The Cookie Trail: Why IG Pros Must Follow the Crumbs

Mark Grysiuk, CIP

Information governance (IG) professionals must understand evolving online tracking technologies and take a leading role in ensuring that their organizations are in compliance with privacy laws and regulations that apply to them. This article provides an overview of the technologies and suggests actions IG professionals should take.

Cookies, which Webopedia.com defines as "a message given to a Web browser by a Web server," have long been a hot topic, generating both positive and negative opinions. They are important to developers and companies for enhancing website functionality and providing valuable information about website visitors. But, they are considered threats to individual privacy when they're hidden and used inappropriately.

This article provides a high-level overview of the purposes and risks of cookies and other web-tracking technologies. It also offers a framework information governance (IG) practitioners can use to incorporate ARMA International's Generally Accepted Recordkeeping Principles® (Principles) into the development of secure web applications in partnership with developers. The article can also be used to educate the user community, who are often oblivious to how their information may be at risk.

What Are Cookies?

Cookies were developed by Netscape Communications Corporation in 1995. The word *cookie* (also known as *browser cookie* or *HTTP cookie*) is derived from *magic cookie*, a term used in programming languages to describe the information shared between "co-operating pieces of software." The Cookie Central website gets a little more technical, defining it as "a text only string that gets entered into the memory of your browser."

Cookies are actually text files, about 4 KB in size, that hold name-pair values that are used to maintain *state,* which is the "application's ability to work interactively with a web user." Without it, user adoption for the Internet might not have happened.

Types of Cookies

Session cookies are stored in temporary memory and deleted when users close the browser. They contain a session ID, which keeps users logged into a website as they navigate from page to page (that is, they maintain "state" on a website).

Persistent cookies, as the name implies, stick around a little longer and are not deleted when users close the browser. How long they survive depends on the expiry date that was set by the website or when the user deletes the cookies.

First-party cookies are cookies placed on a visitor's hard drive by the initial website a person visits.

Third-party cookies are cookies placed on a visitor's hard drive after clicking on an advertisement or other content that is hosted by the initial website that person visited. It is important to note that third-party cookies are not always covered in the privacy policies that govern the original website.

Flash cookies were developed in the early 2000s by Macromedia (later acquired by Adobe Corporation). Officially referred to as locally stored objects, flash cookies improve the operability of Adobe Flash. They can do everything standard cookies do and more:

- They are persistent. No expiry dates are required. This means information can be stored indefinitely unless the developer is mandated by business requirements to include expiry dates.
- Their default size is 100 KB, which is 25 times larger than a browser cookie.
- Where they are stored in a user's machine is system-specific and not controlled by the browser. Thus, a computer running a Windows operating system will store the file in a different location than where a computer running a Mac operating system will store it. For the average user, these locations are difficult to find.
- Since flash cookies are not controlled by the browser, other browsers on the same machine can access the same flash cookies.
- Information collected is dependent on the application. To put it another way: whatever the developer wants, the developer can get!

These characteristics are the reason flash cookies are often referred to as *super cookies*.

The Rise of the Evercookie

A new type of super cookie has emerged: the evercookie. For the average user, evercookies are nearly impossible to delete. Even someone with above-average technical skills may become frustrated by them.

Evercookies use a technology called Persistent Identification Element (PIE), developed by online tracking firm United Virtualities, to recreate or copy cookies to other locations on a user's machine. *Whatis.com* defines *PIE* as "a method of individually tagging users' browsers for the purposes of identification and tracking.... [It] uses a combination of Javascript and Flash to create this tracking substitute.... The method makes it possible for deleted HTTP cookies to be respawned from stored data associated with the unique identifier."

Sam Kamkar, developer of the evercookie, elaborates further: "Cookie data [are stored] in several types of storage mechanisms…available on the local browser. Additionally, if [an] evercookie has found the user has removed any of the types of cookies in question, it recreates them using each mechanism available."

There are at least 10 storage mechanisms, including HTML5 Session, Local Storage, and Silverlight Isolated storage. For a list of available storage mechanisms, check out *samy.pl/evercookie/*.

Other Tracking Mechanisms

Other tracking mechanisms fall into the same family as cookies.

Web beacons (sometimes called *web bugs* or *pixel tags*) are small GIF or PNG transparent images (1 pixel by 1 pixel) embedded in some web pages or HTML-formatted e-mail messages. When a user opens a web page or e-mail message containing a web beacon, regardless of whether it's from a computer or mobile device, a request is sent back to the server, where a record of that request will be stored.

Web beacons are used primarily by third-party advertisers to analyze website traffic and improve the quality of advertisements. When used in conjunction with cookies, they help advertisers build unique profiles about the user.

Device finger printing pertains to information collected

Method	Description
Session fixation	The perpetrator guesses what the user's session ID might be by using e-mail phishing or brute-force searching and induces the victim to log into the web application. To mitigate the risk of such an attack, developers should ensure randomly generated strings are used to create session IDs.
Session sidejacking (or man-in-the-middle attack)	Packet or network sniffing software is used to read network traffic between two people, usually on a public Wi-Fi network. In this vector, the thief steals the session cookies while in transit and uses the information to impersonate the victim. To help prevent such an attack, developers should ensure encryption is used for the entire life cycle of a cookie, not just while in transit.
Cross site scripting (XSS)	Here, attackers use methods such as e-mail phishing or entry points such as search fields, feedback forums, and messaging boards to inject client-side script. When successful, they compromise a web application and gain access to session cookies. To mitigate the risk, developers must ensure end-user input is validated and sanitized.

Table 1: Cookie Hijacking Methods and Mitigation Strategies

online in real time from smartphones, tablets, and other computing devices. Unique characteristics are collected, including operating system, screen resolution, mouse screen position, server domain, the type of cookies stored, a postal code within a four kilometer accuracy radius, and more.

That's just the tip of the iceberg! Like cookies, device finger printing can be used to identify those devices on subsequent visits and build unique user profiles. More information about device finger printing can found at *https://panopticlick.eff.org/*.

IG's Role in Privacy Protection

As the Principle of Protection states, "Information governance program shall be constructed to ensure a reasonable level of protection to information that is personal or that otherwise requires protection." This Principle applies to *all* information collected through any technology, regardless of medium or process. IG practitioners should take an active role in ensuring compliance with this Principle, including the following.

Collaborating with Developers

IG practitioners can point developers to the guidance provided in the proposed standard Internet Engineering Task Force (IETF) 6265 *HTTP State Management Mechanism*. (Although a request for comment about this proposed standard was published in 2011, it has not been finalized.) "[It] defines the HTTP Cookie and Set-Cookie header fields."

That header includes the following attributes that, when used in conjunction with the Principles, reduce liability exposure:

- The Expires Attribute
- The Max-Age Attribute
- The Domain Attribute
- The Path Attribute
- The Secure Attribute
- The HttpOnly Attribute

For newer technologies that utilize HTML5 storage and other collection mechanisms, it's equally important, if not more so, to ensure the Principle of Protection is incorporated early in the systems development lifecycle (SDLC). Though it is beyond the scope of this article to delve into security for HTML5 and other collection mechanisms, IG practitioners can point developers to *http://html5security.org/* and to the Table 2 "Checklist for Developing Transparent Policies and Building Secure Web Applications."

Mitigating Cookie-Related Risks

Vulnerabilities related to cookies *can* be exploited by what seems to be a growing number of attack vectors. Take, for instance, *cookie hijacking* (also called *session hijacking*), which takes place when an attacker intercepts a valid session token, exposing the end user's identity credentials for logging into a remote server.

Table 1 describes three methods for initiating such an attack and ways in which developers can mitigate their risks.

Getting Executive Support

C-level leaders must understand their responsibilities for the security of information. Otherwise, the costs can be enormous.

Although the cost of the December 2014 cyberattack on Sony Pictures Entertainment is not yet known, Reuters News Agency projected it could be as much as $100 million. Robert Smallwood, executive director of the Information Governance Institute, said in a Dec. 30, 2014, LinkedIn blog posting that estimate is “way off.” He speculated that the cost could be as high as $1 billion, factoring in lost revenue; cyber insurance; recruiting, onboarding, and training IT security personnel; reputational ill will; and areas yet to be identified.

Establishing Risk Programs

Thousands of laws around the world regulate privacy, which is why many organizations are establishing privacy risk mitigation programs.

In the European Union (EU), for instance, companies must comply with amendments to the 2002 EU E-Privacy Directive. Nicknamed the “Cookie Law,” the new rules apply to any web-tracking technology.

In Spain, several investigations are under way, with fines having been levied against two companies in 2014 for non-compliance. *Computer Weekly* reported in late 2013 that six EU countries “are investigating Google Privacy policy because of concerns about personal data [collected] and stored in foreign jurisdictions.”

Since 2010, many companies have faced legal action as a result of using flash cookies. *Wired* magazine reported in December 2010 that Quantcast, an online tracking firm, “agreed to pay $2.4 million to settle a class action lawsuit alleging it secretly used Adobe’s…Flash plug-in to recreate tracking cookies.” In 2012, Amazon settled a similar case, though the details were not disclosed.

Applying the Principles

In addition to the Principle of Protection, three other Principles are relevant to the use of cookies and web-tracking technologies:

- Principle of Integrity: If session cookies can be intercepted in transit, unauthorized users can assume someone’s identity and alter the information on that system. By adhering to the Principle of Protection, an organization can protect the integrity of its information.
- Principle of Transparency: Being deceptive about the techniques used to collect information exposes a company to risks associated with consumer backlash. If fraud prevention and enhanced functionality are reasons for using web-tracking techniques, then make it clear in your privacy policy.
- Principle of Disposition: If multi-year expiry dates are set (for HTTP cookies), or if none is set at all (for flash cookies), then organizations are transferring the risk of unauthorized exposure to their customers, whose machines may contain information that is sensitive. If there isn’t a well-defined business reason for storing information on a user’s machine for a long period, then there is no reason to allow cookies or other tracking technologies to persist for several years.

Read More About It

- The Unofficial Cookie FAQ:
  www.cookiecentral.com/faq/#3.3
- Cookie Overview:
  http://itlaw.wikia.com/wiki/Cookie
- HTTP Cookie:
  http://en.wikipedia.org/wiki/HTTP_cookie#Persistent_cookie
- HTTP State Management Mechanism:
  http://tools.ietf.org/html/rfc6265
- “Flash Cookies and Privacy” Study Report:
  http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862
- Session Hijacking Information:
  http://en.wikipedia.org/wiki/Session_hijacking
- White Hat Security Statistics Report 2013:
  www.whitehatsec.com/statistics-report/2014/06/10/statsreport.html
- “Cross-site scripting explained: How to prevent XSS Attacks”:
  www.computerweekly.com/tip/Cross-site-scripting-explained-How-to-prevent-XSS-attacks; HTML5 Security resources: http://html5security.org/
- When do items in HTML5 local storage expire?:
  http://stackoverflow.com/questions/2326943/when-do-items-in-html5-local-storage-expire
- “How should application developers manage cookies?” article:
  http://searchsecurity.techtarget.com/answer/How-should-application-developers-manage-cookies
- Evercookie information:
  http://en.wikipedia.org/wiki/Evercookie
- About Device Fingerprint/Deviceprint:
  http://www.darkwavetech.com/device_fingerprint.html
- See your device fingerprint:
  http://noc.to/#

Developing Policies

In collaboration with the stakeholder community, IG practitioners can use the checklist below to develop transparent policies and build secure web applications.

The Future

From a technological perspective, a lot has happened in the last 20 years in the field of online data collection, now referred to as web analytics. Some readers may have drawn the conclusion that Http cookies' days may be numbered as HTML5 matures in the marketplace. That may be true, especially as more applications become interactive, though probably not any time soon.

There are many intranet and Internet sites that utilize cookies and web beacons and will continue to use them for a variety of well-defined business reasons. Heating up fast is the debate about whether respawning standard browser cookies after the user has deleted them is an acceptable business practice.

The world is changing fast. As such, IG governance and leadership are now a requirement for many organizations. IG practitioners must be ready to step into that role. **END**

Mark Grysiuk, CIP, can be contacted at mgrysiuk@rogers.com. *See his bio on page 47.*

Checklist for Developing Transparent Policies and Secure Web Applications

Question	The Principle	Accountability	Yes, No, N/A
Has a privacy impact assessment been conducted?	Protection	IG Team/ Senior Management	
Is the section about web-tracking technologies in the privacy policy easy to find?	Transparency	IG/Team Senior Management	
Is the policy written so it's easy to understand?	Transparency	IG Team/ Senior Management	
Does the policy explain what information is collected, why it's collected, and how long it is kept?	Transparency	IG Team/ Senior Management	
Are web beacons explained (if applicable)?	Transparency	IG Team/ Senior Management	
Is device finger printing explained (if applicable)?	Transparency	IG Team/ Senior Management	
Is the opt-out feature easily accessible?	Transparency	IG Team/Developer	
Is the application collecting the minimum amount of data that's required?	Transparency	IG Team/Developer	
Is the expires attribute set at the minimum period required?	Disposition	Developer	
Is the max-age attribute set?	Disposition	Developer	
Is the domain attribute properly set?	Protection	Developer	
Is the path attribute set?	Protection	Developer	
Is the secure attribute set?	Protection	Developer	
Is the HttpOnly attribute set?	Protection	Developer	
Are randomly generated strings used for session IDs?	Protection	Developer	
Are expiry dates coded for HTML5 local storage?	Disposition	Developer	
Are all available storage mechanisms being used? If so, are valid business reasons documented?	Disposition	Developer/IG Team	
Is user input validated and sanitized?	Integrity	Developer	
Are cookies encrypted for the entire life cycle of their existence?	Protection	Developer	