Biggish

A solution for the inverse privacy problem

Yuri Gurevich, Neta Haiby, Efim Hudis, Jeannette M. Wing, and Elad Ziklik

Microsoft Research, MSR-TR-2016-24 (May 2016)

Abstract. An item of your personal information is *inversely private* if some party has access to it but you do not. Inverse privacy is ubiquitous. Each interaction you have with commercial and other institutions generates inversely private data. The inverse privacy problem is unjustified inaccessibility of your inversely private data to you. Elsewhere a subset of these authors determined that the problem has a market-based solution that provides consumers with large amounts of their personal data to be mined and processed to benefit them. Here we sketch a particular solution.

1. Preliminaries

The notion of inverse privacy was introduced and analyzed in paper [1]. For reader's convenience we recall here some of the main points of that paper.

Your personal data splits into four buckets:

- *directly private***,** that which you have access to but nobody else does;
- *inversely private***,** that which some party has access to but you do not;
- *partially private*, that which you and a few other parties have access to;
- *public*.

(Strictly speaking there is also a fifth bucket comprising items of your personal data that nobody has access to. For simplicity we ignore the fifth bucket.)

The inversely-private bucket is much bigger than all the other buckets. This is primarily because you generate a data wake with each interaction you have with various institutions. Originally partially private the data wake quickly decays into inversely private because your record keeping ability is no match to that of institutions.

We take the position that you, the customer, has the right to access your inversely private data. (There are rare exceptions that we ignore here.) The inaccessibility to you of your personal data is the Inverse Privacy Problem. The problem is not easy. Why would institutions share your inversely private data with you? What is there for them?

Yet we argue in [1] that the inverse privacy problem can and will be solved in the near future. Your inversely-private data will be, by and large, shared with you in a form convenient to you.

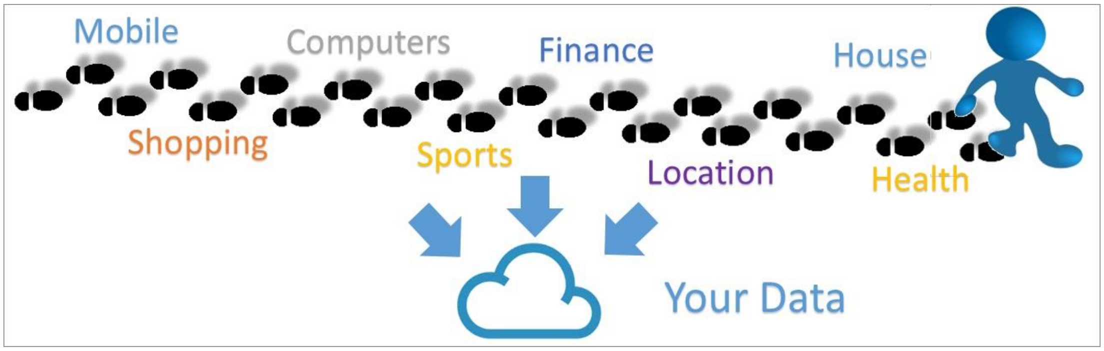
Both in industry and academia there are efforts to solve the inverse privacy problem in some vertical markets. For example, [2] is an industrial effort to solve the financial aspect of the problem, and [3] is an academic effort to solve the health-related aspect of the problem. The purpose of this paper is to sketch a proposal, a vision, to solve the inverse privacy problem in full generality.

2. The Biggish Proposal

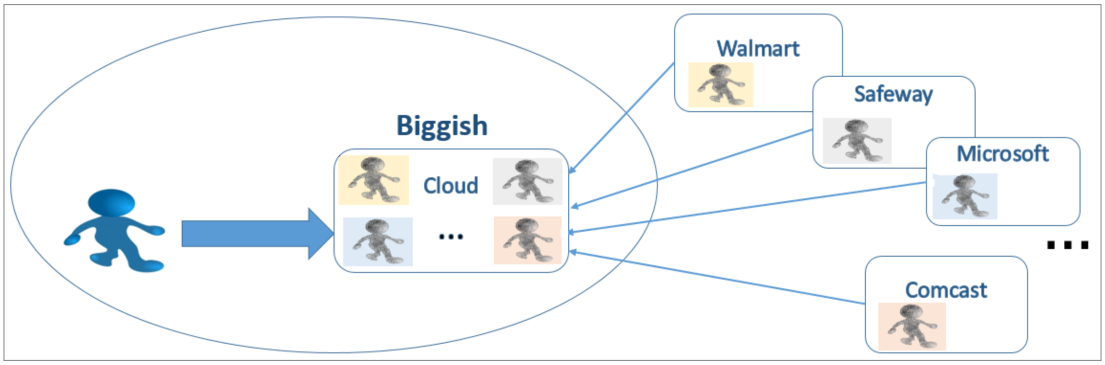
In any solution of inverse privacy problem, you acquire access to large – or "biggish" – amounts of your personal data. What will you do with all those terabytes of data? Where will you keep all that information? How will you use it?

The main goal of Biggish Proposal is to provide you convenient and secure access to your personal data that used to be inversely private. The proposal is a version of an earlier proposal with the same name put forward by Neta Haiby, Efim Hudis and Elad Ziklik.

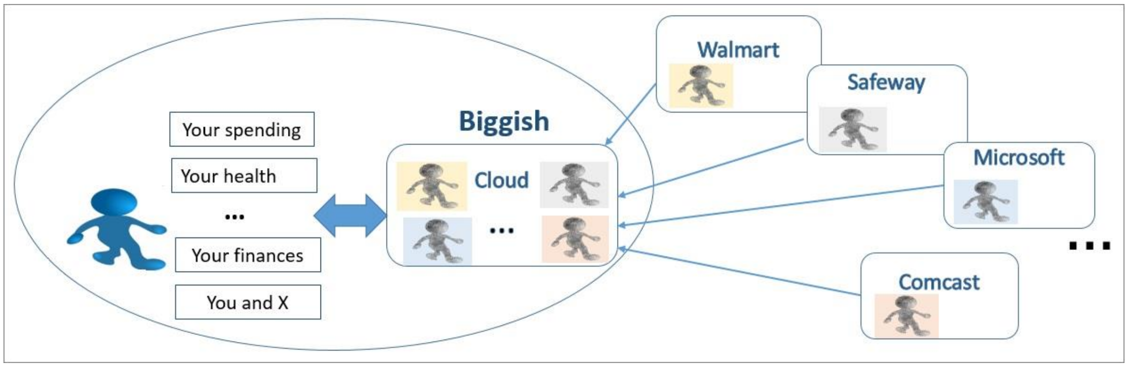
First of all, your data should be collected and securely stored on a cloud.



We may start with a simple share-back scenario:



Eventually Biggish would evolve into a trusted platform to store your personal data. Your detailed purchasing data from various sources is consolidated, and so is your data on health, finances, location, communication, etc.



Third parties will offer various analytics app to run on your data. The apps do not leak data and do not transmit information to the app creators. They are vetted to benefit only you.

Consumers are the main beneficiaries of Biggish. But Biggish seeks a win for every party involved in the process.

For example, Biggish may create an attractive Walmart Loyalty Portal, an incentive for Walmart to share back personal data. The portal will benefit you, as a customer, in various ways: to display analysis of your shopping habits at Walmart and post recommendations for you, to compare your shopping at Walmart with a real or statistical person of your choice. Walmart may use the portal to offer you discount coupons and announce new products, services and benefits.

The market place of analytics apps would benefit vendors and the economy at large.

What about the company or companies that host Biggish? Having all that data is certainly useful. Can that data be used in a way that respects customers' privacy? Yes. How? For the time being, we leave this to reader's imagination. The consumers will surely appreciate the noble efforts to solve the inverse privacy problem on their behalf.

References

1. Yuri Gurevich, Efim Hudis, Jeannette M. Wing, "Inverse Privacy," to appear in Communications of the ACM, also at http://arxiv.org/abs/1510.03311.
2. Intuit's Mint, https://www.mint.com/.
3. Deborah Estrin, "Small Data where n = me," Communications of the ACM, Vol. 57 No. 4, Pages 32–34 (2014).