tokens’ based on the ‘.’ delimiter. We identify 6,729 such unique lexical tags through this process over a set of 1200 benign and malicious domains. We then utilize the frequency of occurrence of each tag in a given domain name as the lexical features of the target. This approach to represent lexical information is commonly referred to as the bag-of-words model. Variants of this model have been used to generate lexical features for use in detecting malicious URL’s in previous studies such as [4, 7].

b) *Statistical Network Features*: Using the DPI records from the cellular carrier we extract the following 12 heuristic features for each target domain: Uplink data transfer volume(or uplink tonnage), downlink data transfer volume (or downlink tonnage),ratio of uplink / downlink tonnage, total tonnage, proportion of failed connections, average URL length, number of connections, number

Data set	**ROC Area-α**	**ROC Area-β**	**ROC Area-γ**
600 malicious 600 benign	0.843	0.744	0.9
540 malicious 600 benign	0.83	0.737	0.897
480 malicious 600 benign	0.838	0.732	0.895
420 malicious 600 benign	0.84	0.73	0.891
360 malicious 600 benign	0.852	0.746	0.897
300 malicious 300 benign	0.84	0.703	0.885
240 malicious 600 benign	0.813	0.703	0.885
180 malicious 600 benign	0.796	0.771	0.857
120 malicious 600 benign	0.824	0.76	0.869
60 malicious 600 benign	0.763	0.728	0.876

Table 5: Comparing ROC Areas

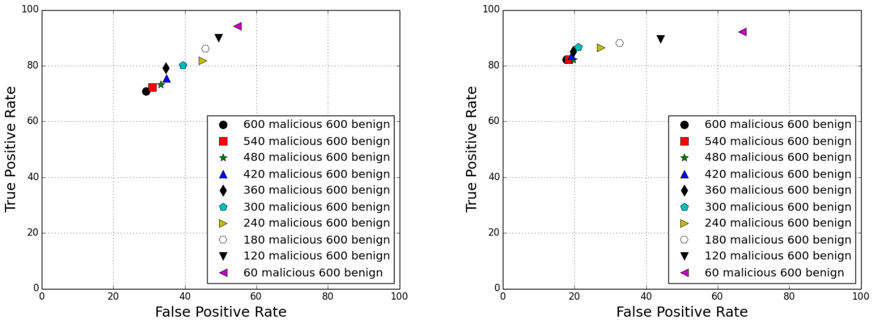
of unique source IP's connecting to the domain, number of failed connections, entropy of destination IP addresses,downlink tonnage entropy and the number of unique tonnage values.

We start our analysis by identifying specific network and lexical features that contribute towards distinguishing between malicious and benign hosts. In order to select such features, we utilize the raw set of attributes described above and apply the Chi-squared statistic evaluation [15]. The Chi-squared score essentially measures the difference between the conditional distributions of a network feature associated with the two classes: malicious vs. benign domains/hosts. On the basis of the results of this exercise, we narrow down our feature set to 53 distinct attributes associated with each malicious/benign domain after removing attributes which have a score of zero. This reduced feature set includes 10 statistical network features and 43 distinct lexical features. Figure 3 shows the cumulative distribution function (CDF) of six selected network features associated with malicious and the benign hosts that exhibited the highest chi-squared scores. It is visually apparent that there is significant difference between the conditional distribution for malicious vs. benign domains/hosts for these network features. Other network features which were selected but not shown include the connection entropy, the destination IP entropy, the downlink tonnage and the number of unique tonnage values.

4.2 Classification of Malicious/Benign Domains

Many of the statistical network features have complex non-linear relationships. This makes the task of classification of domains/hosts into malicious and benign categories non-trivial. To tackle this problem, we use a machine learning approach which can handle such dependent features efficiently. In particular we use the "Random Forest" ensemble learner [16] to create a model with the individual features. This classification method operates by constructing multiple decision trees at a time (15 in our case) and predicts a class by aggregating the predictions of the ensemble. In addition, we use the n-fold cross validation technique to evaluate the accuracy of our model (n=10).

We run our classification experiments on varying proportions of malicious and non-malicious hosts employing *a*) Statistical network features alone (α), *b*) Lexical Features alone (β) and *c*) Statistical network features in addition to lexical features (γ). Figure 4a and Figure 4b present the receiver operating



(a) Only Lexical features

(b) Lexical & Network Features

Fig. 4: Cross-Validation Results

characteristic(ROC) for two of our cross-validation experiments. The ideal ROC would lie close to the upper-left corner with false positive rate close to 0% and true positive rate close to 100%. Note that with the addition of statistical network features to simple lexical features, we obtain a better true positive rate at lower false positive rates for most combinations of malicious and benign hosts.

We observe from Table 5 that the ROC area is higher in the case where we utilize statistical network features along with lexical features (column 3) to perform classification as compared to using the lexical features alone (column 2) or statistical network features alone (column 1) for all proportions of malicious and benign domains. These preliminary results show that statistical network features are complementary to lexical features and hold promise to add to capabilities of existing detection rules to help solve the malicious domain detection problem. A deeper analysis of this result is warranted and we leave it as an important area of exploration for our future work.

5 Conclusions

In this paper, we present a study of malicious mobile traffic by using data obtained from a major US based cellular carrier spanning a one week period that contains over two million subscribers. Our investigation reveals that 0.17% of mobile devices are affected by security threats. This infection rate while still small, is orders of magnitude higher than the last reported infection rate of 0.0009% making this a worrisome problem. We combine multiple disparate data sets to uncover details about the threats affecting mobile devices in the cellular network and their unique characteristics. We also perform a detailed analysis of infection rates in various popular mobile platforms. Our results show that platforms deemed to be more secure by common opinion as iOS and BlackBerry are not as secure as we think. However, Android still remains the most affected platform with an infection rate of 0.39%. Lastly, we characterized the aggregate network footprint of malicious and benign domains associated with the threats observed in our dataset and showed how statistical network features can be used to potentially aid detection of malicious domains/hosts when used in conjunction with other lexical feature sets. Our preliminary results in this direction are promising and we leave more detailed analysis to future work.

Acknowledgement

We would like to thank Parth H. Pathak for his insightful comments on this work. We are grateful to Theo Pan and Josh Vaughen for their help with obtaining VirusTotal data. This work was supported in part by the the Intel Science and Technology Center for Secure Computing.

References

1. X. Wei, L. Gomez, I. Neamtiu, and M. Faloutsos, “Malicious android applications in the enterprise: What do they do and how do we fix it?” in *Proc. 28th IEEE International Conference on Data Engineering Workshops (ICDEW)*, 2012.
2. C. Lever, M. Antonakakis, B. Reaves, P. Traynor, and W. Lee, “The core of the matter: analyzing malicious traffic in cellular carriers,” in *Proc. NDSS ’13*, 2013.
3. H. T. T. Truong, E. Lagerspetz, *et al.*, “The company you keep: mobile malware infection rates and inexpensive risk indicators,” in *Proc. 23rd international conference on World Wide Web*, 2014, pp. 39–50.
4. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, “Identifying suspicious URLs: an application of large-scale online learning,” in *Proc. 26th ACM Annual International Conference on Machine Learning*, 2009, pp. 681–688.
5. H. Choi, B. B. Zhu, and H. Lee, “Detecting malicious web links and identifying their attack types,” in *Proc. 2nd USENIX conference on Web application development*, 2011.
6. A. Blum, B. Wardman, T. Solorio, and G. Warner, “Lexical feature based phishing URL detection using online learning,” in *Proc. 3rd ACM workshop on Artificial intelligence and security*, 2010, pp. 54–60.
7. A. Le, A. Markopoulou, and M. Faloutsos, “Phishdef: URL names say it all,” in *Proc. IEEE INFOCOM*, 2011, pp. 191–195.
8. L. Invernizzi, S. Miskovic, *et al.*, “Nazca: Detecting malware distribution in large-scale networks,” in *Proc. NDSS ’14*, 2014.
9. “The virustotal online scanner,” www.virustotal.com/en/about.
10. D. Maslennikov, “Zeus in the mobile - facts and theories,” www.securelist.com/en/analysis/204792194, 2011.
11. S. El Sawda and P. Urien, “SIP security attacks and solutions: A state-of-the-art review,” in *IEEE Information and Communication Technologies, ICTTA’06. 2nd*, 2006, pp. 3187–3191.
12. D. Geneiatakis, T. Dagiuklas, *et al.*, “Survey of security vulnerabilities in session initiation protocol,” *IEEE Communications Surveys and Tutorials*, vol. 8, no. 1-4, pp. 68–81, 2006.
13. S. M. Patterson, “Contrary to what you’ve heard, Android is almost impenetrable to malware,” qz.com/131436/contrary-to-what-youve-heard-android-is-almost-impenetrable-to-malware, 2013.
14. “Cisco 2014 annual security report,” www.cisco.com/web/offers/lp/2014-annual-security-report/index.html.
15. H. Liu and R. Setiono, “Chi2: Feature selection and discretization of numeric attributes,” in *IEEE 7th International Conference on Tools with Artificial Intelligence*, 1995, pp. 388–391.
16. L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.