



---

AT&T Labs - Research

Dr. Balachander Krishnamurthy

October 20, 2009

Federal Trade Commission  
Office of the Secretary, Room H-135 (Annex P)  
600 Pennsylvania Avenue  
NW, Washington DC 20580

Dear Commission,

We (Balachander Krishnamurthy of AT&T Research Labs and Craig Wills of Worcester Polytechnic Institute) submit the following comments in regard to “Privacy Roundtables – Comment, Project No. P095416.” The substance of these comments are two papers concerning privacy in the World Wide Web. These papers were recently presented and published in leading Computer Science conferences.

The first paper addresses questions about the extent and practices of third-party behavioral advertising over a five-year period. The second paper documents how personally identifiable information is leaked to behavioral advertisers via popular Online Social Networking sites.

The first paper and its abstract:

Balachander Krishnamurthy and Craig E. Wills.

Privacy diffusion on the Web: A longitudinal perspective.

In *Proceedings of the World Wide Web Conference*, pages 541–550, Madrid, Spain, April 2009. ACM.

For the last few years we have studied the diffusion of private information about users as they visit various Web sites triggering data gathering aggregation by third parties. This paper reports on our longitudinal study consisting of multiple snapshots of our examination of such diffusion over four years. We examine the various technical ways by which third-party aggregators acquire data and the depth of user-related information acquired. We study techniques for protecting against this privacy diffusion as well as limitations of such techniques. We introduce the concept of secondary privacy damage.

Our results show increasing aggregation of user-related data by a steadily *decreasing* number of entities. A handful of companies are able to track

users' movement across almost all of the popular Web sites. Virtually all the protection techniques have significant limitations highlighting the seriousness of the problem and the need for alternate solutions.

In addition to the inclusion of the paper itself, we have continued the longitudinal study presented in the paper through September 2009. The updated results show that trends reported in the original longitudinal study continue in the most recently collected data. The updated graphs are in the addendum document.

The second paper and its abstract:

Balachander Krishnamurthy and Craig E. Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the Workshop on Online Social Networks in conjunction with ACM SIGCOMM Conference*, pages 7–12, Barcelona, Spain, August 2009. ACM.

For purposes of this paper, we define “Personally identifiable information” (PII) as information which can be used to distinguish or trace an individual’s identity either alone or when combined with other information that is linkable to a specific individual. The popularity of Online Social Networks (OSN) has accelerated the appearance of vast amounts of personal information on the Internet. Our research shows that it is possible for third-parties to link PII, which is leaked via OSNs, with user actions both within OSN sites and elsewhere on non-OSN sites. We refer to this ability to link PII and combine it with other information as “leakage”. We have identified multiple ways by which such leakage occurs and discuss measures to prevent it.

The data presented in this paper were originally gathered in February 2009. In September 2009, we repeated our methodology and were able to confirm that the leakage of PII via OSNs identified in the original study continues.

We appreciate the consideration of these comments by the Commission and would be happy to answer any questions raised by these comments.

Sincerely,

Balachander Krishnamurthy

# Privacy Diffusion on the Web: A Longitudinal Perspective (Updated Graphs)

Balachander Krishnamurthy  
AT&T Labs–Research

Craig E. Wills  
Worcester Polytechnic Institute

This document contains updated graphs to extend the timeframe of the original study “Privacy Diffusion on the Web: A Longitudinal Perspective” presented at the World Wide Web Conference in Madrid, Spain in April, 2009. The original paper is available at <http://www2009.org/proceedings/pdf/p561.pdf>.

The ten figures in this graph extend the corresponding figures in the original study to reflect data gathered periodically through September’09. Please see the original paper for a description of the methodology used in obtaining the data and analyzing the results.

The updated results show that trends present in the original longitudinal study continue in the updated data.

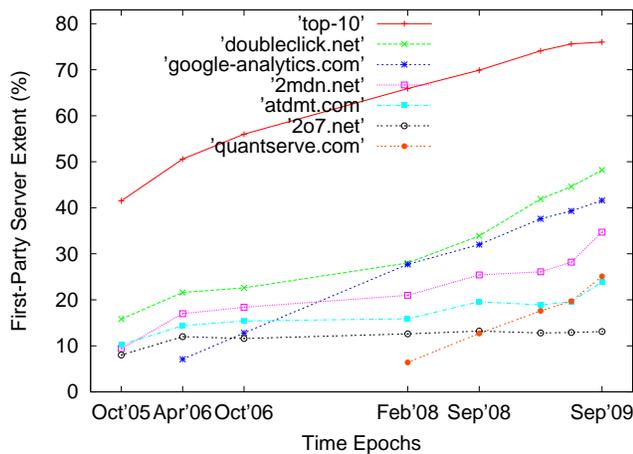


Figure 1: Extent of Top-10 Third-Party Root Domains in Each Epoch and Specific Contribution of Top Domains

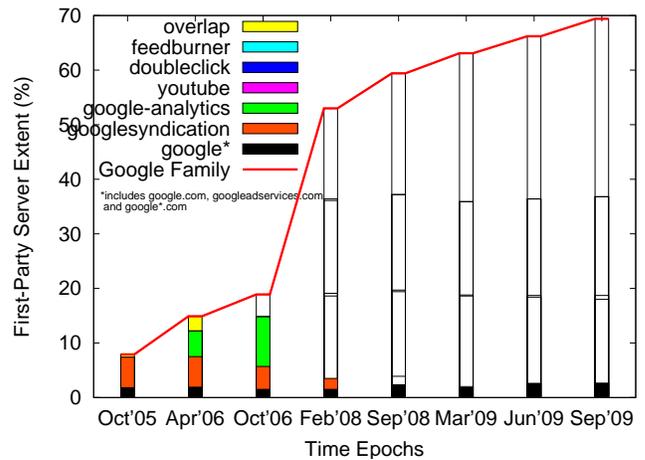


Figure 2: Growth of the Google Family

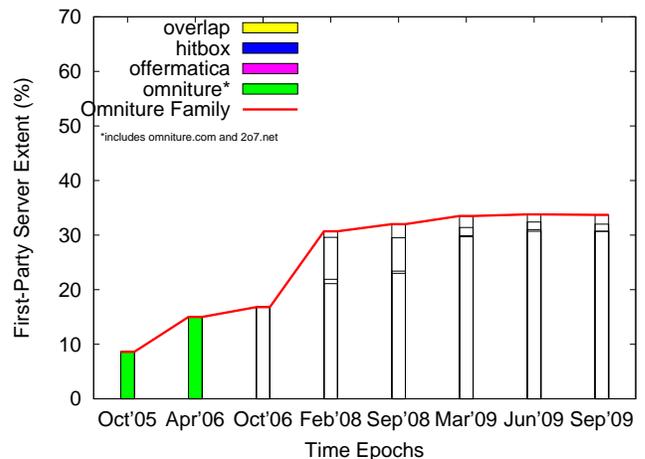


Figure 3: Growth of the Omniture Family

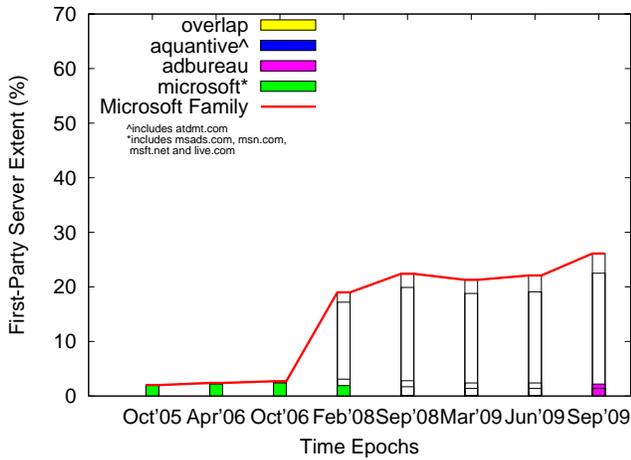


Figure 4: Growth of the Microsoft Family

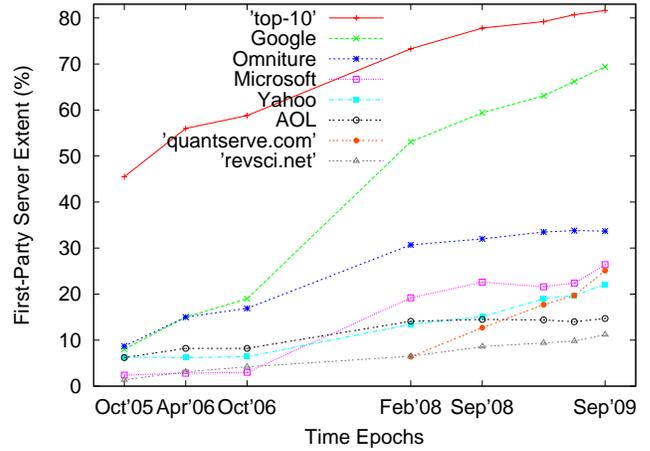


Figure 7: Extent of Top-10 Third-Party Families in Each Epoch and Specific Contribution of Top Families

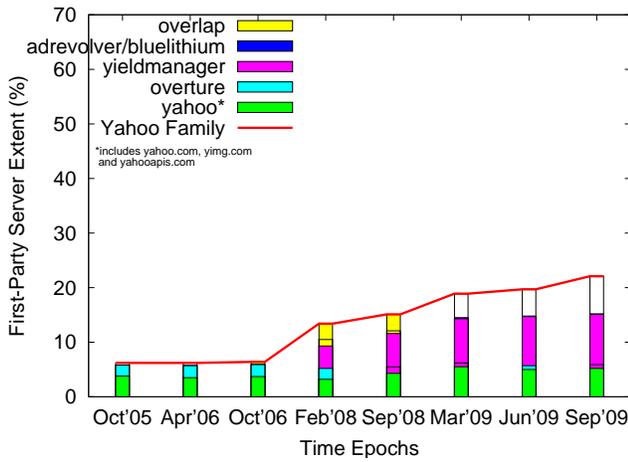


Figure 5: Growth of the Yahoo Family

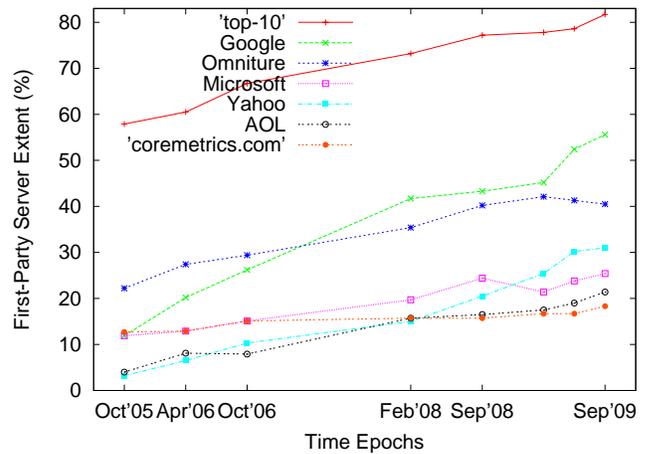


Figure 8: Extent of Top-10 Third-Party Families in Each Epoch and Specific Contribution of Top Families for Consumer Sites

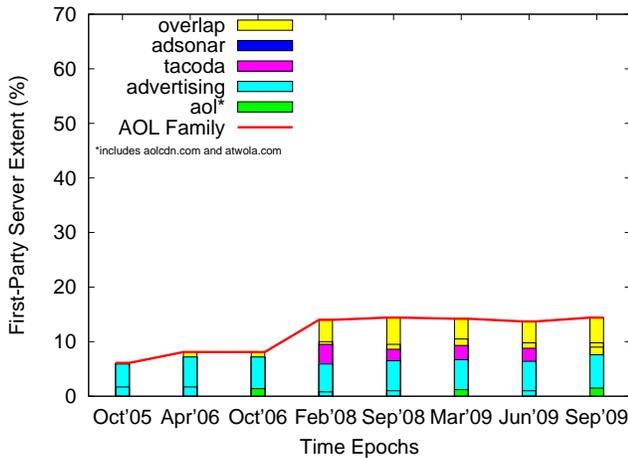
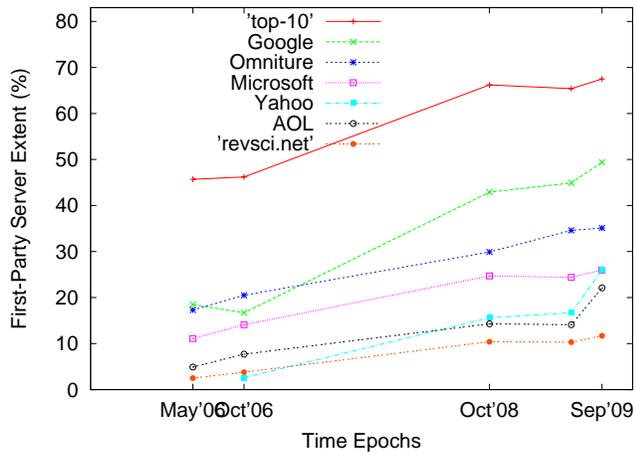
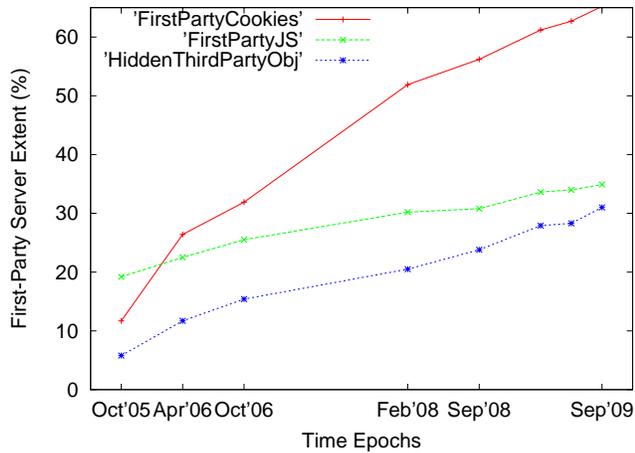


Figure 6: Growth of the AOL Family



**Figure 9: Extent of Top-10 Third-Party Families in Each Epoch and Specific Contribution of Top Families for Fiduciary Sites**



**Figure 10: Growth of Hidden Third-Party Content**