
Papers

The challenge of digital marketing attribution across internet devices

Received (in revised form): 15th August, 2014



Christopher Hogan

manages an analytics team at Macys.com, responsible for marketing, customer and site analytics. He has over 16 years of experience in online media and e-commerce and has held strategic, analytic and operational roles with America Online, Oliver Wyman Consulting and Yahoo! Chris has spoken at numerous industry conferences and has filed multiple patents in the area of digital marketing. He holds a BA with honors from the University of Connecticut and an MBA from the University of Southern California.

Vice President, Marketing Analytics, Macys.com, 14th Floor, 680 Folsom Street, San Francisco, CA 94107, USA
Tel: +1 415 422-1993; E-mail: christopher.hogan@macys.com

Abstract With the rapid adoption of smartphones and tablets, the number of internet-enabled devices per internet user has increased over the last several years. As users employ multiple devices to complete their online objectives, the data used to understand online behaviour are increasingly fragmented. Piecing together a complete picture of activity requires the ability to identify users on every device used. Without user identity, there is no way to connect events on one device to events on another. This has implications for measuring the value of marketing investment as marketing attribution requires the ability to attribute some desired behaviour to marketing activity that occurred in the past. For websites that do not require login or registration and therefore do not have a source of user identity, the accuracy of marketing reporting will decrease as the number of devices used increases. This paper explores the issue of data fragmentation, its impact on the ability to measure marketing effectiveness, and possible solutions for addressing the issue.

KEYWORDS: marketing, digital, attribution, mobile, identity, reporting

INTRODUCTION

The breath-taking growth of smartphones and tablets over the past few years, coupled with the ubiquity of desktop and laptop computers for office use, has fragmented the online experience as internet users increasingly switch between phone, tablet, laptop and desktop computers to accomplish their online objectives. In our ever-connected world, an internet-enabled device is never far from reach. Device usage follows our daily routine, where smartphones

are used during the morning and evening commute hours, desktop and laptop computers are used at the office, and tablets see peak usage during prime-time evening hours.¹

While the development and adoption of new internet-enabled devices has resulted in valuable new features and has led to increased online activity as visitors take advantage of 'found time' they would not previously have spent online,² the use of multiple devices by individual visitors

has created many challenges for digital marketing and analytics professionals who are dependent upon analytic tools developed for a single-device world. The following paper examines some of these challenges through the lens of a large e-commerce retailer.

It is important to keep in mind that not all web sites will be equally impacted by the growth in multiple device usage. Making an online purchase can include an extended research period and the transaction itself may be easier to complete on a larger device. The convenience of mobile devices combined with the utility of desktop devices is likely to result in high rates of multiple device usage for e-commerce sites. Contrast this with media sites such as Facebook and Yahoo! where the objective is generally completed in one session and does not involve a relatively complex transaction. Mobile devices support media objectives with little need for other devices.

IDENTIFYING VISITORS

For e-commerce sites, the growth in smartphone and tablet adoption has given rise to new paths to purchase across devices and has altered usage patterns as customers complete their shopping objectives. Analytics providers are scrambling to keep up with these changes. Cookie-based tracking, still the dominant approach to analysing internet usage, works well for the understanding of behaviour over time if visitors browse on the same device, but cookie-based tracking is inadequate when visitors are active on multiple devices because reporting systems have limited ability to stitch together visit data across devices.

There is simply no way to determine whether a visitor on one device is the same visitor on another device without requiring login, registration or some other way to identify visitors. Even if real identity were available, many reporting systems are not designed to track behaviour over time using real identity. Instead, they use anonymous

cookie IDs to tie visits together. Although cookie-based tracking has its limitations (deleting cookies and using multiple browsers are two notorious sources of noise), it is still the most widely adopted approach for tracking visitor events over time.

Requiring visitors to login to a site makes the task of stitching together a comprehensive view of online activity much more possible since you can match visitors across devices. Regardless of how sites choose to keep track of visitors, whether through email address, username or customer ID, as long as the identifier is the same on every device used, then piecing together activity across devices merely requires joining disparate data using the unique identifier. For sites that do not require login, constructing a longitudinal view of usage across devices can only be done by looking at the subset of visitors who have identified themselves on every device used. And unfortunately this subset may be pretty small.

If a website can identify 20 per cent of unique visitors and the identification rate is roughly the same across devices, then the number of unique visitors that can be identified across two devices can be estimated as 20×20 per cent, or 4 per cent of the population of multiple device users. Add another device to the mix and the number of visitors that can be identified across three devices is $20 \times 20 \times 20$ per cent, or 0.8 per cent of the population. These estimates assume an even distribution of identified visitors across the population of each device, which may not be the case. Nevertheless it is safe to say that understanding usage over time is more difficult with each new device used. And that is a problem, because analysis of Macys.com's visitors found that those who visit on multiple devices spend far more than those who visit on a single device.³ Not only do you want to understand these customers better, you also want to improve their experience.

The inability to match visitors across devices makes it difficult to create seamless

experiences across customer touch points. Anonymous customers who add products to their shopping cart on their phone on their way to work will find an empty cart when they switch to their desktop in the office to complete the purchase. And this can be a real drag on conversion because switching devices to complete a purchase is not uncommon. Analysis of Macys.com's visitors found that those who purchased within a week of using their smartphone were nearly twice as likely to purchase on a desktop or tablet versus their phone.⁴ These findings suggest that smartphones can almost be viewed as 'top-of-the-funnel' devices, where usage results in more qualified traffic when visitors return to the site on a different device.

Multiple device usage presents difficult challenges for site and marketing analysts. Ask most sites that do not require login to report the number of unique visitors, and they will either admit that they do not know or they have not realised the flaws in their reporting. The truth is that many internet businesses have few internal tools that accurately report the number of unique visitors, one of the most fundamental metrics used to measure online activity. Without the ability to match customers across devices, there is no way to identify duplicate customers who visit on multiple devices and no way to determine an accurate count of unique visitors. You can certainly measure the number of unique visitors on a single device, but not across all devices.

And the accuracy of the count of unique visitors decreases as the reporting period increases, because with time it is more likely that internet users will have visited a site on multiple devices. This makes estimating the count of unique visitors challenging even if you know the number of users who have ever visited on multiple devices. Panel-based reporting systems offered by companies such as comScore have the ability to identify duplicate visitors across devices and estimate unique visitors, but panel-based reporting is limited in its ability to provide a

comprehensive and detailed view of internet behaviour for individual businesses. And at the end of the day, panel-based metrics are only an estimate.

The inability to determine the number of unique visitors is troubling, not because knowing the number of unique visitors is critical to success, but because it is representative of the larger issue of data fragmentation. As device usage increases, the ability to see a complete picture of online activity decreases. Today's web analyst must not only analyse data across three or more different versions of a website (desktop, mobile, tablet), they must also piece together data across devices to fully understand online activity because looking at each device in isolation can lead to incorrect conclusions.

Consider digital marketing attribution where the marketing analyst attempts to attribute online activity (eg online orders) to a marketing campaign. Often this attribution includes a 'lookback window' where online activity that occurs today is attributed to marketing campaigns that ran in the past. Seven-day, last-click attribution, perhaps the most popular approach for digital marketing attribution, works by attributing online activity to the most recent marketing channel that can take credit for the desired online activity, as long as the campaign ran no more than seven days prior to the activity.

Regardless of the length of the lookback window or whether the first or last marketing channel receives credit, lookback attribution requires that reporting systems look across a sequence of site visits to attribute online activity to marketing campaigns. Use of multiple devices by individual visitors reduces the accuracy of marketing attribution, because reporting systems have no way to attribute activity across visits if those visits occurred on multiple devices. Email marketing is perhaps most impacted by this inaccuracy, because email marketing has seen the highest rates of mobile growth compared to other forms of digital marketing.⁵

Using email marketing as an example, when a customer clicks through an advertiser's marketing email on their phone but subsequently uses their desktop to place an order, the email campaign does not get credit for that order. Because conversion rates on smartphones tend to be much lower than other devices, it would be easy to draw the incorrect conclusion that marketing investment on the smartphone is less efficient than other devices and should be reduced. To calculate an accurate return on marketing investment, advertisers must include sales that occur across all devices. As previously stated, looking at device performance in isolation can lead to incorrect conclusions.

In addition to mobile penetration, email marketing is different from other channels in that companies typically know who is visiting a site when they arrive via email. Because email subscribers are more likely to be identifiable, email campaign data are useful for understanding the level of inaccuracy introduced as a result of customers using multiple devices. In other words, if I can see that a visitor came to my site via email on their phone two days ago and the same visitor visited the site today through their tablet and placed an order, then I can estimate the level of misattribution introduced through the use of multiple devices.

THE SCIENCE OF DIGITAL MARKETING ATTRIBUTION

Analysis of Macy's email subscribers which sought to quantify the value of incorrectly attributed sales found that sales attributed to email marketing would have been 10–15 per cent higher if sales were attributed across devices.⁶ While these findings are admittedly narrow in applicability, they do suggest that measuring return on marketing investment is indeed impacted by multiple device usage, and that cookie-based marketing attribution may be inadequate in calculating an accurate return on marketing investment, at least in certain circumstances.

It should be noted that email services providers, because they can tie an email address to customer activity across devices, can produce a relatively accurate view of email performance compared to analytics solutions that use cookie-based tracking. That being said, the drawback with all channel-specific reporting tools is that they generally overstate performance for the channel measured. The choice of email marketing for this analysis was not due to lack of data on email performance across devices but because, unlike other channels that cannot tie behaviour to a unique identifier, email marketing data provide a useful benchmark to measure the inaccuracy introduced through use of cookie-based tracking.

Is email representative of other marketing channels? Perhaps not, due to high mobile penetration rates, but marketing investment on devices is certainly increasing⁷ and it is only a matter of time before other channels are also measurably impacted. Further, the level of inaccuracy across all channels grows with the adoption of internet-enabled devices. And the number of internet-enabled devices per internet user is expected to continue to increase even after strong growth in the adoption of smartphones and tablets over the past few years.⁸

So, after years of perfecting the science of digital marketing attribution, the online advertising industry is facing the prospect of losing some of its ability to attribute online sales or other activity to digital marketing investment. And this is a problem because investments that may appear to have performed poorly when looking at a single-device view of the data may in fact be performing well if we look at activity across devices. While certainly not at the level of offline media with respect to measurement precision, it is definitely a step backwards. Yes, marketing attribution has always been an inexact science, and digital marketing is no different, but if projections for the growth of internet devices are correct, the online advertising industry must find new

approaches to ensure a plausible level of accuracy. When your value proposition is based in part upon the perceived ability to precisely quantify results regardless of actual precision, threats to accuracy threaten the continued growth of online marketing.

For individual marketers, the concern is about finding approaches to ensure the optimal allocation of marketing budget. For sites that require login (and social identify platforms such as Facebook Connect have made this option more feasible), it is simply a matter of shifting from cookie-based tracking to identity-based tracking across site visits. It may require some investigation to determine which analytics providers offer this approach, but the data are certainly available to attribute sales across various devices. For sites that do not require login, the options are less straightforward. Nevertheless, let us review a few approaches that may be worth considering.

One approach requires the use of third parties to collect and connect identity across multiple sites. With this approach, a visitor identified on site A is tied to a unique device ID. When that user visits site B on the same device, the third party can retrieve the identity collected on site A to identify a visitor on site B who would otherwise not have been identified. The success of such a scheme is dependent upon the amount of visitor overlap across clients, the identification rate of internet visitors in the client network, and the size of the client network. To help safeguard customer data, 'data management platforms' can offer a layer of data separation and data protection. In addition, 'data onboarding' providers can help augment identification rates. If companies are comfortable with identity 'pooling' and are in a position to hire the right marketing analytics provider, this may be an option for improving the accuracy of marketing analysis.

Panel-based reporting providers also have the ability to match customers across devices, but they do not need to acquire and match

identity through login across each device used and site visited. Instead, customers on the panel volunteer their identity to panel providers and these providers then tie actual identity to a unique device identifier in order to track individual behaviour across devices. Through this approach, panel-based analytics providers are able to offer reporting that accounts for a customer's journey across devices. Panel-based reporting, however, is unlikely to cover all reporting needs or even all marketing channels, meaning multiple sources of data to reconcile. And again, panel-based reporting is only an estimate.

Predicting the identity of visitors based on unique usage patterns is another possible approach. In theory, if I can 'fingerprint' a visitor on a device where identity is known, then I can use that fingerprint to predict identity on a device where the visitor is unknown. With this approach, the data are not necessarily pooled across clients, because the fingerprint can be based on patterns observed on a specific site. Although theoretically possible, whether this approach works in practice is unclear. One advantage to this approach is the ability to use identity information to improve both marketing attribution and site experience, since the data can be stored internally. Contrast this approach with closed third-party reporting systems where user matching is done by centralising identity within external data systems.

Simply estimating lost sales is an option for companies with even the most basic analytic capabilities. Through the internal analysis of granular customer behaviour data, if I know \$1m in sales was not credited to search marketing, for example, based on the group of identified users who searched Google on their phone then switched to their office computer to place an order, then I should be able to estimate sales for the total population of visitors who searched on one device then switched to another. If I know I can identify 20 per cent of the population, and the identifiable population

is representative of the total population, then \$1m can be extrapolated as \$5m in lost sales. While this approach requires little additional data or investment, it is complicated by fact that you need to know the size of the total multiple-device population.

While I can count the number of identified customers using more than one device, there is no way to determine the size of the total population, both identified and unidentified, that is active on multiple devices. After all, I can't match customers across devices without a unique identifier. And if I don't know the size of the total population, I cannot estimate the total value of sales that should be reallocated to search marketing. The count of unique visitors who are active on multiple devices may be attainable through panel-based reporting systems or other providers that match users across devices, so one should not dismiss this option altogether. A mix of internal and external data may improve upon the accuracy of existing reports.

One sure fire approach to reduce reporting error is to increase the percentage of identified visitors. While requiring login is a non-starter for many online businesses, creating stronger incentives for login and registration in order to increase the identification rate may be possible. That being said, the benefits of login and registration are generally well understood at this point and many sites may already have registration as a goal. Does improved reporting accuracy offer an additional incentive to drive visitor login? Yes, but how much more compared to the benefits of improved personalisation and better retention rates?

As each company is different, each of these potential solutions will have its own strengths and weaknesses. There is no one right solution and whether any of these options should be pursued is dependent upon many factors, including business objective, level of marketing investment, identification rate of the visitor base, the

extent to which visitors are active on multiple devices, and the cost of possible solutions, among other factors. As discussed, gauging inaccuracy, and therefore the potential financial upside, is tricky. This makes estimating return on investment difficult. It is safe, however, to assume that the larger the marketing budget the greater the benefit in increased marketing efficiency through improved tracking and reporting.

IDENTITY SERVICES

Those unsure of whether to take action may be tempted to assume that their existing analytics provider will tackle the problem in some future release. While this may be true eventually, it could take some time. The need for real identity, a key component of possible solutions, is less attainable than ever in the wake of the Target data breach⁹ and numerous other newsworthy cases of data theft and abuse, when customers are increasingly concerned about privacy and data security. Increasing public demand for control over individual data is coming at a time when the need for such information is greater than ever in order to understand and support new patterns of online behaviour. These developments certainly make reliance on third-party vendors for visitor identification a riskier proposition from a data governance and customer relations perspective.

Given their high online penetration rates and login-dependent offerings, companies such as Google, Facebook and Apple could in theory offer identity services to other providers or build them directly into their advertising or analytics products. Google, in particular, as the owner of one of the leading web analytics products in Google Analytics, is well positioned to leverage its own customer data to augment Google Analytics client data and improve cross-device tracking. This development seems unlikely, however, as Google is being pressed to satisfy both customer and regulatory demand by

providing its customers with options to disable tracking and data collection. The other major web analytics providers are equally large targets, in IBM and Adobe, and similarly unlikely to jump in with reporting enhancements that would potentially violate customer trust.

All factors considered, there is no clear solution for piecing together a longitudinal view of customer behaviour across devices in order to improve reporting and enable seamless experiences across devices. Until the digital advertising and analytics industries develop techniques to construct a complete picture of behaviour across customer touch points while safeguarding customer identity, a practical first step would be for companies to investigate the extent to which data fragmentation is compromising reporting accuracy. After all, not everyone will be impacted to the same extent. More importantly, acknowledging and understanding the issue is better than assuming your reports are as accurate as ever. If online businesses know they are not seeing the complete picture of performance, they will be less inclined to take action that may ultimately hurt performance.

References

1. comScore (2013) '2013 Mobile future in focus', comScore Research Report, available at: http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/2013_Mobile_Future_in_Focus (accessed 22nd October, 2014).
2. Lookout (2012) 'Mobile mindset study', Harris interactive survey of 2097 adults conducted from 8–10 May, 2012 for Lookout, Inc., available at: https://www.lookout.com/static/ee_images/lookout-mobile-mindset-2012.pdf (accessed 22nd October, 2014).
3. Identifiable customers who visited Macys.com on more than one device spent 92 per cent more on average in December 2013 than identifiable customers who visited on one device. Customers who used a second device but cannot be identified on the second device used are not included in the calculation of average spend for multiple-device customers. The exclusion of these customers may reduce average spend for customers who used more than one device in December 2013.
4. Analysis of Macy.com email marketing customers who used multiple devices during July 2013 found that sales attributed to email marketing would have been 14 per cent higher in July 2013 if sales could have been attributed across devices for customers who visited the site via email marketing on one device but subsequently made a purchase on another device within seven days of their initial visit. Only sales for customers who accessed the site directly on subsequent visits are included in the calculation of unattributed email marketing sales for July 2013.
5. Movable Ink (2014) 'US consumer device preference report', Moveable Ink Research Report, available at: <https://s3.amazonaws.com/movableink-marketing/Movable+Ink-+US+Consumer+Device+Preference+Report-+Q1+2014.pdf> (accessed 22nd October, 2014).
6. Among identifiable customers who made a purchase on Macys.com within seven days of browsing on Macys.com via smartphone in June 2013, 48 per cent of customers made a purchase on a desktop computer and 15 per cent made a purchase on a tablet computer. Because identifiable customers are more likely to purchase on the device on which they are identified, device 'crossover' rates are assumed to be conservative, and non-smartphone purchase penetration for smartphone browsers may be higher than 63 per cent.
7. eMarketer (2013) 'Mobile gains greater share of search, display spending', eMarketer analysis, available at: <http://www.emarketer.com/Article/Mobile-Gains-Greater-Share-of-Search-Display-Spending/1010148> (accessed 22nd October, 2014).
8. Evans, D. (2011) 'The internet of things: How the next evolution of the internet is changing everything', Cisco Systems Research Report, available at: http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.
9. Weiss, E.N., Miller, R.S. (2014), 'The Target Data Breach: Frequently Asked Questions', available at: <http://www.fas.org/sgp/CRS/misc/R43496.pdf> (accessed 17th October, 2014).