Beyond the Radio: Illuminating the Higher Layers of Mobile Networks

Narseo Vallina-Rodriguez*, Srikanth Sundaresan*, Christian Kreibich*†,
Nicholas Weaver*‡, Vern Paxson*‡

*ICSI, †Lastline, ‡UC Berkeley

{narseo,srikanth,nweaver}@icsi.berkeley.edu, {christian,vern}@icir.org

ABSTRACT

Cellular network performance is often viewed as primarily dominated by the radio technology. However, reality proves more complex: mobile operators deploy and configure their networks in different ways, and sometimes establish network sharing agreements with other mobile carriers. Moreover, regulators have encouraged newer operational models such as Mobile Virtual Network Operators (MVNOs) to promote competition. In this paper we draw upon data collected by the ICSI Netalyzr app for Android to characterize how operational decisions, such as network configurations, business models, and relationships between operators introduce diversity in service quality and affect user security and privacy. We delve in detail beyond the radio link and into network configuration and business relationships in six countries. We identify the widespread use of transparent middleboxes such as HTTP and DNS proxies, analyzing how they actively modify user traffic, compromise user privacy, and potentially undermine user security. In addition, we identify network sharing agreements between operators, highlighting the implications of roaming and characterizing the properties of MVNOs, including that a majority are simply rebranded versions of major operators. More broadly, our findings highlight the importance of considering higher-layer relationships when seeking to analyze mobile traffic in a sound fashion.

Categories and Subject Descriptors

C.2.1 [**Computer-communication networks**]: Network Architecture and Design; C.2.3 [**Computer-communication networks**]: Network Operations

General Terms

Design, Measurement, Performance

Keywords

Mobile Networks; Cellular Networks; Middleboxes; HTTP Proxy; DNS; PEP; Mobile Traffic; HTTP Header Injection; Privacy; Android; Measurement

MobiSys'15, May 18–22, 2015, Florence, Italy.

http://dx.doi.org/10.1145/2742647.2742675.

1. INTRODUCTION

Cellular networks are complex, and this complexity extends across multiple layers: from the radio link, where a host of different access technologies compete, to the application layer, which reflects influences such as the effectiveness of DNS lookups and their interaction with Carrier-Grade NATs, and beyond, where regulatory agencies control spectrum allocations that determine carrier coverage. Economic factors, such as roaming agreements between operators, operators trying to monetize user traffic, and the rise of new business models such as Mobile Virtual Network Operators (MVNOs), further muddy the waters. While a large body of work has focused on different aspects of cellular networks, we still lack a holistic understanding of the factors that affect performance, and thus users and to some extent regulators remain in the dark regarding fully understanding user performance, privacy, and security.

Mobile operators deploy and configure their networks in multiple ways: they deploy middleboxes to alter traffic, enhance performance [32], or monetize user traffic using targeted advertising [47]. Operators also establish network sharing agreements with other operators, which leads to their users traversing other networks at least part-way when they reside outside "home" coverage area. All of this typically occurs without user intervention and awareness; indeed, the user has little say in how the operator handles their traffic. Only a small fraction of users take advantage of trusted VPN clients and servers to avoid middlebox interference. At the other end of the spectrum, even aspects technically within the user's control—in particular the APN settings that define how the user connects to the network—potentially affect performance and service quality without their knowledge.

In this work we present a characterization of middlebox behavior and business relationships in cellular networks. In doing so we aim to develop perspectives beyond the substantial body of work that has analyzed network performance using fine-grained metrics such as latency and throughput [36], characterized Carrier-Grade NATs [43], and evaluated DNS performance [33]. We analyze 70 mobile operators in the USA, Canada, Australia, France, Germany, and the UK. We collected our crowd-sourced dataset from Android users running Netalyzr, our comprehensive network troubleshooting tool that gathers a host of measurements regarding cellular networks beyond radio performance. Drawing upon this dataset, we characterize the behavior of network middleboxes such as HTTP proxies and DNS resolvers, and identify business relationships and operational models, including how they affect user traffic, service quality and privacy. We make the following contributions:

1. We demonstrate the widespread use in cellular networks of middleboxes such as HTTP and DNS proxies, many of which are imposed in a manner invisible to users. These middleboxes can potentially modify user traffic, compromise user

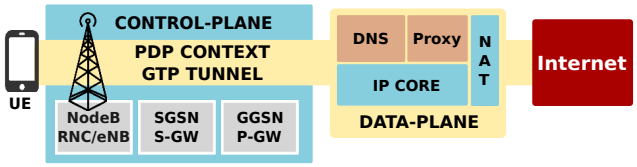

Figure 1: Schematic representation of an MNO's network deployment connected to the Internet, including control- and data-plane. The GTP tunnel hides the control-plane elements.

privacy by injecting HTTP headers that uniquely identify users or reveal their location, and in some cases appear to expose users to security threats by running unpatched software. We show how APN diversity within the same network complicates mobile networks analysis and causes service variability for customers.

2. We identify network sharing agreements between operators, highlighting the implications of roaming and characterizing MVNOs based on their relationship with the parent Mobile Network Operator (MNO). The nature of this relationship has significant implications for analyzing observed network properties and the quality of service delivered to the end user. We show that a majority of MVNOs are simply rebranded versions of major operators; as a result, MVNO subscribers can be vulnerable to inefficiencies and security and privacy issues present in the host MNO network.

This paper describes the implications of middlebox deployment in mobile networks, and business relationships between mobile operators impose on mobile users, researchers, developers, and content providers alike. Our broader theme is that efforts to measure and analyze mobile traffic need to take these dynamics into account to avoid misidentifying the root causes or over-generalizing from limited measurements in an unsound fashion.

2. MODERN MOBILE NETWORKS

Modern cellular networks differ from other access technologies in fundamental ways. In this section, we provide an overview of the technical and economic aspects of the current mobile network ecosystem. We provide a brief introduction of how cellular networks operate and then sketch the popular business models that underpin the mobile market, and how these two combine. Table 1 lists the notations we use in the paper.

2.1 Cellular Network Infrastructure

A primary distinguishing factor for cellular networks is the presence of a clearly delineated control plane. For security and management reasons, this control plane typically remains transparent to the user, and is difficult to monitor and probe directly. While our earlier work has demonstrated how to monitor certain control-plane elements, this requires privileged access to the network or hardware [37]. The data plane tunnels directly to the IP core. We describe both planes in detail.

2.1.1 The Control Plane:

The control plane consists of two logical components:

- *Radio Access Network (RAN).* The handset connects to the base station (Node B in 3G networks), which in turn is controlled by the Radio Network Controller (RNC). In 4G LTE networks, the RNC and Node Bs combine into an enhanced Node B (eNB); this has the advantage of reducing latency to the handset. The RNC is primarily responsible for managing spectrum and battery usage of handsets.

- *Support Nodes.* The Serving GPRS Support Node, or SGSN (S-GW in LTE), is responsible for billing, authentication, mobility management, and relaying packets between the base stations under its control and the gateway (Gateway GSN, or GGSN, in 3G; Packet Gateway, or P-GW, in LTE). It also interconnects decoupled 2G, 3G, and 4G deployments for a given mobile operator. The GGSN serves as the gateway for handsets to the Internet.

APN	Access point name
CG-NAT	Carrier-Grade NAT
GGSN	Gateway GPRS support node (3G)
GPRS	General packet radio service
GTP	GPRS tunneling protocol
MMS	Multimedia Messaging Service
MNC/MCC	Mobile network/country code
MNO	Mobile network operator
MVNO	Mobile virtual network operator
Node B / eNB	Base station (3G) / enhanced Node B (4G)
P-GW	PDN gateway (4G)
PDN	Packet data network
RNC	Radio Network Controller
S-GW	Serving gateway (4G)
SGSN	Serving GPRS support node (3G)
SIM	Subscriber Identifier Module
SUPL	Secure User Plane Location protocol
UE	User equipment, or handset

Table 1: Summary of the most common 3GPP terms and acronyms used throughout the paper.

2.1.2 The Data Plane:

The data plane, represented by the yellow pipe in Figure 1, consists of a direct IP tunnel created by the GPRS Tunneling Protocol (GTP) [13] between the handset and a gateway specified in the APN settings on the device. To connect to the Internet, a mobile device must possess an APN configuration provided by its operator. It determines various parameters of the network connection, including the services the gateway offers (e.g., Internet access and multimedia messaging), or SUPL for assisting location sensors [38]), and IP addressing (IP version, static or dynamic IP address use). However, most mobile platforms like Android allow users to edit the APN settings in case they are not pre-installed in the OS.

Beyond the GGSN resides the IP core, oftentimes behind a Carrier-Grade NAT (CG-NAT). Many mobile operators deploy their own IP core, including middleboxes such as DNS resolvers, HTTP and DNS proxies, NATs, and firewalls. These elements have functional reasons to exist: they accommodate address space depletion (NATs), improve security (firewalls), enhance performance (TCP splitters), or reduce latency (proxies). However, these middleboxes generally remain hidden from the user and operators unilaterally impose their use via in-path deployment or settings locked down in the handset.

2.2 Business Relationships—the 8^{th} Layer

Business relationships form a vital part of the mobile ecosystem. In fact, two classes of mobile operators exist: those who own spectrum (MNOs), and those who do not (MVNOs). This section describes in detail the two business models and network sharing agreements between operators (i.e., roaming), as well as the role played by regulatory bodies in the market emergence of MVNOs.

2.2.1 Mobile Network Operators and Roaming

Mobile Network Operators, or MNOs, are the "traditional" mobile operators. MNOs provide mobile voice and data services after acquiring a radio spectrum license from a government body. They also deploy their own network and support infrastructure.

Due to financial or spectrum availability constraints an MNO might not offer service in parts of the region it operates in. In such cases, the MNO typically enters into a business relationship with another MNO that does have service in that region. This allows the first MNO to provide coverage to its subscribers in the region through "roaming". Roaming is the ability of a cellular customer to automatically use any mobile service when traveling outside the geographical coverage area of the home network, by using a partnering visited network. Roaming agreements, particularly domestic ones, provide a cost-efficient way for MNOs to increase their coverage area without deploying actual infrastructure on the ground; they are thus a cost-saving technique for sparsely populated areas within a "liberalized" (open) mobile market [20]. In the US, such agreements have frequently emerged after the FCC eased restrictions on carriers to obtain automatic roaming agreements in areas where they owned spectrum but do not have infrastructure.

Network sharing agreements increase the complexity of network analysis due to the fact that two ways exist to realize the interconnection between the visited network and the home network. The first one, called *home-routed*, tunnels a user's traffic to their home network by inter-connecting this network with local SGSNs. This option increases the length of the path required by the handset to access the Internet and thus the network latency, but it provides a more homogeneous service as the customers always connect to their home network (i.e., the customer is effectively using its home data-plane, including performance enhancing proxies and DNS resolvers). The second option, *local breakout*, allows a roaming device to connect to the Internet as if it were a local device. As a result, the roaming device uses the data plane of the host network. This solution shortens the data path to the Internet, but it makes the user vulnerable to possible inefficiencies, bugs, and poor practices of the host network.

2.2.2 Mobile Virtual Network Operators (MVNOs)

An MVNO does not have a licensed mobile spectrum; it therefore enters into a business agreement with an MNO to lease spectrum to provide service. This time- and cost-effective approach, sometimes encouraged by regulatory agencies, allows new operators to enter the market and increase competition for the benefit of the consumer. MNOs also promote the creation of MVNOs in order to monetize network capacity that may otherwise remain unused.

MVNOs span a wide range of deployment types and business practices such as branding, marketing, and billing. In this paper we classify MVNOs by the type of service they provide as well as their network infrastructure. The most basic model is known as "light MVNO" or "resellers". In this case, the MVNO acts as a rebranded version of the host MNO, thus fully using its infrastructure. MNOs sometimes promote this model, creating and operating their own light MVNOs as low-cost versions of their own brand (e.g., by not providing customer support to their subscribers, as in the case of GiffGaff in the UK). The other model is known as "full MVNO". Here, the host MNO only provides radio network access; the MVNO deploys their own IP core and customized services. Full MVNOs therefore have more operational freedom from their parent MNO than light MVNOs; however this comes at the cost of making a larger initial investment. As a result, full MVNOs can better tailor their services to target specific communities or demographics, or monetize their user base with alternative business models. Consequently, a large variety of MVNO realizations has emerged.

3. RELATED WORK

Previous cellular network studies focused mainly on performance, rather than feature and behavior characterization. In 2004, Rodriguez *et al.* identified DNS operations as one of the root causes for poor performance on early UMTS networks [32]. To overcome such limitations, they proposed performance-enhancing proxies using techniques such as TCP tuning, aggregating TCP connections, content compression, and DNS optimizations [14,31]. Other studies about web performance focused on client-side optimizations [41, 42], and cache behavior optimization [16, 18]. The work by Rula *et al.*, and the study by Xu *et al.* used active DNS probing from mobile handsets to evaluate the performance of DNS resolvers [33, 48].

More recently, Zarifis *et al.* measured the length of the path between the GGSN and the end of the IP core [49], highlighting the reduced number of ingress points for 3G networks present even in a country as large as the USA. Wang *et al.* identified and characterize carrier-grade NATs present on cellular networks [43]. Their study focused on the security vulnerabilities of such NATs and their impact on the battery life of handsets due to short connection timeouts. Leong *et al.* analyzed the other side of the coin: non-NATed users. In their work, they observed that a malicious user can perform attacks such as data quota drain, DoS flooding, and battery drain to users with routable IP addresses [25]. None of these studies analyzed the deployment of middleboxes such as HTTP and DNS proxies in commercial networks and their impact on users' network connectivity, security, and privacy.

The research community has studied MVNOs mainly from an economic and regulatory perspective. Specifically, there has been considerable interest in understanding how they penetrate a market traditionally dominated by MNOs [15, 23, 35], stressing aspects such as branding and user behavior [34]. From a technical angle the only work comparing MNO–MVNO performance (i.e., TCP throughput and HTTP download time) is the recent study performed by Zarinni *et al.* for two undisclosed MNOs and three associated MVNOs for each in the USA [50]. Our analysis extends and complements this study by analyzing proxy behavior and privacy leaks in terms of operators, cellular technologies, and APN configurations, as well as characterizing business relationships and operational modes. Finally, Mulliner characterized privacy leaks on WAP proxies based on HTTP headers collected by a web server [29]. Our analysis confirms that many of these issues still remain 5 years later.

4. NETALYZR FOR ANDROID

Netalyzr is a free, user-driven network troubleshooting platform we have developed and maintained since 2009. Originally built as a browser-based client, Netalyzr analyzes a broad spectrum of network properties as observed from the edge of the network; it interacts with a suite of custom-built test and measurement servers, looking for a wide range of behavioral anomalies such as DNS transport limitations, hidden proxies, HTTP proxy behavior, network path anomalies, DNS manipulations and performance, outbound port filtering, bufferbloat, and UPnP-enabled gateways. We refer the reader to our earlier work [24, 44–46] for a full description of the tests and for architectural and operational details of the Netalyzr platform.

Due to the growth and continuous evolution of cellular networks, we have adapted and extended Netalyzr as an Android app. Numer-

ous iOS API restrictions make it difficult to port our full test suite to iOS. The mobile app implements the set of tests run by the browser and additionally leverages Android's APIs to extend the test suite and obtain the device's contextual information such as signal-to-noise ratio, the mobile carrier, TLS root certificates, and APN settings. We launched Netalyzr as a free app on Google Play [7] and the Amazon App Store [6] in November 2013. The app has since been installed by 28,000 users in some 120 countries.

4.1 Test description

We next describe the tests relevant for characterizing business relationships and their effect on user experience. We have two broad classes of tests: to characterize the radio and IP network that the user is on, and to characterize the effect of middleboxes, particularly HTTP and DNS proxies, in the network.

4.1.1 Network Identification and IP Core characterization

Due to the complexities inherent in cellular access to the Internet it becomes necessary to identify and decouple the mobile service provider at the radio level, SIM card level, and IP level. For example, a roaming user might be using the radio network of the visited network and the IP core of the home network. As stated earlier, full MVNOs have their own IP core while light MVNOs do not, but both always use an MNO's radio network. We collect three kinds of data with these tests.

IP Addressing. Netalyzr identifies the client's local IP address via Android's APIs and system properties, and uses TCP connections and UDP flows to our echo servers to identify the public IP address of the device. We use the `whois` tool to identify the organization owning the IP address.

Cellular Network Provider Identification. To identify the network service operator we use Android's `TelephonyManager` and `ConnectivityManager` APIs, and extract the APN settings as reported by the handset. This allows us to identify the name of the mobile operator, the name of the operator as reported by the SIM card, the APN providing the service, the cell ID (if permitted by the user), the 3GPP standard providing the service, as well as the MNC and MCC parameters (integer numbers allocated by the ITU that, in combination, uniquely identify the mobile operator at the radio level and the country it operates in—e.g., MCC=310, MNC=410 identify AT&T in the USA [2]).

Location. Android allows us to extract city-level device location if the user allows it. This information helps identify roaming between mobile operators and lets us identify locations with poor network performance.

4.1.2 Proxy Detection

Netalyzr studies HTTP and DNS behavior, including proxy implementation technologies, implementation artifacts, and limitations. The app employs Java's APIs as well as our own customized HTTP and DNS engines in order to analyze these protocols.

In principle we can detect the presence of a proxy any time it permutes a connection's properties. Our basic detection approach is to employ a client and server under our control that exchange precisely known messages; we then look for deviations from the expected. For the present study the most relevant tests for proxy identification and characterization include tests for HTTP and DNS.

HTTP proxies.

Non-responsive server test. TCP-terminating proxies may be deployed in cellular networks for performance improvement [14, 32]. Such proxies are likely to respond with a `SYN-ACK` to a client's connection request before connecting to the intended origin server. We test for this behavior by attempting a connection to a server that replies with a `RST`. If the Netalyzr client's attempt to connect to this server on port 80 initially succeeds, this indicates the presence of a TCP-terminating proxy.

Header modification test. RFC 2616 [17] specifies that systems should treat HTTP header names as case-insensitive, and, with few exceptions, free of ordering requirements. Furthermore, RFC 2615 indicates that any proxy *must* add the `Via` header to indicate its presence to intermediate protocols and recipients [17]. Netalyzr fetches custom content from our server using mixed-cased request and response headers in a known order. Any changes indicate the presence a proxy. This method also allows identifying additional headers added by the HTTP proxy, as in the case of tracking headers [27], and whether intermediate proxies modify traffic using techniques such as image transcoding, which can affect the fidelity of content delivered to mobile clients through CDNs and other cloud infrastructure.

HTTP enforcement test. In addition to standard HTTP, Netalyzr attempts to fetch an entity using the protocol declaration `ICSI/1.1` instead of `HTTP/1.1`. If this request is rejected, we know that the network has a protocol-parsing proxy.

Invalid `Host` header value test. CERT VU 435052 [19] describes how some in-path proxies would interpret the `Host` request header and attempt to contact the listed host rather than forward the request to the intended address. We check for this vulnerability by fetching from our server with an alternate `Host` header of `www.google.com`. The presence of this vulnerability in commercial proxies is alarming as it suggests that operators may not have their middlebox software upgraded, thus potentially having other vulnerabilities not covered by our test suite.

DNS proxies.

Netalyzr checks for DNS awareness and proxying by using custom DNS messages. Our DNS server answers requests for `entropy.netalyzr.edu` with a CNAME encoding the response parameters, including the public address of the device, UDP port, DNS transaction ID, and presence of `0x20` encoding. The client sends such DNS requests directly to the back-end server, bypassing the configured resolver. If the client observes any change in the response (e.g., a different transaction ID or public address), then we have found in-path DNS proxying. After that, Netalyzr makes another direct request, this time with deliberately invalid formatting, to which our servers generate a similarly broken reply. If the request is blocked, we have detected a DNS-aware middleboxes that prohibit non-DNS traffic on UDP port 53.

4.2 Usability Considerations

Besides serving as a crowd-sourced data collector for our study, Netalyzr offers a comprehensive troubleshooting service for mobile device users. We aim to appeal to a broad audience with varying technical sophistication. As a result, we put significant effort (still with potential for improvement) into presenting technical results to the user in an accessible fashion.

Figure 2 shows three screenshots of the application, including a real results summary for an Optus (Australia) subscriber. Users can share their results, seeking help through social media or via email. Indeed, we have received more than 700 emails from Netalyzr users, and have seen users reporting problems to their mobile providers through Twitter.

The tests execute largely sequentially in order to minimize the risk of test-induced connectivity problems. As a result, the execu-

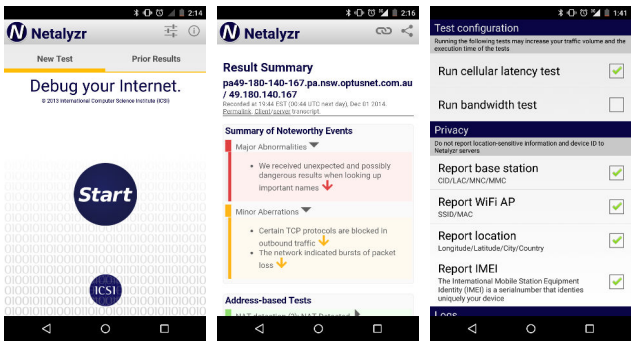

Figure 2: Netalyzr for Android screenshots. Start button activity (left), part of a result summary for a cellular session (center), and app settings and privacy panel (right).

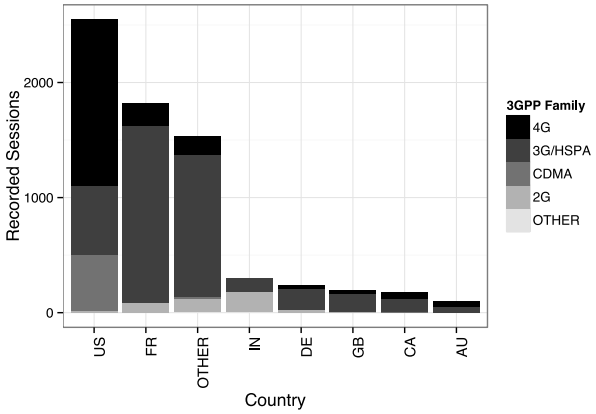

Figure 3: Number of sessions per country per 3GPP technology.

tion time depends on link latency and speed, ranging from 3 minutes in WiFi to 10 minutes in GPRS.

4.3 Privacy considerations

We stress that this research strives to understand the interplay of provider infrastructure and mobile devices, and does not focus on human subjects (confirmed by our institute's IRB). Accordingly, while we collect data on a wide range of network and device properties, little of the data collected has any bearing on the user's privacy. The users may nevertheless control specific aspects of the data collection (e.g., turn off location reporting) if they so desire.

5. DATASET AND METHODOLOGY

Between November 2013, when we released Netalyzr for Android, and September 2014 we recorded 39,110 sessions. In this paper we focus on studying properties of cellular networks, thus we exclude sessions generated over WiFi and WiMax links. We also remove sessions where the handset was tethering, leaving us with 6,918 sessions covering 371 operators in 128 countries. As a result of its crowd-source nature, our dataset has biases in terms of countries, demographics, and technologies from which the sessions originated. We received most of our sessions from the USA and France due to media coverage in those countries.

We potentially have a "geek bias", with 34% sessions coming from rooted phones. Furthermore, CDMA, HSPA variants, and LTE dominate in our datasets. We also observe a long tail in the distribution of sessions run per operator.

We focus our analysis on 6 countries with more than 100 sessions recorded, as shown in Figure 3. We exclude India from our analysis even though we have numerous sessions from it due to the complexity of India's spectrum allocation at a state level—still dominated by 2G standards—and regulatory legislation [22]. Even with the reduced set of countries, we faced significant challenges both in annotating the data and in sanitizing it. We now turn to our approaches for doing so.

5.1 Mobile Operator Identification

Identifying the mobile operator providing the service to a customer is a daunting task. One may think that the operator name reported by the native APIs and the MCC/MNC tuple suffice to associate a session to a mobile operator. However, with the rise of MVNOs relying on such fields can prove misleading: many MVNOs do not necessarily have an MNC value and in some cases they buy access from multiple MNOs to increase their coverage. While some MVNOs, particularly full MVNOs which have their own IP core, have an MNC, the values reported by Android's `TelephonyManager` only identifies the current registered operator (i.e., the MNO providing the radio link).

We developed a technique to identify and classify both flavors of MVNOs by correlating public IP addresses with existing MNOs, `whois` information, DNS resolvers, APN settings, proxy properties, operator name, and the MNC/MCC as reported by Android APIs, along with manual trawling of MVNO Directory databases [4]. For example, this technique allows us to characterize sessions reporting the mobile carrier `HOME` as TracFone Wireless sessions, an MVNO in the USA that allows their subscribers to use either T-Mobile or AT&T's network. By correlating sessions with these two operator names, we observed that they both have two different APNs: `tfdata` and `wap.tracfone`, and their public IP address is owned by Syniverse Technologies. Moreover, the MCC and MNC values report two different MNOs providing the radio link: AT&T (310/410), and T-Mobile (310/260).

This method furthermore allows us to correct errors in the data reported by Android APIs, and also to characterize MVNOs as either light or full MVNOs (§ 2.2.2) based on IP core characteristics. We identified 43 different MVNOs in the countries of interest; this is a small fraction of the total number of existing MVNOs. According to the MVNO Dynamics website [5] 138 MVNOs operated in the first quarter of 2014 in the USA alone, some with marginal market penetration and a short lifespan. Unfortunately, due to Netalyzr's crowd-sourcing approach, we were not able to penetrate and collect traces from all of them.

5.2 Data Sanitization

As a consequence of the crowd-sourced nature of the Netalyzr dataset and the difficulty of controlling the circumstances under which users launch Netalyzr, our dataset presents inaccuracies and corner cases that can potentially bias the results. Consequently, we take into account a variety of factors when vetting our dataset and exclude the following sessions:

1. Sessions collected in testing or engineering mode by network engineers: the MCC code for such sessions is 1 as defined by ITU [2], and the operator name in many cases is "DEFACE".

2. Pre-3G standards such as GPRS, EDGE, and 1xRTT.

3. Sessions in which we could not identify the global IP address of the handset. In a small number of sessions, Netalyzr could

Total number of cellular sessions	**6,918**
Sessions from selected countries	**5,083**
Testing networks	7
Pre-3G standards	207
Erroneous local IP address	21
Static IP addresses	77
User-customized DNS	84
VPNs	84
AdBlock app	51
Femtocells	36
Erroneous operator name	32
International roaming	11
Remaining valid sessions	**4,512**

Table 2: Dataset pruning due to data impurities and unusual cases. Note that sessions may have more than one problem.

not identify the global IP address of the device, possibly as a result of app crashes. Because such sessions considerably limit the fidelity of our analysis, we remove them.

4. Sessions that report static IP addresses. Such sessions are quite rare; mobile networks generally deploy carrier-grade NATs [43].

5. Sessions indicating network customization by users. Android supports traffic tunneling through VPNs (including users with proxy-based TOR clients such as Orbot [8]), and savvy users with rooted devices can configure Google or OpenDNS as their default DNS resolvers. Furthermore, applications such as AdBlock behave as local proxies to block mobile advertising. Such customization modifies the network path and hides properties of the mobile network we are trying to measure.

6. Sessions executed through femtocells, which we identify based on the operator name. Femtocells route traffic directly to the Internet using wired or satellite links rather than the cellular infrastructure.

7. Sessions that exhibit inconsistencies in the mobile operator name as reported by Android's `TelephonyManager` API. This could arise due to APN misconfigurations, erroneous sessions due to app crashes, or re-branding by MVNOs. For instance, some clients still report Orange and T-Mobile in the UK, which merged to form a new MNO known as Everything-Everywhere (EE). Despite our efforts to correct such anomalies by correlating different sources of information (as explained in the following sections), we could not identify the operator providing the service for 32 sessions.

8. Sessions from international roaming users.

Table 2 lists the number of sessions cleaned using the process listed above. After applying our filtering, 4,512 sessions from 70 operators in 6 countries remain that we consider valid for further analysis (49% of the total cellular sessions). Although this provides only a relatively low number of sessions per carrier, it still suffices for identifying structural problems in cellular networks, MNO and MVNOs relationships, domestic roaming agreements, misconfigurations, traffic manipulations, and privacy leaks. Our analyses characterize discrete properties of the provider infrastructure that are unlikely to vary significantly over time from the same provider (unlike metrics such as latency or throughput).

5.3 Dataset Release

We have released the relevant data used for this paper through CRAWDAD [39] after performing the dataset sanitization process explained in § 5.2. We have also made available a larger Netalyzr dataset through PREDICT [21]. We do not include sessions generated by smartphones through VPN tunnels, and remove sensitive information which could identify our users. This includes geographical location, HTTP header modifications—as in the case of HTTP headers injected by HTTP proxies—and some APN fields such as username and passwords. We have anonymized additional sensitive values, such as public IP addresses.

6. CHARACTERIZING MNOs

It is well known that operators commonly deploy proxies in cellular networks for performance enhancement [14, 31]. However, the architecture of MNO networks—particularly the deployment, configuration, and behavior of middleboxes—and network sharing agreements between MNOs can similarly affect performance, data fidelity, as well as user security and privacy. Such design and operational decisions generally remain unknown and unapparent to users, developers and researchers alike. In this section, we analyze these aspects of MNOs. To determine whether a given operator constitutes an MNO or an MVNO, we use documentation from the ITU and national regulatory bodies such as the FCC in the USA, and OFCOM in the UK. MNOs dominate the market [30] and also dominate our dataset, with 86% of our sessions from MNO subscribers.

6.1 Proxy Deployment and Behavior

We focus our analysis on two predominant kinds of middleboxes deployed in cellular networks: HTTP and DNS proxies. While our data also indicate occasional presence of proxies for SMTP, POP3, or IMAP we do not consider these further in this paper.

6.1.1 HTTP Proxies

HTTP proxies intercept and relay HTTP traffic. They are typically used for enhancing performance, and are widely deployed on cellular networks for adapting content, saving bandwidth using caches, compression and transcoding, and for traffic filtering [14, 32].

We confirm substantial presence of HTTP proxies in cellular networks: Netalyzr detected proxies in 59% of sessions. In comparison, our previous analysis revealed that only 14% of clients in wired networks were proxied [44]. Figure 4 shows the percentage of sessions for a given MNO enforcing various flavors of proxies. The color map indicates the percentage of penetration using a gradient from 0% (white) to 100% (red). Where we do not possess sessions conducted using a given cellular technology, we report it in gray, as in the case of Wind (Canada), O2 (GB), and E-Plus (Germany) on 4G LTE.

Our analysis shows that most MNOs generally deploy in-path proxies performing HTTP header re-ordering, HTTP header injection, and header modification. § 6.2.2 discusses HTTP header injection in upstream traffic and their privacy implications in detail. The US operator C Spire modifies non-HTTP traffic going through their proxies. Other MNOs, such as Vodafone and Optus in Australia, Vodafone and T-Mobile in Germany, SFR in France, and T-Mobile in the USA transcode images to reduce data volume. Most proxies remain transparent to the user. RFC 2616 stipulates that proxies *must* indicate their presence using the `Via` general-header field [17]. Despite the fact that proxies are widely present, the `Via` field proves rare: Netalyzr detected its presence in only one ISP (SFR, France) and then only in sessions going through

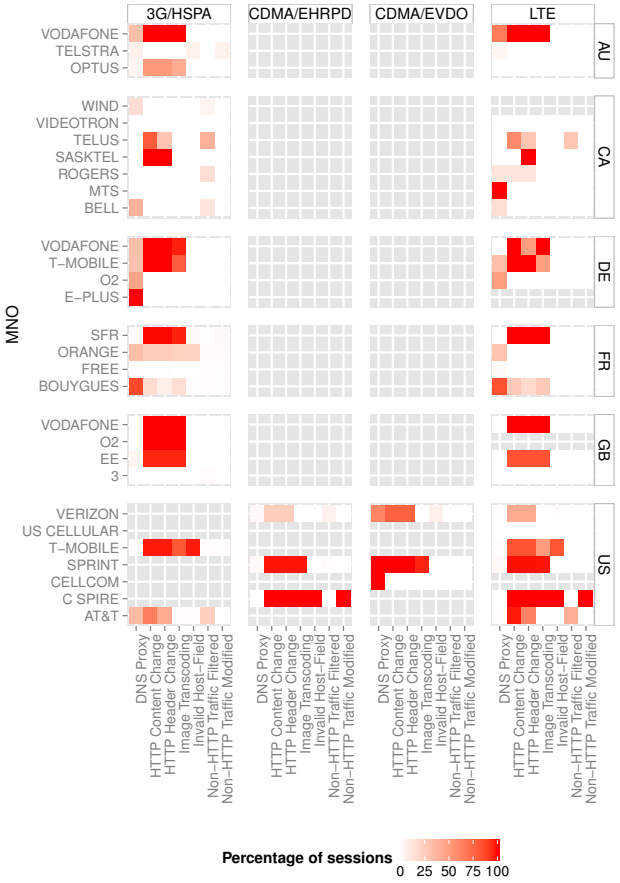

Figure 4: Map showing the deployment and behavior of DNS and HTTP proxies for the MNOs in the considered countries. Each box represents a given property. The color reflects the percentage of sessions with it on a gradient from 0% (white) to 100% (red). Gray indicates we did not identify any session from a MNO for a given technology.

a legacy WAP gateway, `proxy.cwg.net`. Only 2.3% of SFR sessions pass through these gateways. Unfortunately, since these proxies are generally transparent to the user, this behavior affects the fidelity of the data delivered to the customer without their (or the content providers') knowledge or consent. For example, users will not be able to download high-resolution images, even if they explicitly desire to do so.

6.1.2 *DNS Proxies*

Netalyzr sends custom DNS queries directly to our DNS resolver to identify whether the mobile network is proxying DNS traffic, and whether they modify these queries. Netalyzr also sends non-DNS messages over DNS to test if a proxy enforces (its idea of) the DNS protocol, blocking non-DNS traffic on port 53. From our data, we see that a significant number of mobile operators proxy DNS queries. In total, 18% of DNS queries were proxied, of which 70% changed the sender IP address or DNS request ID in queries sent directly to our name servers. The remainder enforced DNS semantics on UDP port 53, blocking non-DNS traffic on this port.

Figure 4 also shows the penetration of DNS proxies among MNOs. We see DNS proxies more commonly in resource-constrained standards such as 3G and CDMA/EVDO, with exceptions such as Bouygues in France, MTS in Canada, and Vodafone in Australia. We speculate that operators rely on such proxies to keep control over the network, particularly relatively resource-constrained ones. For example, AT&T proxies DNS only for a small subset of subscribers that go through their legacy APNs, as explained in § 6.3. Verizon, Cellcom, and Sprint deploy DNS proxies only on their CDMA networks.

Operators also have the ability to interfere with DNS name lookups. Our records indicate that Free mangled and blocked Google Ads (in particular the domains `partner.googleadservices.com` and `ad.doubleclick.net`) due to a peering dispute between Free and Google [12]. Queries for both of these domains resolved to an IP address owned by Free (212.27.40.246, or `white.proxad.net`), serving a blank web page.

6.2 Security and Privacy

As we show in the previous section, a great deal of diversity exists among operators regarding how they manage their networks: some ISPs proxy DNS and HTTP, some do not, while others actively modify user traffic. We next examine the effects of these actions on user security and privacy.

6.2.1 *Security implications*

Proxies in cellular networks are generally enforced without any choice on behalf of the user. This means that any security vulnerability or performance degradation affecting them afflicts all users of that network. Our tests show that middleboxes in T-Mobile and C Spire in the USA as well as Orange in France still remain vulnerable to CERT VU435052 [19], a vulnerability dating back to 2009. While the impact might prove minor in the case of an ISP proxy (the attack primarily enables same-origin bypass within corporate network proxies), the presence of a five-year-old vulnerability naturally raises questions about what other, potentially more serious, vulnerabilities affect these proxies.

6.2.2 *Privacy implications*

Proxies that modify user traffic may not only affect service quality but also violate the users' privacy. We have observed that many operators inject additional headers into every HTTP request generated by their users. The type of headers varies, but the more insidious cases uniquely identify the subscriber, their location, or their IP address. Table 3 summarizes injected HTTP headers containing sensitive information about the users and the operators that add them, as evident in our dataset. Some of these headers, particularly `X-ACR` (by AT&T) and `X-UIDH` (by Verizon), uniquely identify the mobile subscriber.

Tracking-enablers (or "perma-cookies" [27]) are associated with advertisement programs deployed by mobile operators such as Verizon Selects [40, 47]. They allow partner companies to identify the user so they can deliver targeted advertisements more effectively. However, such headers also allow malicious tracking of visiting web users. These privacy leaks violate the trust relationship between mobile operators and their users, and have recently attracted significant press attention [26, 27]. The "perma-cookie" practice prompted us to call out these headers' presence in the Netalyzr session results, and one article recommends using Netalyzr to identify such privacy leaks. While AT&T stopped adding these headers soon after the media exposure, we note that Verizon still continues the practice at the time of this writing. Looking be-

yond the countries we focus on in this study, we identified a similar "perma-cookie" header in sessions from Vodafone Netherlands (`X-VF-ACR`). Other operators such as T-Mobile in Germany, EE and O2 in the UK, and SFR in France also leak private information about their users in other forms.

Proxies widely use headers such as `X-Forwarded-For`, `X-EE-Client-IP`, and `X-Gateway` to identify the client's provider-internal IP address as well as its location. Finally, we have also observed proxies explicitly defining the maximum uplink and downlink speed for a given device, a unique ID of the gateway, as well as the gateway vendor. While these headers might not necessarily compromise user security, they can augment and complement classic HTTP cookies to facilitate user tracking and enable price/search discrimination [28].

6.3 Service Variability Within Providers

A given ISP's middlebox behavior can change depending on the wireless technology used. For example, in the USA Sprint proxies DNS on CDMA/EVDO, but not on LTE. This does not surprise much because the 2G, 3G, and 4G networks often operate independently. However, one would expect that the behavior of such networks to remain uniform within the same ISP and technology. In Figure 4 this should translate to white or red boxes. However, we see that in many cases, the boxes have varying shades or red, indicating that users within the same ISP and using the same technology could experience differing properties for their networks. We did not, however, observe any difference between IPv4 and IPv6 for Verizon, T-Mobile, or Telstra.

Again turning to Figure 4, we observe variability in caching and content modification even within the same ISP and technology. We identify and explain two such cases. For a small fraction of sessions originating from AT&T, the public IP address is owned by WDSPCo, a consortium of some of the major wireless providers in the USA [10]. We also noticed that these sessions correspond to the APN `wap.cingular`, which supports legacy WAP services. These sessions were markedly different in their proxy properties: while their HTTP traffic was not proxied (and the `X-ACR` header not added), DNS lookups were proxied. This clearly suggests this network is different from the network that serves other AT&T customers. Another interesting case study is Orange in France. We see that the properties of the network change depending on the public IP address of the session. For example, sessions that had public IP addresses in the 90.84.144/24 and 90.84.146/24 subnets had an HTTP proxy that does header reordering (92% of sessions), while the other Orange subnets do not (1.0% of sessions). We could not identify any specific APN defining such behavior; we speculate that in Orange's case, service diversity could arise due to a heterogeneous proxy deployment in certain geographical areas.

The ISP controls the network that users traverse; however, APN settings for ISPs are freely available online (via the ISP's website or forums such as xda-developers [11]) and users can easily misconfigure their phones by experimenting with such settings, negatively affecting their own service quality. For example, we have received numerous emails from users reporting app crashes when connecting to the Internet, as a result of using APNs that enforce compression by default while using apps that lack compression support.

6.4 Business Relationships

As described in § 2.2, roaming agreements between MNOs frequently serve the purpose of providing higher network coverage to their customers while incurring minimal deployment costs. In this section we demonstrate that this practice is common in certain countries and happens transparently to the user.

6.4.1 *Method*

We identify roaming sessions by comparing the IP core provider, the provider of the radio link, and APN information when available. This allows us to identify home-routed roaming implementations (cf. § 2.2.1) easily, but not local breakout, as in this case the devices are fully connected to the host network at both the radio and IP level. To differentiate devices doing local breakout from other host network subscribers, we require access to the issuer of the SIM card. Unfortunately, Netalyzr only began collecting this information in September 2014. Furthermore, VPN tunnels to IP addresses of ISPs providing both wired and wireless access can be easily confused with roaming sessions. APN and network interface configuration information sometimes provide an alternative way to identify such cases.

6.4.2 *Home Routing*

Our analysis identified 91 roaming sessions between MNOs in our database, with 92% seen in France between Free and Orange (only for their 3G network), and between other MNOs as reported in Table 4. Android's `TelephonyManager` API flags none of these roaming cases. Free's agreement with Orange is particularly interesting as they account for 81% of all national roaming sessions identified, and for 10% of the total mobile sessions recorded from Free subscribers.

Free subscribers can access Orange's network in locations where Free has no network infrastructure deployed (Free was the last MNO entering the French market, and is still deploying its infrastructure). We can see roaming happening all over France, even in large metropolitan areas such as Paris and Bordeaux. When roaming, the operator name reported by Android is Orange, whereas the APN settings and the IP core belong to Free (their public IP address is in the 37.160.0.0/12 subnet). Moreover, middlebox behavior matches that of customers under Free coverage, as a result of home routing. Consequently, if we based our characterization solely on the operator name, we would have incorrectly mapped Free's Google ad services blocking (described in § 6.1.2) to Orange France.

While users roaming on local breakout implementations are potentially susceptible to inefficiencies and vulnerabilities present in the home network, they can experience significantly higher network latency (e.g., due to increased path-lengths, see § 2.2.1) and performance inefficiencies (e.g., due to the masking of user locations which can impair CDN replica selection) in home routing scenarios. Netalyzr can measure the effect of the different roaming implementations, but we do not currently possess enough data to reach statistically significant conclusions.

We have also identified a reduced number of sessions of MNOs roaming in the US, as listed in Table 4. Cricket has roaming agreements with other MNOs for their CDMA deployment. 4 sessions report "Extended Network" or "Preferred System" as the operator name, while the IP core is Verizon's. We verified that these sessions are actually generated by Verizon subscribers roaming in other MNOs from USA or Canada, in which home airtime rates still apply [9]. We have seen such sessions for CDMA, and in this case, all sessions present Verizon's HTTP "perma-cookies" (even when roaming abroad as in the case of some Verizon subscribers roaming in Canada on Roger), and DNS traffic is proxied and modified.

6.5 Discussion

Middleboxes are not just pervasive in cellular networks, but also transparent to the user. We identify HTTP and DNS proxies in 59% and 18% of our sessions, respectively. Aside from proxying, some middleboxes also actively modify user traffic by modifying HTTP

HTTP Header	Operator	First time seen	Last time seen	Notes
X-ACR	AT&T (US)	2014-05-17	2014-08-26	Unique perma-cookie.
X-EE-Client-IP	EE (GB)	2014-06-11	2014-06-11	Private IP address of the subscriber.
X-Forwarded-For	BOUYGUES (FR)	2014-02-01	2014-02-01	Private IP address of the subscriber.
X-Forwarded-For	O2 (GB)	2013-12-27	2014-07-17	Private IP address of the subscriber.
X-Forwarded-For	SASKTEL (CA)	2014-03-02	2014-08-16	Private IP address of the subscriber.
X-Forwarded-For	SFR (FR)	2013-11-07	2013-12-08	Private IP address of the subscriber.
X-Forwarded-For	T-MOBILE (DE)	2013-11-07	2014-08-28	Private IP address of the subscriber.
X-Gateway	O2 (GB)	2013-12-27	2014-07-17	Network gateway and its location.
X-UIDH	VERIZON (US)	2013-10-23	2014-08-25	Unique perma-cookie.
X-VFPROVIDER	SFR (FR)	2013-11-07	2013-12-08	Operator name.

Table 3: HTTP header identifiers added by different operators.

Country	MNO	Host MNO
France	FREE	ORANGE
	FREE	BOUYGUES
	ORANGE	SFR
USA	AT&T	T-MOBILE
	CRICKET	US CELLULAR

Table 4: List of observed network sharing agreements between MNOs.

headers, HTTP content, transcoding images, or leaking sensitive information about the subscribers. Furthermore, business relationships between MNOs can result in users experiencing middlebox manipulations originating from another operator's network.

Our results highlight the general opacity that shrouds cellular networks. Users have little say in (or, indeed, knowledge about) how their ISP treats or manipulates their traffic. This results in users experiencing differing service quality depending on how in-path middleboxes are configured (often caused by APN misconfigurations), and also leaves users vulnerable to potential security problems or privacy leaks. Transparent roaming agreements between MNOs mean that users do not know which network carries their traffic at any given time. Depending on the actual business relationship, users could either face potential performance hits if their operator uses home routing, or vulnerabilities or tracking of the host network if it does local breakout.

Variability in middlebox performance, often within the same ISP, also complicates research into cellular network performance. It does not suffice simply to obtain the provider name from the operating system; not only can the radio provider and coverage differ, but the path the traffic takes can affect measured service quality. The consequences affect developers, content providers (e.g., ad networks), and researchers. Middleboxes intercept, modify and block traffic, affecting data fidelity. Moreover, they also invalidate server-side performance measurement as a result of the connection-decoupling done by connection-terminating proxies, requiring the control of both endpoints to effectively characterize mobile network performance.

7. CHARACTERIZING MVNOs

In this section we delve into how MVNOs deploy their networks and the potential resulting impact on customers. In § 2.2.2 we developed a basic taxonomy of MVNOs, classifying them as either full or light MVNOs. We first develop these notions further, then classify the MVNOs we see in our dataset according to this taxonomy, and characterize their properties.

7.1 Classifying MVNOs

Figure 5 shows the MVNOs in the six countries we analyze using the method described in § 5.1. For each MVNO we show the MNO providing radio access as well as the provider of the IP core infrastructure (according to `whois`). The line style represents the type of MVNO: we show full MVNOs with solid gray lines and light MVNOs with dotted ones. We validated our findings with information from official MVNO websites (when available) and industry forums [5]. As in § 6 we also describe the network behavior for each of the identified MVNOs, per Figure 6.

7.1.1 Light MVNOs

Light MVNOs are globally most common. We have identified light MVNOs in each country under study, and they are the only type of MVNO we find in Australia, Canada, Germany, and Great Britain. Our results indicate that most light MVNOs provide Internet access through only one parent MNO, per Figure 5, though we also find NRJ in France using both Orange's and SFR's networks.

MVNO sessions generally come from users connected via 3G/HSPA standards. We have not recorded any MVNO session in France over 4G, even when parent MNOs have 4G infrastructure deployed. This suggests that MNOs can block access to their most advanced 3GPP technology. Only 22% of MVNO sessions happened over LTE, as opposed to 40% of LTE sessions generated by MNO subscribers. Furthermore, 88% of the MVNO sessions we recorded over LTE originated from the USA.

We observe that light MVNOs exhibit behavior identical to that of their parent MNOs. In fact, MVNOs such as Fido, Metro PCS, and GiffGaff are owned by Rogers, T-Mobile, and O2 respectively. However, as in the case of MNOs, APN settings can affect behavior. For example, Virgin Australia users can choose between the APNs `VirginInternet` (that proxies DNS traffic and only provides 3G coverage), and `YESINTERNET` (which grants access to the 4G network). The price for these two APN settings varies [1]. Legacy APN settings also propagate through MVNO subscribers. We have identified sessions with the `Via` header `proxy.cwg.net` coming from Leclerc and Prixtel subscribers. Consequently, MVNO users are susceptible to the vulnerabilities of the parent MNO: MetroPCS proxies also present the CERT VU435052 vulnerability [19] found in T-Mobile networks. It remains an open question to what degree MVNO network performance differs from that offered by their host MNOs. A preliminary study for some US operators suggests that MVNO quality of experience proves sub-par compared to the MNO one, for popular applications [50].

7.1.2 Full MVNOs

We have identified four full MVNOs in our dataset: Virgin and Numericable in France, and TracFone and Cricket in the USA.

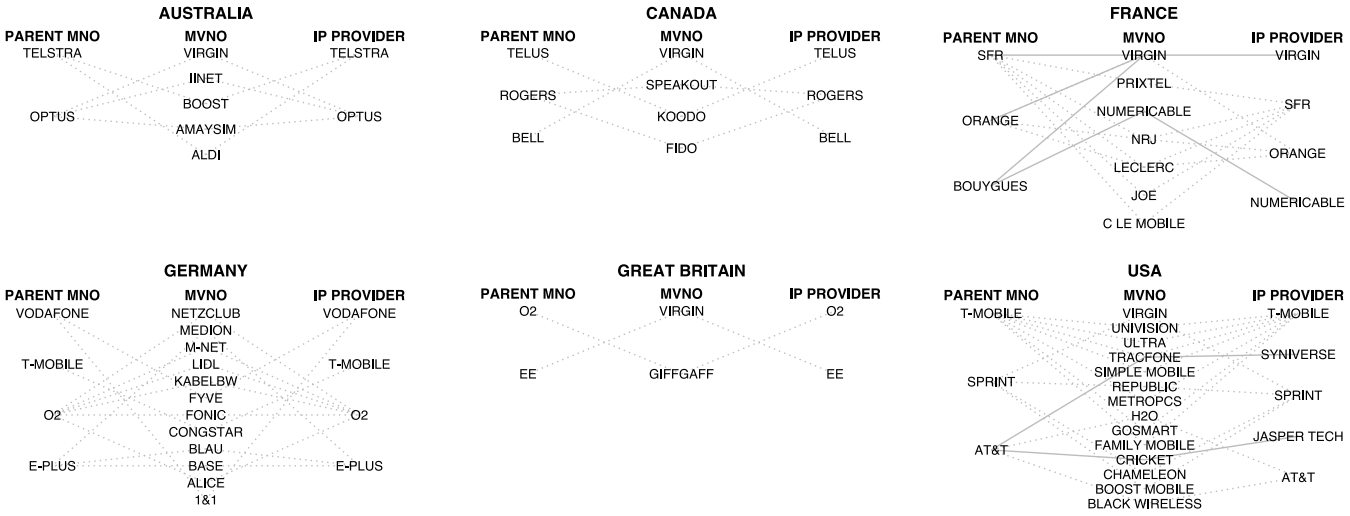

Figure 5: MVNO relationship with parent MNOs and cloud providers for each one of the MVNOs. Solid lines represent full MVNOs, dotted lines light ones.

We represent them in Figure 5 using solid gray lines. Since full MVNOs need to invest larger sums of money to deploy their own IP infrastructure, the fact that we find fewer full MVNOs than light ones does not surprise. Cable and DSL providers such as Numericable can enter the mobile business as MVNOs, forwarding mobile subscribers' data to their IP core.

Full MVNOs differ from light MVNOs in two ways. First, since they have their own IP core, the behavior of such networks can deviate from that observed in their parent MNOs. For instance, when TracFone customers connect through T-Mobile they still use TracFone proxies and thus are not vulnerable to CERT VU435052, unlike other MVNOs who run on T-Mobile. TracFone also performs image transcoding, as opposed to other full MVNOs like Numericable in France. Second, full MVNOs generally use more than one host MNO and IP provider, as shown in Figure 5. Doing so allows full MVNOs to provide coverage to a wider population without being constrained to the network deployment of a single mobile operator, with more competitive prices (due to buying in bulk) than MNOs. TracFone subscribers have the freedom to decide which network they want to use (by entering the APN settings accordingly) while seeing the the same IP core behavior. Nonetheless, most sessions come from a single MNO, AT&T (56 % of sessions), with the rest using Sprint and T-Mobile.

Finally, some MNOs are MVNOs for certain radio frequencies. Both MetroPCS and Cricket have their own CDMA legacy infrastructure, but become MVNOs for their 3G/4G network. While MetroPCS behaves as a light MVNO on T-Mobile's 3G/4G network, Cricket is a full MVNO on AT&T. Interestingly, Cricket partners with Jasper Wireless for their 4G IP infrastructure rather than using AT&T's or their own. Jasper Wireless is a cloud provider partner with a large number of MNOs and MVNOs all over the world [3]. Consequently, the behavior of their CDMA middleboxes differs from the ones they use for their LTE/HSPA networks.

7.2 User Security and Privacy

Light MVNOs' subscribers are susceptible to problems introduced by their host MNO. This holds for GiffGaff in the UK as well as Prixtel and Leclerc subscribers in France, all vulnerable to PII leaks caused by their host MNO proxies provided by O2 and SFR. Table 5 provides a summary. The `X-VFPROVIDER` header added by SFR proxies identifies the name of the mobile operator, even if it is an affiliated MVNO as Prixtel and Leclerc. Full MVNO subscribers can suffer from other network problems, but we have not observed any PII leaks. We cannot verify whether they monetize subscribers' data in other ways.

MVNO	HTTP Header
CONGSTAR (DE)	`X-Forwarded-For`
GIFFGAFF (GB)	`X-Forwarded-For`
GIFFGAFF (GB)	`X-Gateway`
LECLERC (FR)	`X-Forwarded-For`
LECLERC (FR)	`X-VFPROVIDER`
PRIXTEL (FR)	`X-Forwarded-For`
PRIXTEL (FR)	`X-VFPROVIDER`

Table 5: Permanent identifiers found on MVNOs which are inherited from the MNO providing their access.

7.3 Discussion

MVNOs represent an emerging business model that allows providers to offer service without buying spectrum. The MVNO model has several advantages: it lowers the bar for entry for providers, increases competition, and offers the opportunity for MNOs to sell unused capacity. Light MVNOs, which use the host MNO for both radio and IP infrastructure, are more common than full MVNOs, which have their own IP core.

The general problem of opacity extends to MVNOs as well. Light MVNOs are typically little more than rebranded versions of the parent MNO, which renders them susceptible to the problems associated with the parent MNO, including privacy leaks. Full MVNOs do not face this problem as they control their user traffic. However, they are either limited to the coverage area of the host MNO, or, if they have relationships with multiple MNOs, susceptible to differing service quality depending on which MNO currently provides the radio connectivity.

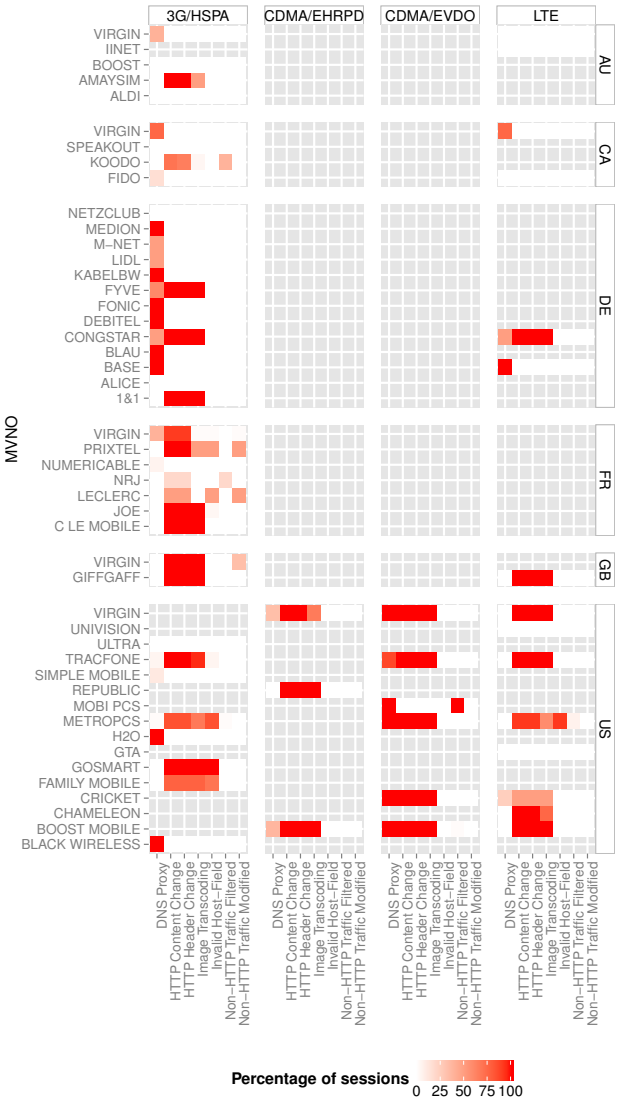

Figure 6: Deployment and behavior of DNS and HTTP proxies for the MVNOs in the considered countries. Only TracFone and Cricket in the USA, and Virgin and Numericable in France, are full MVNOs. The remaining light MVNOs exhibit the same behavior as their parent MNO. (See Figure 5 for details about the MVNO–MNO relationships.)

8. SUMMARY

In this paper, we have provided perspectives on the state of the art of mobile network characterization and middlebox behavior. Looking beyond the predominant radio-centric angle of analysis, we have used Netalyzr's rich test suite to shed light on key properties of HTTP and DNS proxies in provider networks, business relationships, and their implications for users, developers, and mobile researchers alike.

Using data collected by Netalyzr for Android, we evaluated mobile operators from the USA, Canada, Australia, France, Germany, and the UK. We summarize our results in Table 6. We first highlight the difficulty in identifying the mobile operator; simply using the operator name or the MCC/MNC value as reported by Android often does not suffice. We supplement the above with a host of features, including the IP core and APN settings, to identify the operator. Doing so helps us accurately attribute network features and behavior to operators (e.g., attributing to a full MVNO rather than the host MNO). Netalyzr's rich proxy detection suite identified the presence of transparent HTTP proxies in 59% of cellular sessions. We show how proxies can modify or block DNS and HTTP traffic; in some cases affecting data fidelity with techniques such as image transcoding. We report proxies with five-year-old vulnerabilities, indicating a lack of upgrading or patching that suggests the likely presence of other non-identified vulnerabilities. We also find proxies that facilitate tracking of customers by adding new header fields to HTTP requests, as in the case of "perma-cookies". We have identified fundamental differences in proxy behavior even for subscribers from a given operator as a result of supporting different middlebox and APN configurations. Our results underscore the importance of considering proxy behavior to avoid biases while evaluating mobile performance.

In addition to proxy behavior, we identified roaming agreements between operators and different MVNO–MNO relationships by combining radio information, handset settings, and IP-layer information. Each one of the mobile markets in this study has its own peculiarities, but in most of the cases, MVNOs are simple rebranded versions of the actual mobile provider, so the service quality delivered to the customer potentially remains constant. Unfortunately, we could not evaluate whether link performance is degraded

It proves difficult to identify the mobile operator using information provided by Android's APIs; we combine several features, including IP core behavior, to achieve this (§ 5.1). It is vital to do so in order to correctly attribute network behavior.
Middleboxes commonly occur in cellular networks: 59% of sessions have an HTTP proxy, while 18% of sessions have a DNS proxy. These proxies usually remain unapparent to the user, and, even if they reveal themselves, are difficult to avoid. (§ 6.1)
Middleboxes may have security vulnerabilities, modify user traffic (§ 6.2.1), and leak private information (sometimes intentionally) about the user (§ 6.2.2). Savvy users take advantage of trusted and commercial VPN clients and servers to avoid middlebox manipulation and tracking.
Middleboxes may possess heterogeneous configurations within a single ISP, particularly with differing APN settings. This can lead to middleboxes treating similar traffic differently. (§ 6.3)
Roaming agreements are common, particularly in France. However, in many cases such agreements remain transparent to the user, meaning that the device may not flag the fact that roaming on another operator's network occurs. (§ 6.4)
MVNOs reside in every country we study; they mostly constitute rebranded versions of the parent MNO. (§ 7.1.1)
Light MVNOs' subscribers remain vulnerable to the same inefficiencies, security holes, and privacy leaks as subscribers of the parent MNO, as a result of sharing the same IP core. (§ 7.1.1 and § 7.2)
Full MVNOs, which we observe only in France and the USA, deploy their own IP core and thus have great control over their user traffic. Consequently, the treatment of their subscribers' traffic may differ from that of the parent MNO's subscribers. (§ 7.1.2)

Table 6: Highlights of our work.

(e.g., link capacity, or latencies for DNS lookups or HTTP fetches) due to the high variability of the radio links and their best-effort nature. We refer the reader to the work by Zarinni *et al.* for a first performance evaluation in the USA [50].

The Need for Operational Transparency.

As we highlight throughout the paper, most mobile users, developers, and researchers lack visibility into network behavior, due to both middlebox configuration and business relationships, and its impact on performance, service quality, data fidelity, security, and privacy. Practices such as image transcoding can affect data fidelity, and HTTP header injection can compromise privacy without the user's knowledge. The vast majority of HTTP proxies deployed in today's commercial cellular networks do not comply with the HTTP standards, which indicate that any proxy *must* advertise its presence both to the client and the server with the `VIA` general-header [17]. Unfortunately, users and online services have limited power to avoid middleboxes due to their direct imposition on the user's path communication. Only a small fraction of savvy users take advantage of "trusted" VPN clients and servers to avoid middlebox interference in their traffic.

Mobile users also do not generally understand the role of APN settings and how these can affect their service quality. Some users find APN settings via web searches, which can lead to baffling configuration problems due to mismatches with the user's mobile environment. Moreover, mobile users also lack a clear picture of what MVNOs are and how they operate on top of an MNO either as a light or as a full MVNO. In fact, it generally proves difficult to identify the MNO providing the service or determine the operational mode employed by the MVNO. Similarly, most mobile operators notify users with visual notifications or text messages about the possibility of experiencing differing service quality when roaming on a different provider as a result of network sharing agreements. However, as we have seen in this work, a large number of roaming cases happen inapparently to the user.

Conclusion.

Our work has highlighted the challenge of attempting to characterize the behavior of mobile networks based only on fine-grained network measurements. While the presence of proxies in cellular networks may be expected by users, developers and researchers with a good technical background, these middleboxes' purpose, effect on quality of service, and impact on data fidelity are generally not well-defined and difficult to determine empirically. Without an understanding of the broader ecosystem from which the measurements come—including the technical implications of different business relationships—any mobile network analysis risks misattributing network performance, as well as security and privacy issues, to parties other than those actually responsible.

The key implication of our work is that the traditional notion of "the mobile provider" as perceived by users, developers, researchers, and regulators needs rethinking. Providing mobile network connectivity is not a task fulfilled by a monolithic, immutable entity, but rather the result of a complex and dynamic interplay of a set of service providers.

Acknowledgments

As always, we are deeply grateful to our Netalyzr users for enabling this study. This work was partially supported by funding provided to ICSI through National Science Foundation grants CNS-1111672 ("Measuring and Modeling the Dynamics of IPv4 Address Exhaustion"), CNS-1213157 ("User-Centric Network Measurement"), and CNS-1237265 ("Beyond Technical Security: Developing an Empirical Basis for Socio-Economic Perspectives") and by the DHS Directorate of Science and Technology under grant N66001-12-C-0128. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors or originators and do not necessarily reflect the views of the NSF or of the DHS. The authors would like to thank the anonymous reviewers and our shepherd Ramon Caceres (Google) for constructive feedback on preparation of the final version of this paper. We also wish to thank Amazon, Comcast, and Google for their generous support, as well as Moritz Steiner (Akamai), Andrius Auçinas (University of Cambridge), and Jon Crowcroft (University of Cambridge) for their valuable feedback.

9. REFERENCES

[1] APN Settings Virgin Mobile AU. https://community.virginmobile.com.au/t5/Phones-knowledge-base/What-are-the-generic-APN-settings-for-Internet-MMS-and-tethering/ta-p/76.

[2] ITU Operational Bulletin. http://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-OB.1056-2014-OAS-PDF-E.pdf.

[3] Jasper Wireless. http://www.jasper.com/operators.

[4] MVNO Directory. http://www.mvnodirectory.com/overview.html.

[5] MVNODynamics. http://www.mvnodynamics.com.

[6] Netalyzr for Android. Amazon App Store. http://www.amazon.com/International-Computer-Science-Institute-Netalyzr/dp/B00IM1TXMA.

[7] Netalyzr for Android. Google Play. https://play.google.com/store/apps/details?id=edu.berkeley.icsi.netalyzr.android.

[8] Orbot: Proxy with Tor. https://play.google.com/store/apps/details?id=org.torproject.android.

[9] Verizon coverage locator. http://www.verizonwireless.com/b2c/coveragelocator/mapInformation.jsp.

[10] WDSPCo. http://www.wdspco.org.

[11] Xda-Developers. http://www.xda-developers.com/.

[12] France v Google. http://www.economist.com/news/business/21569414-xavier-niel-playing-rough-internet-giant-france-v-google, 2013.

[13] 3GPP. General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gn and Gp interface. http://www.3gpp.org/DynaReport/29060.htm.

[14] R. Chakravorty, S. Banerjee, P. Rodriguez, J. Chesterfield, and I. Pratt. Performance Optimizations for Wireless Wide-area Networks: Comparative Study and Experimental Evaluation. In *Proc. ACM MobiCom*, 2004.

[15] L. Cricelli, M. Grimaldi, and N. L. Ghiron. The Competition Among Mobile Network Operators in the Telecommunication Supply Chain. *International Journal of Production Economics*, 2011.

[16] L. Fan, P. Cao, W. Lin, and Q. Jacobson. Web Prefetching Between Low-bandwidth Clients and Proxies: Potential and Performance. In *Proc. ACM SIGMETRICS*, 1999.

[17] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext Transfer Protocol – HTTP/1.1. RFC 2616, IETF, June 1999.

[18] A. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger. Is There a Case for Mobile Phone Content Pre-staging? In *Proc. ACM CoNEXT*, 2013.

[19] R. Giobbi. CERT Vulnerability Note VU 435052: Intercepting proxy servers may incorrectly rely on HTTP headers to make connections, February 2009.

[20] International Chamber of Commerce. Telecoms Liberalization. 2004.

[21] International Computer Science Institute. Netalyzr dataset. PREDICT. https://www.predict.org/Default.aspx?tabid=104.

[22] R.S. Jain. Spectrum auctions in India: lessons from experience. *Telecommunications Policy*, 25, 2001.

[23] A. Kiiski. Impacts of MVNOs on mobile data service market. In *17th European regional ITS conference*, 2006.

[24] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: Illuminating The Edge Network . In *Proc. ACM IMC* , 2010.

[25] W. K. Leong, A. Kulkarni, Y. Xu, and B. Leong. Unveiling the Hidden Dangers of Public IP Addresses in 4G/LTE Cellular Data Networks. In *Proc. ACM HotMobile*, 2014.

[26] R. McMillan. Verizon and AT&T are the only wireless carriers using Perma-cookies. http://http://www.wired.com/2014/11/permacookie-free/.

[27] R. McMillan. Verizon's perma-cookie is a privacy-killing machine. http://www.wired.com/2014/10/verizons-perma-cookie/.

[28] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting price and search discrimination on the internet. In *Proc. ACM HotNets*, 2012.

[29] C. Mulliner. Privacy leaks in mobile phone internet access. In *Proc. IEEE ICIN*, 2010.

[30] Piran Partners. MVNO Observatory. http://www2.piranpartners.com/component/joomdoc/doc_download/24-piran-partners-mvno-observatory-spring-2014.html, 2014.

[31] P. Rodriguez and V. Friedman. Performance of PEPs in Cellular Wireless Networks, 2003.

[32] P. Rodriguez, S. Mukherjee, and S. Ramgarajan. Session Level Techniques for Improving Web Browsing Performance on Wireless Links. In *Proc. WWW*, 2004.

[33] J. P. Rula and F. Bustamante. Behind the Curtain: Cellular DNS and Content Replica Selection. In *Proc. ACM IMC*, 2014.

[34] D. Shin. Overlay Networks in the West and the East: a Techno-Economic Analysis of Mobile Virtual Network Operators. *Telecommunication Systems*, 2008.

[35] D. H. Shin and M. Bartolacci. A Study of MVNO Diffusion and Market Structure in the EU, US, Hong Kong, and Singapore. *Telematics and Informatics*, 2007.

[36] J. Sommers and P. Barford. Cell vs. WiFi: On the Performance of Metro Area Mobile Connections. In *Proc. ACM IMC*, 2012.

[37] N. Vallina-Rodriguez, A. Aucinas, M. Almeida, Y. Grunenberger, K. Papagiannaki, and J. Crowcroft. RILAnalyzer: A Comprehensive 3G Monitor on Your Phone. In *Proc. ACM IMC*, 2013.

[38] N. Vallina-Rodriguez, J. Crowcroft, A. Finamore, Y. Grunenberger, and K. Papagiannaki. When assistance becomes dependence: Characterizing the costs and inefficiencies of a-gps. *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, 2013.

[39] Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Nicholas Weaver, and Vern Paxson. CRAWDAD data set icsi/netalyzr-android (v. 2015-03-24). Downloaded from http://crawdad.org/icsi/netalyzr-android/, March 2015.

[40] Verizon. Verizon Selects Participation Agreement. http://www.verizonwireless.com/support/terms/products/verizon-selects.html?MID=810452&RID=178021460&CMP=EMC-C-S-RWDS-Announcement1-T1-JulyAug2014-RD&EMHID=E007D0ABF359C5E4BC3832F3C2B204B6&CUHID=E1583E4C8FA54F0CB9FA01EC656DBD5F.

[41] Z. Wang, F. X. Lin, L. Zhong, and M. Chishtie. Why Are Web Browsers Slow on Smartphones? In *Proc. ACM HotMobile*, 2011.

[42] Z. Wang, F. X. Lin, L. Zhong, and M. Chishtie. How Far Can Client-only Solutions Go for Mobile Browser Speed? In *Proc. WWW*, 2012.

[43] Z. Wang, Z. Qian, Q. Xu, Z. M. Mao, and M. Zhang. An Untold Story of Middleboxes in Cellular Networks. *SIGCOMM Computer Communication Review*, 2011.

[44] N. Weaver, C. Kreibich, M. Dam, and V. Paxson. Here Be Web Proxies. In *Proc. PAM*, 2014.

[45] N. Weaver, C. Kreibich, B. Nechaev, and V. Paxson. Implications of Netalyzr's DNS Measurements . In *Proc. SATIN*, 2011.

[46] N. Weaver, C. Kreibich, and V. Paxson. Redirecting DNS for Ads and Profit. In *USENIX FOCI Workshop*, 2011.

[47] WebPolicy.org. How Verizon's Avertising Header Works. http://webpolicy.org/2014/10/24/how-verizons-advertising-header-works/.

[48] Q. Xu, J. Huang, Z. Wang, F. Qian, A. Gerber, and Z. Mao. Cellular Data Network Infrastructure Characterization and Implication on Mobile Content Placement. In *Proc. ACM SIGMETRICS*, 2011.

[49] K. Zarifis, T. Flach, S. Nori, D. Choffnes, R. Govindan, E. Katz-Bassett, Z. Mao, and M. Welsh. Diagnosing Path Inflation of Mobile Client Traffic. In *Proc. PAM*, 2014.

[50] F. Zarinni, A. Chakraborty, V. Sekar, S. Das, and P. Gill. A first look at performance in mobile virtual network operators. In *Proc. ACM IMC*, 2014.