Microsoft

Microsoft Research Ltd.
21 Station Road
Cambridge
CB1 2FB
United Kingdom

Tel: +44 (0)1223 479700
Fax: +44 (0)1223 479999

Thomas Karagiannis
Researcher
thomkar@microsoft.com
http://research.microsoft.com/~thomkar/

November 28th, 2013

Report for Ashwin Rao's Thesis - Improving Transparency and End-User Control in Mobile Networks

The thesis targets the general area of user privacy in the modern mobile ecosystem. As mobile devices in various types and form factors become ubiquitous, private user information appears to slip from users' control in exchange for the promise of a richer user experience, and added device and application functionality. Ideally, this exchange should be transparent and user choices should be conscious through well-defined user interfaces and well-understood metrics. Instead, as the thesis argues, a web of inter-dependences between developers, Internet Service Providers (ISPs), advertisers and other players results in lack of transparency and control, with private user information leaking from one player to the next with unclear (if any) benefit to the end user. Thus, well-designed and easy-to-use systems that provide users with visibility into how their information is used and by whom, but also allow users to control who gets to access this information are going to be invaluable in the future. Ashwin's thesis work is a step towards this direction.

This first chapter of the thesis starts by introducing this web of interdependence between modern applications (apps), ISPs, mobile OS providers and device manufacturers. The thesis does a good job of presenting these players and their relationships as well as the compromises that users are compelled to make to capitalize on the added app functionality. The chapter then provides a comprehensive discussion of previous work in the area and describes why existing solutions are limited because they are constrained to either specific apps or to individual mobile OSes and access technologies. Finally, the chapter presents the thesis' contributions – Meddle, a platform that aims to improve the transparency and end-user control of privacy, a characterization of mobile applications through real user-studies by deploying Meddle in the wild, and an analysis of YouTube traffic patterns. I believe that this chapter clearly sets the scene for the technical part of the work, specifying the problem and the challenges. What perhaps is missing from the discussion without affecting the overall message of this chapter, is an analysis about the role of the cloud, with the cloud providers being yet another player in this web of interdependences. Most applications are cloud-based today which implies a wealth of user data being stored at a few dominant cloud provider offerings and infrastructures – essentially, cloud providers might or might not, indirectly gain visibility into user data and app trends as do rest of the aforementioned players.

The second chapter describes the architecture of Meddle, which is based on redirecting mobile traffic through VPN tunnelling. Meddle can thus monitor all incoming and outgoing traffic from the device. This technique allows Meddle to be OS, device, technology and ISP agnostic. The thesis further argues that the design is also

Registered Office: Microsoft Limited, Microsoft Campus, Thames Valley Park, Reading RG6 1WG

Registered in England no: 03369488
VAT no: GB 642353552

Microsoft

Microsoft Research Ltd.
21 Station Road
Cambridge
CB1 2FB
United Kingdom

Tel: +44 (0)1223 479700
Fax: +44 (0)1223 479999

app agnostic. This is harder to achieve in my opinion as Meddle needs to understand some application semantics, for example when traffic is encoded. Still, I believe the design overcomes a series of limitations of previous works and provides an efficient way to monitor device traffic. Besides the design, I found the discussion on feasibility quite interesting, highlighting non-obvious aspects and trade-offs of Meddle design, especially regarding the location of the proxy and the encoded traffic. Further, the chapter does a good job to quantify the overheads which in most cases appear acceptable – it would also be interesting here to see some real deployment numbers. Finally, I enjoyed the discussion of the legal issues and appreciated the use of the Acceptable Use Policy. Essentially, the discussion implies that with Meddle, trust shifts from other players to the operator of Meddle servers. Meddle appears to take extensive steps to protect user privacy. Overall, the discussion in this chapter reveals a deep understanding of the examined area and the underlying trade-offs Meddle has to tackle.

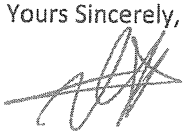
Chapter 3 presents an analysis of mobile app traffic with Meddle. The thesis shows that HTTP is the main source of app traffic, and provides a methodology to classify HTTP traffic flows to web services. This is a very challenging task and I believe the thesis provides a nice starting point to this end by considering information from the User-Agent and Host fields. The thesis shows high classification accuracy in controlled experiments and in real traffic which is impressive. Further, the thesis introduces a methodology to look at and classify SSL traffic which is important as the use of HTTPS increases. Finally, the chapter studies the leakage of PII information both in controlled experiments and in the wild and shows that Meddle can indeed identify a series of leaks. This chapter provides significant contributions both in terms of novel measurement methodologies, but also regarding the trends of mobile app traffic and evidence of leakage of private information in the wild.

Motivated by the fact that YouTube traffic was dominant in the captured traces, Chapter 4 focuses on analysing the traffic patterns of the popular video streaming service. The chapter provides an extensive, comprehensive study across browsers and devices, and shows how rate control mechanisms evolved over time. The chapter provides valuable insights into YouTube traffic, highlighting at the same time that traffic characteristics can be quite diverse depending on the browser or device. This is an important contribution as it implies that studies focusing only on a few browsers or devices might get a distorted picture.

Finally, I found that the appendix of the thesis also presents interesting insights, for example the notes about how Google search evolved and the results about the (lack of) benefits of compression in various cases.

In summary, I believe that the thesis deserves to be defended. The thesis is very well-written, organized and presented, and the work tackles a challenging problem. The thesis presents a series of important contributions both in terms of providing a system to help users of mobile devices control their privacy, but also regarding insights about the traffic of modern applications.

Yours Sincerely,



Thomas Karagiannis
Researcher
Microsoft Research

Registered Office: Microsoft Limited, Microsoft Campus
Thames Valley Park, Reading RG6 1WG

Registered in England no: 03369488
VAT no: GB 642353552