Meddle: middleboxes for increased transparency and control of mobile traffic
Ashwin Rao, Justine Sherry, Arnaud Legout, Arvind Krishnamurthy, Walid Dabbous, David Choffnes

To cite this version:

HAL Id: hal-01281734
https://hal.archives-ouvertes.fr/hal-01281734
Submitted on 2 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Meddle: Middleboxes for Increased Transparency and Control of Mobile Traffic

Ashwin Rao
INRIA

Justine Sherry
UC Berkeley

Arnaud Legaut
INRIA

Arvind Krishnamurthy
University of Washington

Walid Dabbous
INRIA

David Choffnes
University of Washington

Categories and Subject Descriptors
C.2 [Computer Communication Networks]: General

Keywords
Mobile, Networking, Measurement, Middlebox, Performance.

1. INTRODUCTION

Mobile networks are the most popular, fastest growing and least understood systems in today’s Internet ecosystem. Despite a large collection of privacy, policy and performance issues in mobile networks [3, 6] users and researchers are faced with few options to characterize and address them. In this poster we present Meddle, a framework aimed at enhancing transparency in mobile networks and providing a platform that enables users (and researchers) control mobile traffic.

In the mobile environment, users are forced to interact with a single operating system tied to their device, generally run closed-source apps that routinely violate user privacy [6], and subscribe to network providers that can (and do) transparently modify, block or otherwise interfere with network traffic [13].

Researchers face a similar set of challenges for characterizing and experimenting with mobile systems. To characterize mobile traffic and design new protocols and services that are better tailored to the mobile environment, we would like a framework that allows us to intercept and potentially modify traffic generated by mobile devices as they move with users, regardless of the device, OS, wireless technology, or carrier. However, implementing this functionality is difficult on mobile devices because it requires warranty-voiding techniques such as jail breaking to access and manipulate traffic at the network layer [3]. Even when using such an approach, carriers may manipulate traffic once it leaves the mobile device [13], thus rendering some research impractical. Furthermore, researchers generally have no ability to deploy solutions and services such as prefetching and security filters, that should be implemented in the network.

In this poster, we present Meddle, a framework that combines virtual private networks (VPNs) with middleboxes to provide an experimental platform that aligns the interests of users and researchers.

Meddle relies on VPN tunnels to access the mobile traffic regardless of the device, OS, wireless technology, and carrier. Meddle can thus provide a continuous and comprehensive view of how mobile devices interact with the Internet. Once packets arrive at a Meddle server, we use a variety of middlebox approaches to interpose on mobile-device traffic. Meddle offers new opportunities for measuring and characterizing mobile traffic, and designing new in-network features to improve the mobile experience. For example, by accessing network traffic regardless of the wireless technology we can analyze how different operating systems and apps offload their traffic from cellular networks to Wi-Fi. To improve the user experience, we implement packet filters to block ads; unlike existing packet filters for mobile devices, the packet filters provided by Meddle do not require jail-breaking the mobile device. Furthermore, Meddle provides a vantage point for separating mobile-network performance from server-side performance, thus improving bottleneck identification for mobile applications. Meddle also enables researchers to investigate what-if scenarios for the impact of new middleboxes as if they were deployed in carrier networks. For example, Meddle can be used to deploy anonymization systems such as Privad [5].

2. MEDDLE ARCHITECTURE

Meddle uses VPNs as a portable mechanism to tunnel the data traffic from mobile devices to a machine where users and researchers
Table 1: Traffic share and volume from three devices when using Wi-fi at home, Wi-Fi at work, or a cellular connection. Meddle provides a comprehensive view of mobile traffic regardless of the access technology.

<table>
<thead>
<tr>
<th>Device</th>
<th>Home (Wi-Fi)</th>
<th>Work (Wi-Fi)</th>
<th>Mobile (Cell)</th>
<th>Volume (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Android</td>
<td>38.6%</td>
<td>58.7%</td>
<td>2.7%</td>
<td>436 MB</td>
</tr>
<tr>
<td>iPhone</td>
<td>47.3%</td>
<td>23.1%</td>
<td>0.7%</td>
<td>105 MB</td>
</tr>
<tr>
<td>iPad</td>
<td>92.8%</td>
<td>7.2%</td>
<td>N/A</td>
<td>812 MB</td>
</tr>
</tbody>
</table>

Table: Traffic share and volume from three devices when using Wi-fi at home, home at work, or a cellular connection. Meddle provides a comprehensive view of mobile traffic regardless of the access technology.

can exert control over network flows. VPNs also reduce the barrier to entry for deploying Meddle because Android, BlackBerry, and iOS, which represent more than 86% of the mobile device market [4], have native VPN support. As shown in Figure 1, when a mobile device connects to the Internet, we tunnel its traffic via a nearby Meddle server in a similar way to how CDNs use DNS to redirect Web clients to nearby content caches [2]. On each Meddle server we implement custom services for users such as packet filtering, caching, and intrusion detection. Meddle thus takes two well-known technologies – VPNs and middleboxes – and combines them in unintended ways for the mobile environment.

3. CASE STUDIES

We believe that the research enabled by Meddle will form a positive feedback loop in which new, proven research artifacts become additional incentives for user adoption, thus enabling further research. We have a Meddle prototype which has been running since August 27, 2012. Our prototype currently serves mobile devices that include an Android phone, an iPhone, an iPad, and an iPod Touch. The participants in our study are using their own devices to interact with Meddle, as part of an IRB-approved study to characterize mobile traffic.

**Meddle example.** We have implemented a DNS-based filter to block ads, analytics, and mediation sites. We believe that such a filter is an important incentive for a user based study because Vallina-Rodriguez et al. [12] observe that ads account for 5% of daily traffic from more than 50% of Android users in a large European ISP. Our ad blocking engine relies on the publicly available list of domains for ads and analytics [1]; we augment this list of domains using the recent research on mobile ads [6, 7]. From our initial deployment, we observed a 0.05% to 0.8% reduction in total traffic at each mobile device due to our ad blocking engine. We are investigating whether this difference from previous work is due to incomplete filters, user-specific behavior or other factors.

**Measuring Traffic Oﬄood.** In Table 1, we summarize the traffic volume observed from an Android phone, an iPhone, and an iPod Touch. Participants in our study are using their own devices to interact with Meddle, as part of an IRB-approved study to characterize mobile traffic.

**Measuring Traffic Oﬄood.** In Table 1, we summarize the traffic volume observed from an Android phone, an iPhone, and an iPod, for 15 days since October 16, 2012. The three users generated 1.35 GB of traffic that passed through our Meddle server. In Table 1, we observe that the iPad usage consumes more cellular bandwidth compared to the Android user. Further, we note that the vast majority of mobile device traffic is traversing Wi-Fi networks instead of cellular networks. We plan to use this comprehensive view of network traffic to investigate the marginal impact of oﬄooding cellular traffic onto Wi-Fi networks.

**Overhead.** We observe low overheads in terms of power consumption, data quota, and network latency, when tunneling data traffic via a VPN. We observed a 10% increase in power consumption when streaming an HD video from YouTube to an Android device and an iPhone via one of our Meddle servers. We measured the overhead of the tunnel in terms of data overhead from IPsec headers and keep-alive messages, finding that it ranges from 8–12% for an Android phone and an iPhone. Our test traffic included Web searches, interaction on social networks, map searches, playing a game, online shopping, downloading popular apps, emailing and reading the news. To mitigate potential additional network delay from routing traffic through Meddle we envision a DONAR-style deployment where users are dynamically redirected to a Meddle server. We observed that PlanetLab nodes have a latency between 3 ms and 13 ms, with a median of 5 ms from the mobile-network egress points. We used the data collected from 10 mobile phones located throughout the US for this measurement study. Thus, when compared to RTTs of 10s or 100s of milliseconds that exist in mobile networks, the additional latencies from traversing a Meddle server is expected to be relatively small or even negligible.

4. FUTURE WORK

We are currently recruiting users for an IRB approved study on the network usage profiles of mobile phone users. To enhance the transparency of mobile networks, we plan to allow users to observe how their installed apps use the network and with whom these apps share (or leak) information; a system similar to Mozilla Collusion [8]. We also plan to test algorithms for content coalescing, caching, prefetching, and oﬄooding the work of processing the DOM to speed up page load times [9, 10, 11]. We are currently working towards making the Meddle system publicly available along with a framework to enable researchers to contribute to the Meddle system and analyze the data collected by the Meddle servers.

5. REFERENCES