The irregular musings of Lou Montulli

Early web guy, engineer and entrepreneur

TUESDAY, MAY 14, 2013

The reasoning behind Web Cookies

I get a fair number of questions about cookies from individuals and the press. I thought I would try and explain some of the motivation and history behind Web cookies as well as some of the design behind them. Feel free to post more questions and I will try and expand this article with more details.

See my recent post on 3rd party cookies as well!

Motivation

HTTP is the networking language behind your browser. HTTP, or Hyper Text Transport Protocol, was designed and introduced as part of the WWW project that started with Tim Berners-Lee at CERN and was expanded upon by several universities, corporations and private individuals. The WWW and HTTP are open standards which are published and given into the public trust in order to foster interoperability and to create products that can become ubiquitous.

One of the problems faced in the early years of the web was how to create websites that had "memory" for individual users. The uses of "memory" on a website are many: shopping carts for shopping, personalized content, logging in, and many other interactive features require memory. The problem in 1994 was a lack of mechanisms to identify a user individually. HTTP was designed to be fast and efficient and part of its design was to connect to a website, grab a document and then disconnect. This freed up the website to serve other customers, but it also meant that there was no concept of a session. Without a session, each time a user clicked to move to a different page they would become just another random user with no way to associate them with an action they had done just moments ago. This is a bit like talking to someone with Alzheimer disease. Each interaction would result in having to introduce yourself again, and again, and again.

History

By the summer 1994 the idea of adding some form of "memory" to HTTP had been kicking around the WWW design groups for a while. I'm not sure exactly how long but I would guess for 1 or 2 years. There were some interesting proposals, but one of the popular ones that kept coming up was to add a unique identifier to every web browser so that a web site could individually identify each user use that to build a session. I was very much against this concept because the unique identifier could be used to track a user at every website. Without another proposal the idea of adding sessions or memory remained dormant.

Sometime around July of 1994 I went to a meeting to talk about a shopping server that another group at Netscape was working on. They needed to build shopping cart functionality into the server, but they didn't have a good way to make it work with the existing technology. They explained what they wanted to do and I had to shake my head and agree with them and tell them that they could not do it in any reasonable way with what existed. I promised to think about the problem and get back to them. At

BLOG ARCHIVE

► 2015 (1)

► 2014 (2)

▼ 2013 (15)

► June (5)

▼ May (10)

Value based home music recording

Wow, Flickr is giving away 1TB of storage for pers...

It's the end of <blink>!

Why blocking 3rd party cookies could be a bad thin...

Google "Play All" music service initial review

The reasoning behind Web Cookies

A short history of the "about:" URL

The origins of the <blink> tag

I'm trying out blogger

Fine cabinetry meets not so modern gaming

ABOUT ME

Lou Montulli

View my complete profile

Class, this blog has disappeared so I recovered these essays from the Waybackmachine.

some point during the next week the general concept of Web Cookies formed in my head. The idea of allowing a single web site to send a session identifier to the browser that would get sent back only to the server appealed to me and prevented cross site tracking. I wanted to create something that had more utility so that we could do a lot more than just shopping carts, so I extended the concept of the session identifier into a general payload that would get sent back to the server.

Design

With a rough sketch in my head I worked on the general design of cookies. The goal was to create a session identifier and general "memory" mechanism for websites that didn't allow for cross site tracking. I discussed the design with other members of the engineering team and got useful feedback, especially with regard to denial of service attracts and other security concerns. John Giannandrea was especially helpful with the design. The end result was the cookie specification that still defines 95% of what cookies do today. A Web server can return information to the browser that the browser should give back to the server each time it contacts the server in the future. Within the Web Cookie, a server can embed a random session identifier, a username, a shopping cart, or anything it wants as long as it doesn't try to send something too big or set too many cookies. There are also restrictions on how cookies can be transferred, either over secure connections or not and if the cookie should be erased when the user closes the browser or reboots their computer.

The name

I had heard the term "magic cookie" from an operating systems course from college. The term has a somewhat similar meaning to the way Web Cookies worked and I liked the term "cookies" for aesthetic reasons. Cookies was the first thing I came up with and the name stuck.

Usefulness

We released the Netscape browser in the fall of 1994 and within a year it was the most popular browser in the world. We also released the Cookie specification so that other browsers could implement it and so that web sites would know how to use it. With most of the world using a browser that supported Web Cookies, web sites started using cookies for many different things and in ways that we could not have predicted. Most of those uses were fantastic, some of them were concerning.

By 1996 and 1997 the web had grown dramatically and was a big business. Most websites offered their content for free and were advertising supported. Advertisers were looking for ways to increase the effectiveness of their ads and they looked to Web Cookies to help them with that.

An Unforeseen Problem

A problem that I missed during the Cookie design phase was an interaction between cookies and embedded content within a webpage. Webpages start as a single HTML document on one server. That one HTML document contains references to other resources that are loaded to display the site that the user sees. Images, videos, more text, and plug-ins are all references within an HTML document and any of those resources can be loaded from anywhere in the world. This referencing technique is one of the things that make the Web so amazingly powerful. When Web Cookies are combined with embedded references that point to other websites they are called "3rd party cookies" and they represent a new way in which users can be tracked across multiple web sites.

Advertising uses

Big ad companies, in particular DoubleClick, started using 3rd party cookies to track a browser uniquely across all of the sites that used DoubleClick to serve ads. The use of the tracking wasn't to actually identify the name of a user, but to make sure they didn't see the same ad every time or to track the number of unique users that saw a

particular ad. Eventually the ads were customized to reflect the browsing habits of the unique identifiers. If a browser had looked at kayaks and racquetball racquets in the past then they were given ads for sporting goods and kayaks and racquets.

Around 1996 ad tracking via Cookies became a hot topic. Users were upset with the practice and asking for change. Tracking across websites was certainly not what cookies were designed to do, they were designed with the opposite intention, but what could be done at that point to fix the problem? One of the solutions we came up with was to add controls to give each user control over what cookies to accept and from whom. The big question at the time was whether or not to disable 3rd party cookies completely or not. The decision had wide ranging effects since advertising was paying for most of the web and disabling 3rd party cookies would also disable many legitimate uses for cookies on embedded resources.

An uncomfortable decision

In the end the decision to disable 3rd party cookies or keep them on was left to me. I agonized about the decision for weeks and in the end I chose to keep them. My logic was two fold:

> Any company that had the ability to track users across a large section of the web would need to be a large publicly visible company. Cookies could be seen by users so a tracking company can't hide from the public. In this way the public has a natural feedback mechanism to constrain those that would seek to track them.
>
> If 3rd party cookies were disabled ad companies would use another mechanism to accomplish the same thing, and that mechanism would not have the same level of visibility and control as cookies. We would be trading out one problem for another.

Today, I still believe that it was the correct decision. Governments have an ability to regulate the collection of data by large visible companies and has shown a willingness to do so. The public has a responsibility to keep pressure on both the companies that have the ability to track users and governments to enact reasonable privacy regulations and enforce them. Most importantly, there are other mechanisms that can replace Web cookies for tracking if they are universally disabled, and those mechanisms would be much harder to observe and disable.

Durability

The design of Cookies have remained fairly unchanged for the last 19 years. The biggest changes have been to how users can view and control their cookies. Many people have proposed alternatives, but none have caught on. Cookies are not perfect, but they have certainly proved good enough and much of the functionality of the web depends on them.

Summary

Web Cookies find themselves in the midst of a very controversial area, and have gained a level of notoriety because of it. Even though they were designed to protect privacy they have been used in ways that are sometimes infringing upon it. Many efforts have been taken to protect the privacy of Web users and Cookie controls have been put into the hands of every user. The future of Cookies and user privacy is surely to continue as an interesting news item for a very long time and rightfully so. The nature of the advertising business is to collect as much information as it possibly can, so the public needs to push back when it goes too far.

Posted by Lou Montulli at 4:47 PM

Labels: Cookies, Software Development

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

The irregular musings of Lou Montulli

Early web guy, engineer and entrepreneur

FRIDAY, MAY 17, 2013

Why blocking 3rd party cookies could be a bad thing

3rd party Cookies are used to facilitate targeted advertising and can be used to track your browsing across multiple Web sites.

I am not comfortable with being tracked across the web by Cookies, and Cookies were specifically designed to try and prevent this sort of behavior, so why would I be against disabling the major mechanism that is used for web tracking?

The answer is pretty simple:

- The evil you know is better than the one you don't.
- This is probably a race we can't win.

The evil you know is better than the one you don't

Right now we know that advertising companies use 3rd party cookies to track behaviors and serve custom ads. The companies that use these methods are well known to us and use similar methods. They almost exclusively use 3rd party cookies. With this knowledge it is very easy to disable tracking: Go into your browser preferences and disable 3rd party cookies. For those who care there is a simple and reliable method to disable tracking.

Now lets suppose that someone decides to turn off 3rd party cookies by default in most browsers. What do you think will happen? Will the advertising industry just close up shop and move on to other businesses? Will web tracking just cease to happen? Of course not! Advertising companies will move to other technical methods for user tracking. They exist already, they are just a bit less convenient than 3rd party cookies to use. The new situation is one that will be much harder for an individual user to disable tracking.

This is probably a race we can't win

Think of this battle similar to the challenges of DRM. Companies spend lots of time and money to design a super secure system to protect their digital content and within a week or two the DRM is broken and the content is all over the internet. DRM is typically broken by small groups of unpaid hackers who do it for a hobby and they win just about every time. Now think about how difficult it will be to stop a multi-billion dollar industry from winning a similar war. If you are a browser company how are you going to stop well funded companies from coming up with technical ways to serve targeted ads? There are already multiple ways to do tracking without 3rd party cookies and countless others could be created if enough effort is put into it.

What can we do?

My suggestion is to handle this politically not technically. Keep working on Do-Not-Track and other mechanisms. Keep working on legislative methods to restrict the policies. Those are wars that we can win and those policies can have real teeth to

ABOUT ME

Lou Montulli

View my complete profile

them. One of the benefits of the 3rd party cookie tracking techniques is that it is really easy to see who is doing the tracking. If we make the societal choice to prohibit tracking then it will be easy to track down offenders.

User tracking is a hot issue and should continue to be discussed and debated. It is important to keep in mind that targeted ads are paying for almost all of the things that are on the Internet today. If we make a wholesale decision to stop serving ads are we all prepared to start paying for all of the services that we are getting for free now? Perhaps many people would agree to some level of targeted advertising if they saw the reasons and cost benefits clearly enough. What I don't think will be useful is to start a technical war that disables the current working methods for disabling tracking and replaces them with poorly understood and less visible techniques.

Posted by Lou Montulli at 11:18 AM

Labels: Cookies, Software Development

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)