

Online Reach — Are Your Visitor Counts Half-Baked?!

By Gary Angel

comScore, one of the leading web-traffic measurement companies, dropped a bombshell of sorts on the web analytic community when they released a study last week concerning the effects of cookie deletion on web visitor counting. This may not sound like the sort of topic to stir up angry debate, but the implications of the study are serious and far-reaching. They could potentially change your view of how much traffic your site (or partner sites) actually gets, how well your online campaigns are performing, how reliable many of your web analytic reports may have been and even how you are going to proceed with online measurement in the future.

Let's start at the beginning. A cookie is a small data file placed on your computer when you visit a site. The cookie is put there by the site your visiting, and it's used to identify your specific computer as one that has visited the site before and tie together behavior from visit to visit.

But before I go on, that last statement needs clarification. Because while it technically applies to every kind of cookie there are two different cookie types and the way they work makes for a subtle difference in understanding this last statement. The first and most common type of cookie is called a 1st Party cookie. And it's really what I've been describing. A 1st Party cookie is placed on your machine by the site you visited. Simple.

But the clever folks who do advertising and measurement on the web had a kind of sneaky idea. If they place a single object on a site you visit (like a tiny image) and that object is served by them not by the site your landing on, then the image request is made to THEIR computers. And, along with that request, they can drop a cookie on your machine. And they can read THEIR cookie, whenever you visit ANY site that has one of their images on the page. This is called a 3rd Party cookie. It doesn't work any different than a 1st Party cookie — but it does mean that the owner of the 3rd Party cookie can see something that the owner of the 1st Party cookie can't. Namely, what site(s) you visit — with the emphasis on the plural.

Who drops third party cookies? Pretty much every Internet advertising and measurement company. Google, Yahoo, MSN, Omniture, WebSideStory, DoubleClick, ValueClick, yada, yada, yada. Keep in mind, the 3rd Party cookie is anonymous — it doesn't know who the visitor is. It isn't a threat to privacy and it isn't spam.

But despite all these "is nots", the 3rd Party cookie has gotten something of a bad name. And, unlike 1st Party cookies that often serve a clearly useful function (like enabling portal customization), 3rd Party cookies have no tangible consumer

benefit. So as browsers evolved and Spyware companies latched onto the idea that 3rd Party cookies weren't good, problems arose in 3rd Party cookie measurement.

You see, a user can configure their browser to reject all cookies, no cookies or just one kind of cookie. Since many users realize that 1st Party cookies are actually useful, they choose to accept 1st Party cookies but not 3rd Party cookies. This led to a rapid increase in 3rd Party cookie rejection — a problem that meant many visitors were slipping through the measurement cracks. In addition, Spyware companies built lists of common 3rd Party cookies from advertising and measurement companies and generally set themselves up to delete these cookies whenever they found them.

Bad news for measurement. But the web analytics vendors (at least) had a pretty good response. They changed the way their systems worked so that the cookie being used for measurement was actually one dropped by the site you're visiting. Suddenly, it's a 1st Party cookie. It's much less likely to be rejected and much less likely (we thought) to be deleted. All is again good in the world of web analytics.

Because, you see, the cookie is vital to the web analytics world as the sole way of tracking visitors. Not page views, which are easily tracked. Not visits, which can be tracked sans cookie. But visitors. Cookies are the only web analytics technique in popular use for tracking the behavior of **visitors** and for counting how many visitors a site actually has.

Here's a short list of measurement tasks you can't do without a cookie:

- Count Visitors
- Track Campaign Success over time
- Distinguish between New and Repeat Visitors
- Distinguish between Prospects and Customers
- Analyze behavior across more than one of your sites
- Analyze your Sales Cycle
- Measure over-time Engagement / Loyalty / Frequency of Visitors

The list could be extended indefinitely, but the basic idea is this. Without cookies, web measurement has no understanding of any "visitor" behavior — only visits and page views.

Which brings me to the comScore study. comScore's study tracked one very large internet portal site and one large ad-serving network. The portal site was dropping a 1st Party cookie and the ad-serving network was dropping a 3rd Party cookie. Here is a summary of their results from the comScore Press Release:

The study examined the degree to which users cleared cookies from their computers, thereby causing servers to deposit new cookies and potentially leading to overstated estimates of unique users when relying on cookie-based server data.

Average Computer Receives 2.5 First-Party Cookies per Site Each Month

comScore observed that 31 percent of U.S. Internet users cleared their first-party cookies during the month. Within this user segment, the study found an average of 4.7 different cookies for the site. Among the 7-percent of computers with at least 4 cookie resets, comScore counted an average of 12.5 distinct first-party cookies per computer, accounting for 35 percent of all cookies observed in the analysis.

Using the total comScore sample as a basis, an average of 2.5 distinct first-party cookies were observed per computer for the site being examined. This indicates that Web site server logs that count unique cookies to measure unique visitors are likely to be exaggerating the size of the site's audience by a factor as high as 2.5, or an overstatement of 150 percent.

comScore Cookie Deletion Analysis — 1st Party Cookies*

December 2006

Source: comScore, Inc.

Number of Cookie Deletions/Resets	Percent of Computers	Avg. No. of Cookies per Computer	Percent of Cookies
Total Sample	100%	2.5	100%
1 or more	31%	4.7	58%
4 or more	7%	12.5	35%

**Excludes log-in cookies*

“While past studies from other research companies have shown a similar proportion of computers that clear their cookies, the comScore study is the first to highlight the disproportionately high percentage of cookies represented by those computers,” commented Dr. Magid Abraham, President and CEO of comScore. “For example, with just 7 percent of computers accounting for 35 percent of all cookies, it’s clear that a certain segment of Internet users clears its cookies very frequently. These ‘serial resetters’ have the potential to wildly inflate a site’s internal unique visitor tally, because just one set of ‘eyeballs’ at the site may be counted as 10 or more unique visitors over the course of a month. The result is a highly inflated estimate of unique visitors for sites that rely on cookies to count their audience.”

Third-Party Cookies Deleted at Nearly Same Rate as First-Party Cookies

comScore's analysis of third-party cookies revealed an average of 2.6 distinct cookies per computer in December, indicating a similar rate of overstatement as the first-party cookies. For those computers where at least one cookie reset occurred, the number of third-party cookies observed was slightly higher than first-party cookies at 5.5.

Everyone in the web analytics community has long understood that people sometimes delete their cookies. However, there are three aspects of the comScore study that are particularly significant. First, the percentage of cookie deletions in a month is on the high-end of expectations (31%). Second, the frequency of cookie deletion is much higher than expected — meaning that cookies aren't just deleted once a month. Indeed, the biggest impact comes from a group of visitors who appear to delete cookies after every session (a settable option in Firefox). These “serial deleters” — if they are frequent visitors to a site — will get counted over and over again as distinct visitors. Third, 1st Party cookies fare no better than 3rd Party cookies when it comes to deletion. This last result is very surprising to most of us — and indicates that cookie deletion is not typically a function of spyware protection software (much more likely to delete 3rd Party cookies) but of browser settings or manual erasing.

Since comScore focuses on visitor counting, most of the discussion around these results has also focused on this. And it's an important issue. Some in the web analytics community argue that errors in estimating Reach aren't that important — since these numbers are used for comparative and trending purposes.

Unfortunately, that isn't really true. First, most companies these days are evaluating Reach by marketing channel — so it's important to know how the web site compares to other possible marketing efforts. And second, even on a comparative basis, there is every reason to believe that different sites (and even the same site over time) will be differently impacted. At least two (and probably more) factors will obviously drive differences in site measurement: the degree to which a site has lots of heavy repeat users will significantly impact the degree of overstatement in counting and the degree to which the user population uses Firefox will probably have a significant impact as well.

Because of this, a portal site with a very high return rate will likely show a much higher degree of traffic inflation than an eCommerce site. A daily news site will show higher traffic inflation than a health portal. A site for technical users that is heavy on Firefox adopters might be much worse off than a cooking site. Sites that raise privacy issues or concerns (like pornography or medical sites) may be harder hit. And so on. Worse, if you change your site in ways that increase Engagement, this change will artificially increase your Reach numbers as well. Meaning your own site may not be comparable to itself over time.

In addition, this implies that no **simple** adjustment to web analytic numbers is applicable. You can't just take web analytic visitor counts and divide them by some correcting factor.

As troublesome as the implications are for Reach, they are even worse when you consider the impact on your broader web measurement issues. Remember the list of tasks that rely on cookies? Many of these analyses (including campaign attribution) are going to be seriously distorted unless the analyst controls carefully for the cookie deletion problems. Even with thoughtful planning, some types of analysis (such as how New Visitors behave) may be nearly impossible for some sites.

Are there good alternatives to using half-baked cookies? Not always. Certainly, there are alternatives for measuring your Reach as comScore is no doubt happy to point out. But from a web measurement perspective, there aren't many easy solutions. However, being aware of the extent of the problem may be at least some protection. A deep understanding of the problem may keep you from acting on an analysis based on flawed data and secondly it may help you screen-off your analysis from the effects of cookie deletion. There are, in many cases, techniques that will accomplish this. In addition, understanding the extent of the problem may cause some sites to focus on the possibility of "opt-in" measurement — getting visitors to register and log in consistently. It may also drive the measurement community toward a better technical solution.

No matter how you look at it, the comScore study isn't good news. But the right answer is probably not shooting the messenger. Web analytic data has always had a high degree of messiness. But the fact that data quality isn't perfect doesn't also mean that any level of data imperfection is acceptable. An analyst must be able to understand the issue and all its measurement implications to be able to make a sensible judgment about what to do differently and how to adjust site measurement. If you fob off the comScore study with a Gallic shrug, "That's just the way web data is..." then it won't just be your cookies that are half-baked!