Toxic Cookies

Hal Berghel
University of Nevada, Las Vegas

The W3C has shown what can happen when bright, well-intentioned people become policy czars.

Over the years, I have referred to cookies as Web guano. I'm now of the opinion that I may have been too charitable. Cookies are far more hazardous to your digital health than anyone could have predicted 20 years ago.

I last wrote about this subject more than 10 years ago in reaction to what I considered to be inexcusable inattention to security and privacy issues in the IETF RFC2965 standard (H. Berghel, "Caustic Cookies," *Comm. ACM*, May 2001, pp. 19-22). However, the term caustic no longer does justice to the problems cookies cause. They've since moved on beyond carcinogenic, ducked into mephitic, and are now decidedly toxic—unsafe for any human consumption. And it didn't need to be this way.

COOKIE RECIPES

Cookies were created to overcome the statelessness of HTTP for Web commerce applications. The shopping cart contents had to be stored somewhere, and offloading this responsibility to the customer's hard drive required far less in the way of network and programming resources than retaining it on the merchant's servers.

I don't really have a problem with that as long as all principals—online merchants, Web applications and browser developers, search engines, operating systems, computer manufacturers, and, most of all, end users and customers—are on the same page regarding deployment and disclosures. But as some of us have explained for these many years, not everyone is behaving nicely.

Lou Montulli invented the basic HTTP cookie recipe while he was developing e-commerce applications for Netscape Communications in 1994. (Some of you might remember Lynx, Montulli's early multiplatform, hypertext Web browser.) Montulli extended the "magic cookie" programming metaphor in response to a need for client-side Web memory—the shopping cart—for one of the last releases of the original Netscape Mosaic in October 1994, just months before the browser evolved into Netscape Navigator.

The general idea was straightforward: as part of an HTTP response to a browser, a server-side platform uses a "set cookie" header to leave small amounts of digital guano (cookies) on the user's hard disk. The set cookie attributes are transaction-oriented data, such as user ID, name, date, server domain, pages visited, shopping cart contents, and, potentially, any personally identifying information (PII) the user provides during the session. While this information is stored on the user's side, it's also creating a server-side memory. How these two interrelate in any particular cookie context is anyone's guess because there are no legally enforceable standards.

Persistent cookies

Montulli's 1998 patent (US5774670) defined persistent cookies. This concept was almost immediately extended well beyond the original idea of adding client-side

Published by the IEEE Computer Society

memory to a stateless Internet protocol. Authentication cookies followed for use in monitoring current user information and connection status.

Consider what information the client and server must exchange for a commercial interest to authenticate a user. In a nutshell, it's information that the user/customer doesn't want leaked, so the process must rely on the host website's security integrity for protection.

What has the past 20 years taught us about relying on payment card systems and e-commerce environments to protect our security? Does Heartland Payment Systems ring a bell? TJX? CardSystems Solutions? For a convenient refresher, take a look at the Identity Theft and Financial Fraud Reading Room (www.itffroc.org/rr.html) or my column in *Computer*'s January 2012 issue ("Identity Theft and Financial Fraud: Some Strangeness in the Proportions," pp. 86-89).

The only reliable way to protect against such PII compromise is to prevent use of the data in the first place. A sobering thought for a world that lives on plastic.

Third-party cookies

Third-party cookies are a different story altogether. They're set with "foreign" domain tails—that is, they're different from those of the site being visited. When embedded ads in a webpage are allowed to store their own cookies, commercial interests can reconstruct or track the browsing behavior, hence the term "tracking cookies."

From the consumer's point of view, third-party cookies present an interesting case study for perfecting really bad ideas. Ad networks primarily use them to track movement between websites.

Cookies were admitted into the HTTP standards without any user awareness requirement. Although third-party cookie blocking was the default in section 4.3.2 of the original RFC 2109, the draft standard for HTTP state management, the browser developers didn't follow the standard.

Other options

The basic cookie mix serves many appetites. Session or transaction cookies are (we hope) dispatched at the end of the browser session. Secure cookies are created during SSL exchanges—for example, HTTPS sessions. There are ill-behaved, out-of-band cookies such as supercookies and Zombie cookies as well.

Although there are many variations on the cookie theme, in my opinion, all but the original recipe are half-baked.

SHEETS OF COOKIES

This quote from Wikipedia gives some idea that cookie sharing is out of control:

> The United States government has set strict rules on setting cookies in 2000 after it was disclosed that the White House drug policy office used cookies to track computer users viewing its online anti-drug advertising. In 2002, privacy activist Daniel Brandt found that the CIA had been leaving persistent cookies on computers which had visited its website. When notified it was violating policy, CIA stated that these cookies were not intentionally set and stopped setting them. On December 25, 2005, Brandt discovered that the National Security Agency had been leaving two persistent cookies on visitors' computers due to a software upgrade. After being informed, the National Security Agency immediately disabled the cookies.

We're being abused not only by e-merchants but also by our own government.

Do Not Track

Tracking cookies have been with us for quite a while. The concept is to enable server-side systems to monitor online customer behavior. Do Not Track (DNT) is now an accepted IETF HTTP header field. If a browser has DNT enabled, then tracking is prevented, right? Not at all.

When a browser sends an HTTP request to a webserver, the dialog is organized around message header fields such as GET or POST. Per IETF RFCs 2616 and 4229, some headers are considered core and must be supported to achieve IETF HTTP compliance. Others, such as DNT, are outside the core and optional.

Respecting a browser's Do Not Track request is entirely voluntary and can be ignored without penalty under IETF standards.

To put it simply, respecting a browser's DNT request is entirely voluntary and can be ignored without penalty under IETF standards. Ask yourself where the motivation for this idea came from.

The European Union has taken a more reasonable approach in its compliance model (www.ico.gov.uk/for_organisations/privacy_and_electronic_communications/the_guide/cookies.aspx). The EU's 2009/136/EC amendment to section 5C of the 2003 E-Privacy Regulation was developed to "protect the privacy of Internet users—even where the information being collected about them is not directly personally identifiable."

Not surprisingly, such concern for the security and privacy of individuals is anathema to US business interests, whose focus is primarily on increasing the consumption of goods and services and selling PII.

URL PEARLS

The current IETF cookie standard is RFC 6265 (http://tools.ietf.org/html/rfc6265), released in August 2011. Section 8 in this RFC outlines the security vulnerabilities of cookies (bearing the euphemism "pitfalls"). This includes cross-site scripting, cross-site request forgeries, session fixation vulnerabilities, ambient authority and confused deputy attacks, and replay attacks, to name but a few.

Currently, RFC 6265 "recommends" but does not require either encryption of cookie payloads or the use of secure channels such as HTTPS. Although modern browsers support user-configurable security improvements, neither of these restrictions has been integrated into browser designs.

Interestingly, cookie vulnerabilities were anticipated when the original standard was drafted in October 2000 (http://tools.ietf.org/html/rfc2965, section 7). The point is that informing and protecting end users when vulnerabilities are known should have been a priority concern when the standards were being set. Did you know that cookie vulnerabilities were anticipated in 1997? Those of us who wrote about this vulnerability at the time were either ignored or vilified by Internet snake charmers.

Basic explanations of cookies can be found at www.berghel.net/col-edit/digital_village/apr-01/dv_4-01.php; Wikipedia: http://en.wikipedia.org/wiki/HTTP_cookie; and the All About Cookies website: www.allaboutcookies.org.

Philosophically, the US views restrictions on cookies as primarily an economic issue—unrealized sales, decreased advertising revenue, increased overhead, decreased efficiencies—whereas the EU regards it as a matter of civil liberties—the right to be left alone, the right to privacy.

From a historic perspective, the W3C has tried unsuccessfully for several years to establish a DNT standard. The problem was conceptual confusion: it spent its time asking, "What does 'do not track' mean?," as if this were some Wittgensteinian grand challenge. It would have been further along if it had approached the problem from Humpty Dumpty's point of view.

The fault is that in trying to be everything to all parties, including regulators, users, and privacy zealots, as well as the 10,000-pound gorilla in the room, the business interests (merchants, advertisers, analytics services, and so on) lost their way and engendered some nonsense into the process.

The W3C has added a level of agenda-based obfuscation that parallels the gun rights debate. Introducing a sprig of DNT and a pinch of public policy/econobabble into a huge vat of self-serving business interests still yields a huge vat of self-serving business interests.

DNT isn't just deceptively simple—it's paradigmatically simple. The term means just what it says: full stop. The W3C behaves as if DNT's meaning can be found in cost-benefit studies. Linguistic absurdity and the suspension of common sense will never frame an intelligent discussion on DNT.

Of course, the business interests' mantra is that any restraint on using other people's data that would affect their profit is, by definition, over-regulation. Such is the rhetoric of Edward Bernays public relations knock-offs whose ideological mentor thought that strategies to get more people to smoke cigarettes was inspired.

Perhaps DNT isn't the appropriate operational metaphor. Maybe we should define a continuum that extends from "track me a little, but don't scar the cheeks" to "have your way with me, you global commerce she-devil."

It's possible to imagine a middle ground here. Microsoft had the right idea with DNT1, which was turned on by default in IE10. However, the application failed because the company didn't build a consensus, and, as a consequence, Web merchants thwarted its effort to protect user privacy.

By statute, Canada doesn't allow tracking, which is the only intelligent starting point when viewing users as anything more than consumers. This is all easily accomplished with that bête noir of the Web advertising and analytics crowd, the opt-in checkbox or the EU's enhanced browser settings.

Microsoft would have been far more successful if it had built "tracker tracking" features into IE to let the end user see what the servers were doing, taking a swerve around the W3C altogether.

Add-on wars

The 1990s "browser wars" made it clear that there was a lack of orthodoxy concerning compliance with W3C recommendations. As a consequence, there was no assurance that what you saw in the browser was what the webpage author intended. I coined this WYSINWOS—what you see isn't necessarily what's on the server. This disparity led me to develop the World Wide Web Test Pattern beginning in 1994 (www.berghel.net/webtestpattern).

Particularly annoying to the W3C was Microsoft's zeal at innovation. The W3C was trying to get developers to work through the approval process, while Microsoft was going its own way.

To this day, some Web portals are still designed around IE. How many times have you visited websites where the text didn't fit nicely into the text box provided? Well, the

browser wars are back, but this time the fight is over add-ons.

Speaking of which, these are three that I like adding on to Firefox: Adblock Plus (www.youtube.com/watch?v=oNvb2SjVjjI), a tunable add-on that eliminates most Web ads; Empty Cache Button 2.2 (https://addons.mozilla.org/en-us/firefox/addon/empty-cache-button), which does just what its name implies in memory or on disk; and https-finder (https://code.google.com/p/https-finder), which automatically detects and enforces HTTPS connections whenever possible.

At this point, the add-on community is doing more for browsing privacy than either the W3C or the IETF.

In 2001, when referring to Web barbarians at the electronic gates that penetrate our digital zones of privacy, I wrote, "for wont of a simple technical patch to overcome the statelessness of TCP/IP, we have created a cookie monster."

At that time, I assumed that Web merchants were for the most part acting responsibly. I no longer hold that belief—especially with respect to the Web advertising and analysis community. We need a federal statutory wake-up call ensuring that the data demands of marketing, behavioral analytics, and the like do not trump a citizen's expectation of privacy.

I'll return to the more technical side of cookie abuse in a future column. For now, I'll conclude as I did 12 years ago: "The problem society has to deal with is whether the collection of personal information about an individual without the individual's informed consent should be tolerated."

Hal Berghel, *Out of Band column editor, is a professor of computer science at the University of Nevada, Las Vegas, where he is the director of the Identity Theft and Financial Fraud Research and Operations Center (itffroc.org). Contact him at hlb@computer.org.*