Taming the Cookie Monster
How Companies Track Us Online

Paul A. Watters

Executive Summary

This paper reviews the technical need for cookies in web applications, as a means of managing state. It concludes that when used appropriately, session cookies can be privacy-preserving if they ensure that only authenticated users can access application functions and data for which they are authorised.

The paper also examines persistent cookies and concludes that while some may be helpful, others have significant negative privacy implications.

The paper recommends 5 cookie policies be considered:

1. Give users the choice to indicate wherever customisation or personalisation is required, rather than storing persistent cookies. Sessions should be managed using session cookies, and all user data should only be stored on the server-side.

2. Require explicit informed consent to be obtained from users for persistent or tracking cookies to be stored.

3. Require cookies to be stored adhering to an approved standard (such as RFC2109).

4. Users should be presented (upon request) with a copy of data being recorded about them that is subsequently used for personalised advertising.

5. Ensure that cookie standards which specify controls to prevent the compromise of cookies on browsers are verified on each browser release.

In addition, the paper suggests that in order to be able to effectively investigate privacy breaches of a technical nature, the Office of the Information Commissioner should be granted powers to seize digital evidence directly and subject information systems to forensic analysis.

The paper advises consumers how to configure their browsers and install browser plugins that prevent the collection and dissemination of their website usage by third-parties.

Finally, the paper concludes that to better inform policy development in Australia, a study to determine the prevalence of tracking cookies targeting Australian users should be undertaken, and an audit taken of measures already in place to obtain explicit informed consent when personal data is stored as a cookie.
Introduction

All distributed applications need some technical means of identifying individual clients, to ensure that only data intended for an authorised user is provided to that user and no other user. This process is known as state management. In web applications, this is typically achieved by storing a small piece of data on the client called a ‘cookie’. This cookie is passed by the client to the server as part of a HTTP request; the server processes the cookie and can then use its value to determine the client’s state vis-à-vis an application running on the server. Although state management can be achieved by means other than cookies, they are the standard means for managing state in web applications specified by the W3C Network Working Group (RFC 2109, February 1997). Note that cookies have been used in other distributed technologies – such as the X11 windowing system – and are not unique to the Web.

Session Cookies

The most straightforward use of a cookie is client identification by using a session cookie. Consider a simple example involving authentication: an internet banking application allows bank customers to perform online transactions, view balances, and so on. Before accessing any of these functions, a user must enter a valid username and password, which is transmitted to an authentication function on the server; if the user can be authenticated, a session cookie is created and passed to the client. In subsequent accesses to the server application during the current session, the cookie is passed by the client back to the server. On the server side, many different objects can be retained in memory while a user session is in progress, such as the contents of a shopping basket, or the details of a funds transfer. Once the user has completed their session, perhaps by clicking on a ‘logout’ button or similar, the session will be invalidated. Used appropriately, session cookies ensure that only authenticated users can access application functions and data for which they are authorised. In such cases, the use of a cookie can be considered privacy-preserving.
When properly used, session cookies are legitimate entities that can play a key role in securing web applications. Cookies can be controlled by applications on the server to meet operational requirements.

From a technical perspective, there is generally no need to use cookies at all; other techniques (such as URL rewriting) could be used to uniquely identify a client. Indeed, for Java applications on the Web – if cookies are disabled on the client side for some reason – the HttpSession API will revert to URL rewriting. Could this approach not solve any concerns with cookies if such a simple replacement is available, especially given the policy concerns which are discussed below?

One of the key concerns with URL rewriting of HTTP GET requests is that they often end up being logged through intermediate hosts, such as proxy servers. This could then lead to session hijacking, where an unauthenticated user could intercept the session identifier. Of course, an intermediate host (such as a router) whose network interface was operating in promiscuous mode could also potentially intercept such data – or even a cookie being transmitted – in real-time. So, perhaps the best advice to avoid interception of state data for all web applications is to transmit all data over the Secure Sockets Layer (SSL) which encrypts web traffic between the client and server. This is the type of ‘secure channel’ recommended by RFC2109, especially for data of a personal or financial nature. Thus, simply ‘getting rid of cookies’ at the technical level is unlikely to make a practical difference without clear policy outcomes in mind.

Persistent Cookies

Since session cookies last for only a short time, what are the other cookie types, and what are the main privacy concerns for them? The other key cookie type is a persistent cookie, which – as the name suggests – lasts for longer than the lifetime of a single session. Persistent cookies can have many uses, and they can be potentially very helpful. For example, if a user visits a multilingual travel booking site, and selects their language as ‘Chinese’ – and if this preference is stored as a persistent cookie by their browser - the next time they visit the site, the home page will automatically display Chinese text. Or, if a site requires registration, a persistent cookie might store the fact that someone using their browser has already registered, and prompt them to login.

Figure 1 shows an example of how Firefox allows users to view and manage all of the cookies that are stored by each particular site. In this example, several cookies (name/value pairs with an expiry date and other metadata) have been stored by drive.com.au, such as a list of recent searches. Thus, when a user loads the www.drive.com.au page, presumably these cookies are read and used to customise and tailor the user experience.

Most browsers provide a means to remove all stored cookies, or to delete them on a per-site basis (how to do this is covered later in this paper).

A complicating factor for rich web application platforms (such as Adobe Flash) is that they may store their own persistent cookies, but not within the browser’s cookie container. This makes them impossible to remove directly using a browser; thus, users may believe they have removed cookies through the browser,
but they may in fact persist within a platform’s own customised storage. Furthermore, unlike browsers, the cookie structure, processes and safeguards may not be public and will not have been discussed or approved using a community process as per RFC2109.

Tracking Cookies

Persistent cookies are sometimes also referred to as tracking cookies since there are many opportunities to use them to track broadly what any user is doing across multiple web applications. The key privacy issue which has been raised in relation to tracking cookies is the use of third-party cookies by advertisers and marketers.

When a user visits a website such as www.samplebank1.com, a third-party cookie may be set by an advertising service such as sampletracker.com. If the sampletracker.com site also sets cookies on another bank (samplebank2.com), or any other site, it is able to build up a profile based on the users’ behaviour. sampletracker.com may then use this profile to target advertising to be more relevant to that user, based on their social, economic and demographic profile.

For example, if a user visits a number of real estate websites and a banking website, they may be served up advertisements for a conveyancing service convenient to their location. Ultimately, sampletracker.com may be able to amass sufficient usage data over the long term for them to be able to determine the identity of the user, and serve up even more personalised content. This includes the unique identifiers which may be associated with IP address data that is available to sampletracker.com.

Facebook has the concept of “social plug-ins” that allow users on a third-party site, for example, to ‘Like’ other users on Facebook, without leaving their site. This is achieved by the use of cross-domain requests, where a web developer loads the code to enable the Liking (using the Cookie stored in the user’s browser), such as:

```html
<script src="http://static.ak.connect.facebook.com/js/api_lib/v0.4/FeatureLoader.js.php" type="text/javascript"></script>
```

**Figure 2** shows an example of a large collection of cookies with long expiry dates (5 years) which have been stored by a site called esomniture.com. A visit to www.esomniture.com simply displays ‘HERE’. A WHOIS search reveals that the site is actually owned by Adobe. It is unclear why Adobe have chosen to use a different domain for these cookies, providing no information to the public about what the site is used for, what these cookies are tracking, or how this data is being used. WHOIS data indicates that Adobe owns 1,596 domains.

**Figure 2.** Example of cookies esomniture.com stores, which (according to WHOIS) is owned by Adobe
Technical Implications

Using session cookies over SSL provides an effective means to manage user state in distributed applications. However, URL rewriting over SSL could provide the same functionality, where SSL would provide protection against session hijacking. It is recommended that this approach be adopted in future development where feasible, to minimise the number of cookies being stored.

Using persistent cookies does fulfill a technical need, but what are the alternatives if they are blocked or otherwise disallowed? One solution would be to require all users to authenticate wherever customisation or personalisation is required. Sessions could then be managed using session cookies, and all user data could be stored on the server-side. Thus, there are no insurmountable challenges to replacing persistent cookies from a technical perspective. However, user acceptance is likely to be a major challenge, since users who have become used to personalisation provided through cookies may reject the need to remember “yet another” password for every site they visit. Also, moving more processing and storage of data away from the client may introduce additional server-side overhead.

Furthermore, there is a risk of cookie theft if cross-site scripting requests are used on a site, and if cookies are not set to HttpOnly, ie, if they can be accessed by scripting languages such as JavaScript or JScript. More broadly, any personally identifying data stored as a persistent cookie on the client is susceptible to retrieval by malicious users through malware or malicious browsers. If criminals were able to steal a user’s cookies, they might use them for session hijacking, impersonation of the user as an account holder, especially if the cookies have a long lifetime. However, an assessment of the risk relative to opportunity is needed here; so-called “man-in-the-browser” attacks (like the Zeus Trojan) have historically targeted banking sessions to hijack sessions and steal money, rather than trawling for personal information.

A related issue is browser fingerprinting, which might allow a third-party to uniquely identify a user’s particular configuration of browser user agent, browser plugins, timezone, screen depth and other attributes as unique. The EFF Panopticlick website (panopticlick.eff.org) allows users to test whether their configuration is unique (from a pool of more than two million browsers).

Policy Implications

Different jurisdictions have proposed a number of responses to deal with cookie issues in relation to privacy.

In the US, the Federal Trade Commission has taken action in 32 cases since 2011 for violation of Section 5 of the FTC Act; the most recent case involved the placement of tracking cookies by Google Inc to target ads towards users of Safari browsers, resulting in a $22.5m settlement.

In Europe, as per the recent e-Privacy Directive, European websites must now obtain the consent of users to store “non-essential” cookies on their computers, whether their users are based in the EU or not. It’s not clear what ‘non-essential’ means, and no doubt this will be tested in the courts. It may be better to define a policy response in terms of session or persistent cookies (i.e., using technical nomenclature) to ensure that businesses are clearly directed not to use persistent cookies, for example, unless they have users’ explicit informed consent, including what information persistent cookies will allow the organisation to collect, and how the data will be used. The UK Information Commissioner’s Office has published the Guidance on the rules on use of cookies and similar technologies which provides a helpful start.

An alternate approach that would allow greater industry flexibility would be to ensure that users are presented with a description of what data is being collected about them, and ensuring that they consent to its use and aggregation for the purposes of personalising advertising. Some users may desire to see advertising which is targeted and relevant, others may disagree with such data being passed to third parties, often located outside the jurisdiction of the user. An effective policy should provide choice and include a range of safeguards to prevent against the technical compromise of cookies stored on client computers. This may require liaison with the relevant Web and Internet engineering bodies to ensure more secure designs going forward.

As per Principle 2 of Australia’s National Privacy Principles (Privacy Act 1988 (Cth)), organisations that collect personal information (including sites which are using persistent cookies for tracking) must make users aware that such data is being collected, the purpose for which it is being collected, and under what conditions the data may be disclosed to others.
The definition of ‘organisations’ in the Privacy Act excludes private sector ‘entities that are small businesses, unless they disclose personal information about an individual to anyone else for a benefit, service or advantage’ (Privacy Act, s 6D(4)(c)).

At present – apart from sites operating under European jurisdiction – this does not appear to be common practice. While businesses may argue that the uses of data are often contained in very lengthy user agreements (which are routinely not read or not fully understood by users), the notion of obtaining citizens’ informed consent is a prerequisite for authorisation of an activity. Notification should be provided to users in context, so that users are informed of how their information will be used when they make the decision to use a product or features.

Informed consent, in this context, means that users are not just told that small pieces of data might be stored about them, but how this data might be used in the future and what the consequences and risks of its use might be. Informed consent is the basis of the Privacy Act; the key issue in relation to cookies is whether users are actually giving their ‘informed’ consent.

In some cases, it may be difficult or impossible to obtain informed consent, especially where users may not be in possession of all the relevant facts, or may have insufficient reasoning ability to make informed decisions. Thus, as a child requires parental or guardian consent to be prescribed medication, should a similar model be adopted to manage children’s access to websites, or to tracking data? However, this kind of regulation is arguably ineffective, or far less effective than actual parental supervision.

A summary of the Top 5 cookie policy recommendations stemming from this paper is shown in Table 1, below.

### Advice on Investigations

Through the Office of the Australian Information Commissioner (OAIC), the Information Commissioner has responsibility for conducting investigations in relation to interferences with privacy, as defined by the Privacy Act 1988 (Cth). The Act provides the Information Privacy Principles which apply to the handling of personal information by most Australian Government agencies, and the National Privacy Principles which apply to private sector organisations. Determinations of the Information Commissioner are enforceable through the Federal Court or Federal Magistrate’s Court; the first determination (since the OAIC was established) under Section 52 was made in December 2011, where the Wentworthville Leagues Club sent subpoenaed documents to an ex-partner rather than to the applicant. Nine determinations have been made under s52.

In relation to cookies, an organisation (such an advertising company) is effectively a collector of personal information (as defined by the Privacy Act), and could be the target of investigation for a privacy breach. ‘Personal information’ includes ‘personally identifiable information’. The Privacy Act defines

---

**Top 5 cookie policy recommendations**

1. **Give users the choice to indicate wherever customisation or personalisation is required, rather than storing persistent cookies. Sessions should be managed using session cookies, and all user data should only be stored on the server-side**

2. **Require explicit informed consent to be obtained for persistent or tracking cookies to be stored.**

3. **Require cookies to be stored adhering to an approved standard (such as RFC2109).**

4. **Users should be presented (upon request) with a copy of data being recorded about them that is subsequently used for personalised advertising.**

5. **Ensure that cookie standards which specify controls to prevent the compromise of cookies on browsers are verified on each browser release.**

---

**Table 1.** Top 5 cookie policy recommendations
‘personal information’ as ‘information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion’. The Act covers the collection and use of personal information, so all uses of tracking cookies appear to be within the scope for the Commissioner to investigate.

The Commissioner has broad powers to collect evidence and inspect documents, including:

- under Section 44 of The Act, the Commissioner can request information or documents from any person as part of an investigation;
- under Section 45 of The Act, the Commissioner has the power to examine witnesses under oath;
- under Section 68 of The Act, the Commissioner may enter premises and inspect documents.

Unless there is documentary evidence available, the Section 68 powers may not be as helpful as using Section 44 and 45 together. For example, an investigation of an advertising company’s information systems could reveal the extent to which persistent cookies are used for tracking and for the subsequent identification of users, by requesting information (Section 44) and requiring the provider to take an oath that its contents are correct and complete (Section 45). An investigation could consider requesting the following (idealised) items to determine if persistent cookies are being used for tracking:

Requesting information about the extent to which IP addresses from multiple sites on which advertising appears are being grouped together. In SQL, such queries may be constructed using the SELECT statement, eg,

```
SELECT * FROM TRACKING WHERE IPADDRESS='141.132.343.1' where 141.132.343.1 is the client’s IP address
```

Requesting information about the extent to which an aggregated dataset is being used to personally identify users. Personally-identifying information (as defined by the US Office of Management and Budget) includes identifiers such as full name, national identification or insurance number, car registration details, credit card number, biometric data, genetic data, place of birth and even digital identifiers. Some or all of this information could potentially be used to recover all data known about a person, whether that information is public or private.

In both cases, an investigation could uncover whether the collector has made attempts to comply with the National Privacy Principles in relation to the manner and purpose of collecting this personal information using lawful and fair means. From a policy perspective, it may be more useful to consider an extension of the Commissioner’s powers to seize digital evidence directly and subject information systems to forensic analysis to determine the extent to which data about individuals is being collated and used.

Advice to Consumers

The best advice to consumers is for them to configure their browser or install browser plugins that prevent the collection and dissemination of their website usage by third-parties, and to regularly delete cookies that are unnecessary.

A good example of third-party anti-tracking technology is AVG’s Do Not Track plugin (http://www.avg.com.au/do-not-track/). This technology works with all major browsers – using the HTTP ‘DNT’ header - and allows users to block access to their usage data by third parties. Furthermore, it provides detailed reports to consumers on each site that they visit, about the tracking activity that is linked to each page. The plug-in indicates to users whether the trackers are collecting personal or non-personal data, and whether data is being retained or shared about them. Links are then provided to the privacy policy and ‘opt-out’ features for some of the most common trackers. The nice thing about such tools is that the number of advertising systems is relatively small, and they all tend to use fixed DNS hostnames so that their presence on webpages can be easily identified.

However, these tools only reject browser-based cookies, and not those served up by rich application environments such as Adobe Flash. It should also be noted that the DNT header is not respected by most websites, and that some online entities have signalled their intention to ignore it.
For specific browsers, the following steps can also be taken:

**Firefox**: under Privacy->Tracking, select “Tell web sites I do not want to be tracked”, and possibly select ‘Never remember history’ to prevent cookies being remembered. Users can also ‘clear their recent history’ or ‘remove individual cookies’.

**Internet Explorer**: under Tools->Delete Browsing History, users can remove all cookies; select Tools->InPrivate Browsing to not remember history; under Tools->Tracking Protection, select Automatically Block for third-party advertising and Enable.

**Chrome**: under Tools->Clear Browsing Data, users should ensure that ‘Delete cookies and other site and plug-in data’ is checked.

**Safari**: From the menu, select ‘Private Browsing’, and under Preferences->Privacy, select ‘Block Cookies’->’Always’. Users may also consider selecting ‘Limit website access to location services’ if they don’t want their location data to be disclosed.

To remove Abode Flash cookies, users should install a plug-in such as the BetterPrivacy add-on, use the Flash Settings Manager, or disable Flash entirely.

In blocking access to persistent/tracking cookies, users should remember that they will receive less-personalised advertising, and will need to re-enter their own preferences each time they access a site. For example, on the CNN website, users can choose to view the International or US editions by default, and their selection is stored in the SelectedEdition cookie. If users block such cookies from being stored, they will need to manually select their edition every time.

**A clearer picture**

While we know much about how third party tracking cookies are used, and can make specific technical and policy recommendations in this paper, it would be insightful to sample the population of Australian websites to determine the extent to which they use tracking cookies, and for what purposes. The study should include an audit of measures already in place to obtain explicit informed consent when personal data is stored as a cookie. This may help the development of further policy recommendations and also provide guidance and scoping on the scale of privacy breach investigations which may be needed in the future. For example, a recent UK study found that there were 14 tracking cookies per page on average websites from the Top 50 in the UK (based on Alexa rankings), with the majority coming from third-party companies, and with 49% being persistent.

**Conclusion**

Cookies were introduced as a convenient means to track user state in web application sessions, but their use has now been extended to support personalised advertising involving third-party providers, where users may not be giving their informed consent to have data about them used in this way.

Although some users may prefer to receive targeted advertising, there is a risk that personal information could be stored and re-used without user consent, or for a user’s identified preferences to be disclosed inappropriately or against their wishes.

The rise of third-party rich application frameworks (such as Adobe Flash) add an extra level of complexity for users who actively wish to manage what data is stored about them and tracked using cookies.

Given the EU’s recent directive on cookie use and storage, Australia should consider undertaking further analysis of the technical implications of restrictions through policy on the use of cookies, and only enact changes which are enforceable and meaningful to users.

Furthermore, Australian Internet users concerned about their privacy should use advanced settings and other software to reduce the amount of personal information tracked by websites they visit.