General
Internet-Draft
Intended status: Informational
Expires: July 19, 2014

L. Johansson
SUNET

January 15, 2014

Linkability Considered Harmful
draft-johansson-linkability-bad-00

Abstract

Current debate on pervasive monitoring often focus on passive attacks on the protocol and transport layers but even if these issues were eliminated through the judicious use of encryption, roughly the same information would still be available to an attacker who is able to (legally or otherwize) obtain access to linked data sets which are being maintained by large content and service providers.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on July 19, 2014.

Copyright Notice

1. Introduction

This I-D is submitted as a position paper for the joint IAB/W3C STRINT workshop 2014. The author wishes to call attention to the fact that linked data sets are a source of information, sometimes every bit as useful as anything that can be gleaned from passive monitoring of Internet traffic. Such data sets are routinely generated and maintained by service and content providers and are often a source of secondary (or even primary) income for those that own and generate them.

In the current discussion on pervasive monitoring we often overlook the fact that even as more encryption is used, making passive attacks harder, focus may simply shift to active attacks on owners of linked data sets. We should strike at the root of this problem by making it less appealing to maintain these data sets and by offering users a measure of control over how their information is used and shared.

2. A Simple Example

Service providers (we use this term in a general sense, and not with a view to any particular protocols etc) typically manage users and billing records. This leads to a data set being created for every user of that service. Most services employ a simple pattern for user enrollment which relies on an email address as a means of (supposedly) uniquely identifying the user. The email address has become the defacto user identifier on the Internet.

When a user pays for the service a pair of linked data sets is created: the user data at the service provider is associated (via the credit card information) to the user data held by the credit card company. The value of the linked data, aswell as the risk to the user, is higher than the value/risk involved in the two data sets taken as separate entities. For instance the linked data says something about the buying habits of the user (based on the use of the particular credit card) which in itself is valuable information.

Linking increases the risk to the user aswell. With every service that stores the users credit card the risk of exposure to active attacks increase as events in recent years have made it painfully clear.

If this example seems overly simplified or even naive to bring up, consider the simple observation that when we visit a store in the physical world we have the ability to "browse", i.e to view and select among the offered goods without having to identify ourselves or prove our ability to pay for any of the goods in the store. This aspect of the real world has not been translated into the online world where prospective customers are routinely fingerprinted and our behaviour tracked even when we have shown no intention of engaging in a business transaction with the store owner.

Naturally there must be ways to "conduct business on the Internet", but there are ways to enable business without the need for linkable attributes. In fact there are ways to enable business using non-linkable attributes in such a way that the risk to business owners are reduced.

3. From Linkable Identifiers to Attributes

The way to avoid linking is simple (and yet so hard in practice): avoid the use of linkable attributes. In our e-commerce example above, the credit card number is a linkable attribute. However in this case the credit card is strictly speaking not needed at the service provider. When the user provides her credit card information to the service provider she is actually providing an authorization to the service giving the service provider the right to obtain payment from the credit card company.

Instead of using the credit card number as an implicit grant (of a right to obtain payment), a token that isn’t linkable across identifier domains could be used to represent an explicit grant issued on behalf of the user by the credit card company to the service provider. This is a simple example of a general pattern: instead of using a linkable user identifier, provide access to an attribute representing some property of the user that used to grant specific access.

Some credit card companies have actually taken first steps towards this by involving the user directly in a second factor authentication (typically to reduce the risk of fraud). This practice follows a model for 3:rd party authentication services (aka identity providers) commonly used in the enterprise and R&E community. Experience from the R&E identity federation community shows that access control using identity providers and non-linkable pseudonomous identifiers is by no means problem free, but can be made to work in many situations.

4. Incentives

There are strong incentives for service providers to enrich the value of their data set using attribute linking. The value of the attribute naturally increase with the inverse of the size of the set of users who share that attribute: the more specific the attribute the more valuable it is, because it can be used to identify a user with a higher degree of certainty.

Unfortunately there seem to be few costs associated with keeping large linked data sets around - stolen user credentials in the 10s of thousands rarely result in more than a brief notice in the news anymore. To date the IETF community have focused on how to avoid the use of long-term credentials (passwords) to reduce the effects of such attacks. We need to broaden our scope to find ways to disincentivize the (over)use of linkable attributes.

5. "Conclusions"

Part of the Internet economy seems to be based on linked data sets and linkable attributes. Changing this will require creating negative incentives for service providers, making it less attractive to keep data around aswell as establishing technical mechanisms that allow service providers access to the attributes they do need in order to conduct their business wo having to rely on linkable attributes. Success will depend on carefully engineering the negative incentives to match the technical mechanisms in order to promote good behaviour.

Last but not least the Internet community needs to decide what it considers acceptable behaviour wrt to attribute linking.

6. Acknowledgements

Many thanks to Linus Nordberg for important contributions and lots of interesting discussions on this topic. Many thanks also to Lucy Lynch who is the source of much wisdom, the "I’m just browsing" response to identification on the web in particular.

7. References

Author's Address

Leif Johansson
SUNET

Email: leifj@sunet.se