Affiliate Crookies: Characterizing Affiliate Marketing Abuse

Neha Chachra
nchachra@cs.ucsd.edu

Stefan Savage
savage@cs.ucsd.edu

Geoffrey M. Voelker
voelker@cs.ucsd.edu

Department of Computer Science and Engineering
University of California, San Diego

ABSTRACT

Modern affiliate marketing networks provide an infrastructure for connecting merchants seeking customers with independent marketers (affiliates) seeking compensation. This approach depends on Web cookies to identify, at checkout time, which affiliate should receive a commission. Thus, scammers "stuff" their own cookies into a user's browser to divert this revenue. This paper provides a measurement-based characterization of cookie-stuffing fraud in online affiliate marketing. We use a custom-built Chrome extension, AffTracker, to identify affiliate cookies and use it to gather data from hundreds of thousands of crawled domains which we expect to be targeted by fraudulent affiliates. Overall, despite some notable historical precedents, we found cookie-stuffing fraud to be relatively scarce in our data set. Based on what fraud we detected, though, we identify which categories of merchants are most targeted and which third-party affiliate networks are most implicated in stuffing scams. We find that large affiliate networks are targeted significantly more than merchant-run affiliate programs. However, scammers use a wider range of evasive techniques to target merchant-run affiliate programs to mitigate the risk of detection suggesting that in-house affiliate programs enjoy stricter policing.

Categories and Subject Descriptors

K.4 [**Computers and Society**]: Electronic Commerce—*Security*

Keywords

Measurement; Online Advertising; Affiliate Marketing; Security

1. INTRODUCTION

Affiliate marketing is a popular form of pay-per-action or pay-per-sale advertising whereby independent marketers are paid a commission on "converting traffic" (e.g., clicks that culminate in a sale). Heralded as the "the holy grail" of online advertising a decade ago [17], affiliate marketing has become prevalent across the Web, complementing more traditional forms of display advertising.

Often described as a "low-risk" proposition for merchants who pay out only upon successful completion of sales, affiliate marketing attracts significant investment from almost every major online retailer, some of whom also invest in multiple third-party affiliate advertising programs. Similarly, it is an attractive proposition for independent marketers as they can create online content (e.g., book reviews) that can be monetized simultaneously as a means to attract likely converting traffic and to host contextual advertising. Of the two approaches, affiliate marketing is frequently the more profitable option with earnings typically between 4 and 10% of sales revenue [2, 18].

Like almost all economic activity on the Web, affiliate marketing also attracts the attention of fraudsters looking to make easy cash. Affiliate fraud garnered widespread media attention in 2013 with the indictment of Shawn Hogan, an EBay affiliate indicted for wire fraud of $28M through the use of a technique called *cookie-stuffing* [8] whereby the Web cookies used to determine the likely source of user traffic are overwritten without the user's knowledge. There have been multiple similar legal disputes over affiliate marketing since then [6]. Besides media attention, affiliate marketing has also been a subject of academic research to understand the incentives in the ecosystem and the extent of affiliate fraud [7, 16]. In this paper, we characterize popular cookie-stuffing techniques. From crawling likely sources of cookie-stuffing, we find that large affiliate networks such as CJ Affiliate (formerly Commission Junction) and Rakuten LinkShare (recently renamed to Rakuten Affiliate Network) are implicated in cookie-stuffing orders of magnitude more than affiliate programs run by merchants themselves, such as the Amazon Associates Program. Lower attempted fraud coupled with the much higher use of evasive cookie-stuffing techniques against in-house affiliate programs suggests that such programs enjoy stricter policing, thereby making them difficult targets of fraud.

We also find that retailers in the *Apparel*, *Department Stores*, and *Travel and Hotels* sectors of e-commerce are disproportionately targeted by affiliate fraud, usually through domains typosquatted on the merchant's trademarks. Finally, we evaluate data from a two-month *in situ* user study with 70+ users and find that affiliate marketing is dominated by a small number of affiliates while cookie-stuffing fraud is rarely encountered. Overall, our targeted crawl and user study both suggest that the problem, while real, appears to be less prevalent than suggested by previous reports.

2. BACKGROUND

Online merchants benefit from affiliate marketing through customized and targeted advertising for their products. For example, when an affiliate reviews a bicycle on a blog dedicated to biking, the bicycle merchant can receive sales from the readers of the blog without the merchant having to produce an advertising creative or directly advertise to the blog subscribers. Instead, the merchant pays a commission to the affiliate for each such sale made.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
IMC'15, October 28–30, 2015, Tokyo, Japan.
© 2015 ACM. ISBN 978-1-4503-3848-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2815675.2815720.

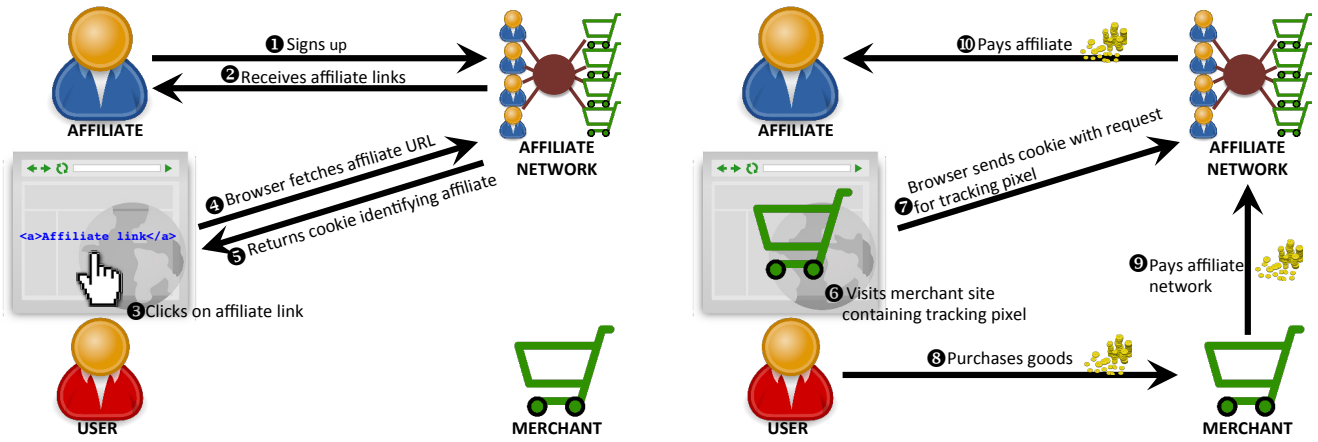

Figure 1: Different actors and revenue flow in the affiliate marketing ecosystem. The left half of the figure depicts a potential customer receiving an affiliate cookie, while the right half shows the use of the affiliate cookie to determine payout upon a successful transaction.

To recruit *affiliates* for advertising goods and services, an online merchant can either run its own *affiliate program* or join one run by a larger *affiliate network*. While some merchants like Amazon and HostGator run their own affiliate programs, most online retailers (particularly those whose expertise is on the brick-and-mortar side of the business) market through large affiliate networks such as CJ Affiliate and Rakuten LinkShare. Figure 1 provides an overview of the affiliate marketing ecosystem where a merchant is part of a large affiliate network that acts as a link between *affiliates*, who are typically content *publishers*, and *merchants*.

Affiliates who sign up for an affiliate network can choose to market for one or more merchants who are members of the network. Affiliate networks generally assign unique identifiers to all affiliates (*affiliate IDs*) and merchants (*merchant IDs*). Upon signup, affiliates receive special links from the affiliate network that encode the identifiers for the affiliate and the merchant for whom the affiliate is advertising. An affiliate includes these *affiliate links* in published content (e.g., a product review site) such that, when a potential buyer visits an affiliate's Web site and clicks on the link, it redirects the visitor to the merchant site via the affiliate program. The affiliate link `GET` request to the affiliate program returns an HTTP cookie (i.e., an *affiliate cookie*) that associates the user's visit with the corresponding affiliate. These cookies uniquely identify the referring affiliate for up to a month after the initial visit. If the user visits the merchant site during this period and completes a transaction, the affiliate network can identify the referral using the affiliate program's tracking pixel on the merchant's site. The referring affiliate usually earns between 4 and 10% on a completed transaction as a commission from the affiliate network who, in turn, is paid by the merchant for sourcing the sale.[1] In-house programs work similarly, with the network replaced by infrastructure maintained by the merchant itself.

Affiliate cookies remain in a user's browser until they expire, are overwritten by a different affiliate's cookie, or the user deletes them manually. If a user clicks on links for the same merchant from multiple affiliates in the same affiliate program, the cookie is overwritten and only the last affiliate to refer the user earns a commission.

These behaviors — that the presence of a cookie determines payout and that the most recent cookie "wins" — are at the core of the *cookie-stuffing* technique that allows fraudulent affiliates to obtain illicit commissions. In Figure 1, instead of using the affiliate URL as a clickable link, a fraudulent affiliate may cause the browser to directly fetch her affiliate URL on a page controlled by her without any explicit clicks from the user, thereby tricking the affiliate program into returning a cookie that then identifies the fraudulent affiliate as the referrer for the user's transactions. As a result, not only does an affiliate program pay a non-advertising affiliate, but the fraudulent cookie overwrites any existing affiliate cookie that may have already been present, thereby potentially stealing the commission from a legitimate affiliate. Furthermore, cookie-stuffing fraud is typically completely opaque to an end user and goes against the advertising guidelines issued by the Federal Trade Commission for marketers, which require declaration of any financial relationship with advertisers [9]. As a result, most affiliate programs explicitly forbid cookie-stuffing. For instance, the HostGator affiliate program states that "sales made through cookie stuffing methods will be considered invalid" [10].

In prior work, Moore et al. found several typosquatted domains that belong to fraudulent affiliate marketers [13]. Kapravelos et al. also found popular browser extensions that were cookie-stuffing major affiliate networks like the Amazon Associates Program [11]. Snyder et al. studied the extent to which users encountered affiliate fraud from the HTTP request logs for a public university [16]. In another study, Edelman et al. explored the incentives of different players in the affiliate marketing ecosystem, and also used crawling to identify affiliate programs defrauded through adware, typosquatting, and search engine optimization (SEO) [7]. Our work furthers this line of work by characterizing cookie-stuffing techniques and the range of targeted networks and retailers. In addition, we perform a user study to characterize the prevalence of affiliate marketing and cookie-stuffing fraud.

3. METHODOLOGY

In this section we describe how we measure cookie-stuffing fraud against six large affiliate programs: CJ Affiliate, Rakuten LinkShare, ShareASale, ClickBank, Amazon Associates Program, and HostGator Affiliate Program. While Amazon and HostGator run their own affiliate programs, the remaining four are consistently top-rated affiliate networks [15], which include well known merchants such as Nordstrom, Lego Brand, GoDaddy, etc. First, we study the structures of affiliate URLs and cookies used by these programs

[1]This is a general description. The details of the commission, the allowed duration for conversion and the implementation details (including the affiliate URL and cookie structures) can vary considerably among affiliate programs.

so that we can identify the affiliate network, the targeted merchant, and the affiliate's ID. We then use a custom-built browser extension to identify affiliate cookies received while browsing, and use it for the large scale crawling and the user study.

3.1 Identifying Affiliate URLs and Cookies

Broadly, we identified affiliate URLs and cookies either by signing up for these programs ourselves, or by finding this information online. Table 1 shows how we parse out affiliate and merchant IDs from some example affiliate URLs and cookies. For CJ Affiliate, we only show how we identify the publisher ID because we are unable to identify the corresponding affiliate ID. Every CJ affiliate can have multiple publisher IDs, one for each site used for publishing affiliate marketing creatives. However, every publisher ID is uniquely associated with a single affiliate. As a result, we use the terms publisher ID and affiliate ID interchangeably when discussing CJ Affiliate in the following sections.

Finally, the merchant is easy to identify because an affiliate URL eventually redirects to the merchant domain.

3.2 User Study

In our user study, we examine how often users click on affiliate links while browsing the Web, and identify affiliate cookies using a custom-built browser extension for Google Chrome called AffTracker.[2]

AffTracker gathers information about every single affiliate cookie it observes in the `Set-Cookie` HTTP response headers while a user is browsing. Upon detection of an affiliate cookie, AffTracker parses out the affiliate and merchant identifiers and the rendering information, including size and visibility, for the DOM element that initiated the affiliate URL request. AffTracker also records the redirect chain for the requests that result in affiliate cookies. Besides notifying the user about the cookie, AffTracker also submits this information to our server which stores it in a Postgres database.

By advertising to friends and colleagues, we obtained browsing data from 74 installations between March 1, 2015 – May 2, 2015. Using a locally generated unique ID, we can attribute affiliate cookies to specific users without collecting any personally identifiable information (PII).

While we can identify final attributes of DOM elements that cause a browser to fetch the affiliate links, we cannot automatically determine how such DOM elements are generated. Upon manual inspection we came across several affiliates who use JavaScript or Flash to dynamically generate hidden images and iframes that then request affiliate URLs. However, we are unable to quantify this phenomenon. Also, our user study does not have a completely random sample of users, and is likely biased towards savvy computer users. We discuss the results of our user study in Section 4.3.

3.3 Crawling

To characterize cookie-stuffing at scale, we visited over 475K domains to search for stuffed cookies. As described in Section 2, a user should only receive an affiliate cookie upon clicking on an affiliate URL. While crawling we do not click on any links and therefore every affiliate cookie we receive is deemed fraudulent. Since it is infeasible to crawl every Web page, we narrowed our visits to four different sets of URLs where we expected to come across affiliate fraud.

For every crawl we used a slightly modified version of our publicly available extension, AffTracker, which automatically grabs a new URL from a queue on Redis, a persistent key-value store. Upon completion of a visit, the extension submits results to our server and purges the crawler browser of all history, cookies, and local storage. We purge the browser because we found affiliates who save state in browsers to rate-limit their cookie-stuffing to evade detection by affiliate programs. For example, an affiliate, *jon007*, who controls `bestwordpressthemes.com`, sets a custom cookie called `bwt` which is valid for a month. As long as this cookie remains valid in a browser, `bestwordpressthemes.com` does not request HostGator affiliate cookies for the user. Also, inspired by Shawn Hogan who, according to EBay, rate-limited his cookie-stuffing by only requesting an affiliate cookie once per IP [4], we use 300 proxies to mitigate IP based detection by fraudulent affiliates.

We crawled the following four sets of domains to gather affiliate cookies. Except Alexa top domains set, the remaining three sets are purposely biased towards domains where we expect to find higher concentration of cookie-stuffing.

Alexa Top Domains. We crawled the Alexa [1] top 100K domains as of April 16, 2015 to find popular domains stuffing cookies.

Reverse Cookie Lookups. After identifying the cookie names used by different affiliate programs, we performed reverse lookups on the cookie-search interface [5] on `digitalpoint.com`, a webmaster community that indexes all of the cookies its crawler encounters. Overall, we gathered 9.5K domains that the Digital Point crawler observed performing cookie-stuffing over the last 2 years.

Reverse Affiliate ID Lookup. Using the cookie-stuffing affiliate IDs discovered from our Digital Point domain crawl, we queried an aggregator site, `sameid.net`, that indexes domains by Amazon and ClickBank affiliate IDs seen on a domain. By iteratively crawling domains queried from the newly discovered cookie-stuffing affiliate IDs, we visited a total of 74.5K domains.

Typosquatted Domains. In our first three sets we observed that much of the cookie-stuffing fraud was on domains typosquatted on merchant domains names. Thus, we crawled all typosquatted `.com` domains for over 7K domains belonging to major e-retailers. Similar to Edelman et al. [7], we interpret the use of typosquatting to redirect users to merchant sites without any explicit clicks as cookie-stuffing, and therefore, fraudulent. Typically users visiting typosquatted domains intend to visit the merchant site rather than the typosquatted domain itself. Thus, typosquatting a domain does not bring new customers to merchant sites via direct affiliate marketing of products. Therefore, we recognize cookie-stuffing via typosquatted domains as fraudulent.

We acquired the set of domains belonging to e-retailers from a public API offered by Rakuten Popshops.[3] The downloaded data includes merchant lists for Commission Junction, ShareASale, and Rakuten LinkShare affiliate networks. By calculating the Levenshtein distance [12] for merchant domains against all `.com` domains in a zone file from April 19, 2015, we found over 300K typosquatted domains with an edit distance of one. We visited this set of domains as well.

We gather the same features about affiliate cookies through crawling as from the user study because we use the same browser extension, AffTracker, in both cases. Furthermore, we only visit top-level pages of domains and therefore miss any cookie-stuffing in domain sub-pages. Also, Google Chrome disables popups by default, a feature we left unchanged because it emulates a user's browser more faithfully. However, this behavior likely caused our

[2] affiliatetracker.ucsd.edu

[3] `https://www.popshops.com`

Affiliate Program	URL	Cookie
Amazon Associates Program	`http://www.amazon.com/dp/tag=<aff>&...`	`UserPref=.*`
CJ Affiliate	`http://www.anrdoezrs.net/click-<pub>-...`	`LCLK=.*`
ClickBank	`http://<aff>.<merchant>.hop.clickbank.net/`	`q=.*`
HostGator	`http://secure.hostgator.com/~affiliat/...`	`GatorAffiliate=.*.<aff>`
Rakuten LinkShare	`http://click.linksynergy.com/fs-bin/click?...`	`lsclick_mid<merchant>=".*|<aff>-.*"`
ShareASale	`http://www.shareasale.com/r.cfm?...`	`MERCHANT<merchant>=<aff>`

Table 1: Examples of affiliate URLs and cookies for different affiliate programs.

crawler to miss any affiliate fraud where a fraudster opens a popup to load an affiliate URL.

Finally, while we can detect fraudulent affiliates stuffing cookies, we are unable to discern whether an affiliate network has already identified the fraudster. We have seen examples of ClickBank and LinkShare affiliate sites where affiliate links show an error message about affiliates having been banned, but some networks do not break banned affiliate links to prevent bad end-user experience for their URLs.

4. RESULTS

In this section we first analyze the affiliate networks and merchants most impacted by cookie-stuffing fraud. Next, we survey various cookie-stuffing techniques used by affiliates, and then present the results from our user study on the prevalence of affiliate marketing and cookie-stuffing fraud.

4.1 Networks Affected by Cookie-Stuffing

Using data collected from our crawls, we identify the affiliate networks most targeted by cookie-stuffing. Overall, we received 12,033 affiliate cookies from 11.7K domains. Table 2 summarizes our results.

We found that CJ Affiliate and Rakuten LinkShare are the most targeted programs, comprising 85% of all fraudulent cookies we observed. We identified affiliate IDs for all but 1.6% of these cookies. Every fraudulent CJ affiliate stuffed almost 50 cookies, while every LinkShare affiliate stuffed 41 cookies. However, fraudulent affiliates in Amazon and HostGator affiliate programs only stuffed 2.5 cookies per affiliate on average, the fewest of all affiliate programs in our study, suggesting that fraudulent affiliates target networks much more than they target in-house affiliate programs.

Generally, affiliate networks present greater cookie-stuffing opportunity to fraudulent affiliates because a single affiliate can simultaneously defraud multiple merchant members of the network. Our data from the Popshops API (Section 3.3) contains almost 2.4K merchants in CJ Affiliate, and 1.3K merchants in Rakuten LinkShare, and as Table 2 shows, each fraudulent affiliate targeted more than three merchants in LinkShare on average. We received 10 and 15 cookies on average for every targeted merchant in CJ Affiliate and Rakuten LinkShare, respectively.

Using the Popshops data as ground truth, we classified the defrauded merchants in all of the major networks (Figure 2) in our study except ClickBank and 420 CJ Affiliate cookies. We found that on the whole, *Apparel and Accessories* e-retailers are targeted the most across all three affiliate networks while the second most impacted group of merchants are *Department Stores*, again abused in all three networks but to a greater extent in LinkShare. *Travel and Hotel* sites were the third-most defrauded group. These three sectors have a large number of merchants and we found almost 11 stuffed cookies on average for every targeted merchant. On the other hand, the *Tools and Hardware* category only contains four impacted merchants but we observed almost 45 cookies for each of them on average, the highest of any category. Home Depot, a CJ Affiliate member, was the most impacted merchant in this category with 163 stuffed cookies.

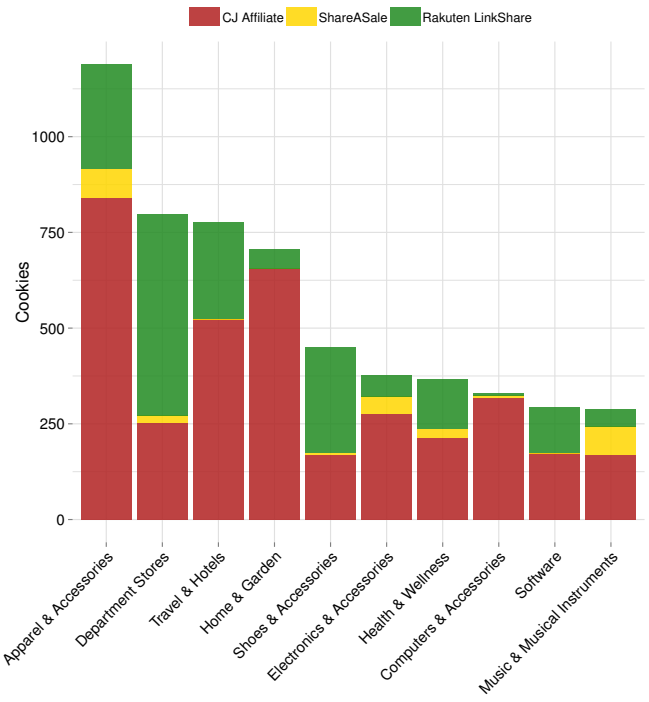

Figure 2: Stuffed cookie distribution for top 10 categories of impacted merchants.

We also found 107 merchants who were defrauded across two or more networks. *Chemistry.com*, a member of CJ Affiliate and LinkShare, was the most targeted merchant participating in more than one affiliate program.

4.2 Prevalence of Cookie-Stuffing Techniques

We use the collected data to characterize the use of various cookie-stuffing techniques against each of the affiliate programs under consideration. As described in Section 3, whenever the browser requests an affiliate URL, our extension finds the DOM element that caused the fetch. Visiting a fraudulent affiliate's Web page can cause the user's browser to fetch an affiliate URL without any explicit click from the user by setting it as the `src` attribute of images, iframes, or script tags. All of these DOM elements can request third-party content on a Web page. Our extension is able to detect all affiliate cookies by observing the `Set-Cookie` HTTP headers, and distinguish between these techniques automatically.

Affiliate Program	Cookies	Domains	Merchants	Affiliates	Techniques			Avg. Redirects
					Images	*Iframes*	*Redirecting*	
Amazon Associates Program	170 (*1.41%*)	122	1	70	28.8%	34.1%	37.0%	1.64
CJ Affiliate	7344 (*61.0%*)	7253	725	146	0.29%	2.46%	97.2%	0.94
ClickBank	1146 (*9.52%*)	1001	606	403	34.4%	13.5%	52.0%	0.68
HostGator	71 (*0.59%*)	63	1	29	43.7%	19.7%	35.2%	0.87
Rakuten LinkShare	2895 (*24.1%*)	2861	188	57	0.28%	0.41%	99.3%	1.01
ShareASale	407 (*3.38%*)	404	66	34	0.25%	0.0%	99.8%	0.74

Table 2: Affiliate Programs affected by cookie-stuffing.

Table 2 shows the percentage of cookies corresponding to each of these techniques for every affiliate program.

At a high level, we observe that fraudulent affiliates use a much larger variety of techniques uniformly for affiliate programs run by the merchants themselves compared with larger affiliate networks, which are targeted primary via redirects to merchant sites without user clicks. Table 2 also shows the average number of intermediate domains requested after the initial page visit but before the affiliate URL, i.e., a value of zero means that an affiliate URL was directly requested from the crawled page. Intermediate referrers can be used to obfuscate the original source of a fraudulent affiliate URL request because an affiliate program only sees the final referrer when determining the legitimacy of such a request. Affiliates defrauding Amazon Associates Program use more intermediate domains on average. Since the cost of getting intermediate domains is higher, it is likely more expensive for fraudulent affiliates to defraud Amazon Associates Program presumably due to stricter policing by Amazon.

Redirecting. Fraudulent affiliates redirect users to affiliate URLs without any clicks by either using 301 or 302 HTTP response status codes, or using Flash or JavaScript to redirect the browser to the affiliate URL. In each of these cases, the original page we crawled only resulted in one stuffed cookie per visit. Such redirects delivered over 91% of all stuffed cookies, most of which resulted from typosquatted domains. In fact, we received 84% of all affiliate cookies from 10.1K typosquatted domains.

Of the 10.1K cookies from typosquatted domains, 93% (9.4K cookies) are from domains typosquatting on merchant domain names while 1.8% resulted from typosquatting on subdomains. For example, `liinensource.com` redirects to Rakuten LinkShare merchant `linensource.blair.com`. We manually inspected 30 of the remaining 520 typosquatted domains that resulted in affiliate cookies and found that these domains can be broadly classified into three types. One-third of these domains are contextually related to the final landing page. For example, `0rganize.com` redirects to CJ merchant `shopgetorganized.com`, and `bhealthypets.com` and `healthypts.com` redirect to CJ merchant `entirelypets.com`. Another third appear to be expired CJ offers, and thus did not redirect to any merchant site. The remaining one-third cookies result from typosquatted domains selling traffic to traffic distributors like `pureleads.com`, `7search.com`, and `blendernetworks.com` that eventually redirect through an affiliate URL. We revisit traffic distributors in our discussion on referrer obfuscation in this section.

Iframes. Iframes are often used to render third-party content on a Web page, which is otherwise forbidden by the Same-Origin policy implemented by browsers. Most major affiliate programs disallow the use of iframes, as it is a commonly used mechanism to facilitate cookie-stuffing. For example, HostGator explicitly prohibits iframes: "iframes may not be used unless given express permission by HostGator, sales made through hidden iframes or Cookie-stuffing methods will be considered invalid" [10]. Similarly, Amazon Associates Program prohibits framing any Amazon link on a page [3].

We received 420 cookies from content rendered in iframes on third-party sites. Generally, a server can prevent a Web site from framing its content on a page by using the `X-Frame-Options` HTTP header with its value set to `SAMEORIGIN` to only allow rendering content on a page with the same origin as the frame, or `DENY` to completely disallow framing content on any Web site [14]. We determined that Google Chrome and Firefox browsers honor the `X-Frame-Options` header and do not render the iframe content, but both browsers save the cookies nonetheless. Thus, iframe based stuffing is effective despite the use of `X-Frame-Options` header.

We found that 17% of the cookies received from iframes set `X-Frame-Options` to either `SAMEORIGIN` or `DENY`, including every Amazon Associates Program cookie. Table 2 shows that iframes still accounted for over a third of the stuffed Amazon cookies. Unlike Amazon, only 2% of CJ cookies and 50% of LinkShare cookies were accompanied by a restrictive `X-Frame-Options` header.

We also used the style and size information gathered in our crawl to determine how, if at all, a user would have seen the corresponding iframe on the crawled page. We gathered this information for 46% of the iframes. Of the 191 iframes, 64% explicitly set the height or width to either 0 or 1px; 49 (25%) iframes have `visibility:hidden` or `display:none` set, thereby making the iframe invisible to an end user. Additionally, seven iframes use CSS classes to hide the iframe DOM element. Of these, three have the same affiliate ID *kunkinkun* and the CSS class `rkt` specifies `left:-9000px`, which positions the iframe outside the viewport, and therefore invisible to the user. The same affiliate also defrauds Amazon Associates Program using the same technique and ID *shoppertoday-20*. We also found two examples where iframes were made invisible by setting the `visibility` CSS property on their parent DOM elements. The 49 remaining iframes were not hidden, and most of them correspond to ClickBank.

Images. Images can also be used to fetch third-party content on a Web page. 504 cookies in our data set were requested as images. Of these we have recorded rendering information for 91% cookies. Unlike our iframe data, we found that every single DOM element either had width or height set to 0 or 1px, or style set to `display:none`, effectively hiding the image from the end user.

We also found six cookies were requested by hidden `img` elements embedded within `iframe` elements. For example, `best blackhatforum.eu`, a domain with Alexa rank 47,520, stuffs cookies for three different LinkShare merchants (`UDemy.com`, `microsoftstore.com`, `origin.com`), one CJ merchant (`GoDa`

ddy.com), and Amazon. All of these affiliate URLs are requested as hidden images of height and width zero pixels *inside* iframes with src set to lievequinp.com, which is then observed by affiliate programs as the referrer. As a result, the affiliate programs do not observe the actual cookie-stuffing bestblackhatforum.eu domain in the request for affiliate URLs thereby making detection of cookie-stuffing difficult. As shown in Table 3, fraudulent affiliates use iframes to defraud Amazon and HostGator more often than affiliate networks, suggesting greater difficulty in evading detection by these in-house programs.

Scripts. Even though script tags can be used to fetch third-party content from affiliate URLs by setting the src attribute, we only found two such stuffed cookies. However, upon manual inspection of several cookie-stuffing domains we found that scripts are often used for dynamic generation of hidden images and iframes that then request the affiliate URLs.

Referrer Obfuscation. Next, we analyze the extent to which fraudulent affiliates hide the actual cookie-stuffing domain behind innocuous domains. Referrer obfuscation is used to make cookie-stuffing via any technique, such as images, opaque to the affiliate programs.

Of the 12K cookies we gathered in our crawl, 84% were fetched via at least one intermediate domain. In fact, 77% of all cookies were fetched via a single redirect, 4.5% via two redirects, and another 2% via three or more redirects. Only the last redirect is seen by the affiliate program in the HTTP Referer header.

We analyzed the intermediate domains, and found that a significant portion of the redirects go through a very small variety of domains. The most common intermediate domains we observed are cheap-universe.us, flexlinks.com, dpdnav.com, pgpartner.com, 7search.com and pricegrabber.com. Of these, flexlinks.com belongs to an affiliate program called FlexOffers, while the other domains are likely traffic distributors buying traffic and then monetizing via affiliate fraud. Over 25% of the cookies in our data contain a redirect through at least one of these traffic distributors. In fact, 36% of all CJ cookies contain at least one of these domains.

4.3 Prevalence of Affiliate Marketing

As described in Section 3, we gathered affiliate cookie data from 74 users over a two month period to study how often users click on affiliate links and how often they receive stuffed cookies. Only 12 users received any affiliate cookie, encountering a total of 61 cookies for 23 distinct merchants. Over a third of these cookies resulted from affiliate links on dealnews.com and slickdeals.net. Thus, while almost 84% of the users did not receive any cookie at all, 12 users received an average of 5 cookies per user. Table 3 shows the high level results.

Amazon Associates Program was the most popular affiliate program in our study, accounting for almost 51% of the cookies. As shown in Table 3, CJ Affiliate was the second most popular affiliate program, followed by Rakuten LinkShare. Our users did not receive any affiliate cookies from the ClickBank or HostGator affiliate programs. This distribution is different from the networks targeted by cookie-stuffers, where CJ Affiliate is targeted significantly more than Amazon.

Notably, none of these affiliate cookies were rendered within hidden DOM elements. We manually inspected all of them and verified that none of the source affiliate Web sites are stuffing cookies. To rule out the possibility that users were protected by ad-blocking extensions that often disallow third-party cookies, we gathered the lists of extensions on their browsers and found that only four users use any such extension. From our user study we found that users rarely encounter cookie-stuffing fraud, and affiliate marketing is dominated by a small number of affiliates. While the set of participants of our user study is likely biased towards technologically savvy users, our results are consistent with Snyder et al. [16] who found that cookie-stuffing was a very small percentage of the HTTP traffic of a large public university.

Affiliate Network	Cookies	Users	Merchants	Affiliates
Amazon Associates Program	31	9	1	16
CJ Affiliate	18	5	2	7
ClickBank	0	0	0	0
HostGator	0	0	0	0
Rakuten LinkShare	9	3	6	5
ShareASale	3	2	3	2

Table 3: Affiliate Programs that AffTracker users received cookies for.

5. CONCLUSION

In this study we characterized the abuse of affiliate marketing for monetary gains through the use of techniques broadly classified as *cookie-stuffing*. Overall, even through targeted crawling of domains with higher likelihood of encountering affiliate fraud, we observed only a limited amount cookie-stuffing in our study.

Most merchants interested in affiliate marketing have a choice to either run their own affiliate programs, or join a large affiliate network with thousands of other merchants and affiliates. Since affiliate networks have a larger number of merchants, they provide greater opportunity to fraudulent affiliates to simultaneously target multiple merchants. In fact, of the affiliate fraud we did identify, we observed that large affiliate networks are targeted disproportionately more compared to the merchant-run affiliate programs who are targeted to a smaller extent, but through more sophisticated and costly cookie-stuffing techniques such as the use of intermediate domains to obfuscate the cookie-stuffing domains. These results suggest that the in-house affiliate programs are better placed to police their affiliate programs due to greater visibility into the affiliate activities and the revenue flow, and possibly shorter turnaround time to take action against a fraudulent affiliate upon detection. However, running an in-house affiliate program requires expertise and cost investment not necessary for outsourcing the logistics of running an affiliate program to an existing network.

Finally, we conducted a user study to determine the extent of affiliate marketing encountered by users during their daily Web browsing. We found that a small number of affiliates dominate affiliate marketing, and, as with the crawling results, our users rarely encountered cookie-stuffing fraud.

Acknowledgements

We thank our shepherd Paul Barford for his help with improving the paper, and the anonymous reviewers for their valuable feedback. We are also grateful to all the users of the AffTracker extension without whom our user study would not have been possible. We are in debt to Cindy Moore for her help in setting up the affiliatetracker.ucsd.edu domain, and to Tristan Halvorson for managing the COM zone file data. This work was supported by National Science Foundation grant NSF-1237264 and by generous research, operational and/or in-kind support from Google, Microsoft, Yahoo, and the UCSD Center for Networked Systems (CNS).

6. REFERENCES

[1] Alexa. Does Alexa have a list of its top-ranked websites? https://support.alexa.com/hc/en-us/articles/200449834-Does-Alexa-have-a-list-of-its-top-ranked-websites-.

[2] Amazon. Associates Program Advertising Fee Schedule. https://affiliate-program.amazon.com/gp/associates/help/operating/advertisingfees.

[3] Amazon. Associates Program Participation Requirements. https://affiliate-program.amazon.com/gp/associates/help/operating/participation/.

[4] Civil Cover Sheet. http://www.benedelman.org/affiliate-litigation/ebay-digitalpoint-hogan-kessler-thunderwood-dunning-complaint.pdf#page=8, 2008.

[5] Digital Point. Cookie Search. https://tools.digitalpoint.com/cookie-search.

[6] B. Edelman. Affiliate fraud litigation index. http://www.benedelman.org/affiliate-litigation/, 2015.

[7] B. Edelman and W. Brandi. Risk, Information, and Incentives in Online Affiliate Marketing. In *Journal of Marketing Research*, 2014.

[8] J. Edwards. How eBay Worked With The FBI To Put Its Top Affiliate Marketers In Prison. http://finance.yahoo.com/news/ebay-worked-fbi-put-top-120500693.html, 2013.

[9] Federal Trade Commission. Guides Concerning the Use of Endorsements and Testimonials in Advertising. https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-publishes-final-guides-governing-endorsements-testimonials/091005revisedendorsementguides.pdf.

[10] HostGator. Affiliate Terms of Service. http://www.hostgator.com/tos/affiliate-tos.

[11] A. Kapravelos, C. Grier, N. Chachra, C. Kruegel, G. Vigna, and V. Paxson. Hulk: Eliciting Malicious Behavior in Browser Extensions. In *Proceedings of the 23rd USENIX Security Symposium*, 2014.

[12] V. I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. In *Soviet Physics Doklady*, 1966.

[13] T. Moore and B. Edelman. Measuring the perpetrators and funders of typosquatting. In *Financial Cryptography and Data Security*, 2010.

[14] Mozilla Developer Network. The X-Frame-Options response header. https://developer.mozilla.org/en-US/docs/Web/HTTP/X-Frame-Options.

[15] mThink. The Top 20 Affiliate Networks 2014. http://mthink.com/the-top-20-affiliate-networks-2014.

[16] P. Snyder and C. Kanich. No Please, After You: Detecting Fraud in Affiliate Marketing Networks. In *Proceedings of the Workshop on the Economics of Information Security (WEIS)*, 2015.

[17] The Economist. Pay Per Sale. http://www.economist.com/node/4462811.

[18] The New York Times. Surviving the Dark Side of Affiliate Marketing. http://www.nytimes.com/2013/12/05/business/smallbusiness/surviving-the-dark-side-of-affiliate-marketing.html.