Web Audit Report

Trent Chismar, Angela Castronuovo
Ryan Brockway, Josh Cominelli

Grammarly is a service which proofreads a user's work as they type it. It comes in many different forms, from browser plugins, to an add on for Microsoft Office. In this audit, we are focusing on the Grammarly website, grammarly.com. Using Google Chrome's browser development tools, we examined the HTML and Javascript on various parts of the website. We also used the tools to examine the cookies that Grammarly sets for the user. This information gave us more of an insight on how Grammarly interacts with users, as well as other websites

Cookies are widely used on websites for a number of reasons: for tracking, for authorization, and for identification of users. Every time the user clicks on a different Grammarly web page, the page sets between 35 and 100 cookies. Grammarly.com's numerous cookies mostly consist of tracking cookies for advertising purposes. These include cookies from Facebook, Google Analytics, Quora, Taboola, Adsymptotic, Hotjar, Linkedin, and Bing. On the Grammarly home page, a csrf-token cookie, which is active for a year, was found. It is interesting to note that the csrf_token is not an HTTPOnly cookie, so it is more vulnerable to XSS attacks [1]. The home page also consisted of various Facebook cookies that result in more personalized ads for the user as well as improving Facebook's advertising. Analyzing keyboard page [2] led to the discovery of cookies from Adsymptotic.com along with invisible tracking pixels. After researching the cookies, we found that their policy states that they "may share this aggregate data with its affiliates, agents and business partners. This information includes internet protocol (IP) addresses, browser type, internet service provider (ISP), referring/exit pages, operating system, date/time stamp, and clickstream data…" [3]. This is a little concerning that all of this user information is getting distributed to many different companies. Additionally, Grammarly sets one of Doubleclick.net's cookies, called DSID, which sets the specific user identity that is hashed and encrypted. Doubleclick.net is part of Google Analytics and its information is given to well-paying companies. This also is a little concerning because a company that receives so much of our personal data is selling it to the highest bidder [4] . Grammarly's cookies did not show any obvious vulnerabilities, as the user ID cookies and cookies having to do with payment was secure and was not long lasting.

Grammarly uses multiple different outside services that track visitors to the site. The most prominent of these is Hotjar. This company allows Grammarly to get a better understanding of how users are interacting with the site. Grammarly can do this by inserting Hotjar's scripts into the site where they record what pages the user visits, their keystrokes, and even mouse movements. [5] These recordings introduce another avenue that hackers can exploit to steal user's personal information. This personal data can include usernames and passwords, but for the premium customers, it also could include their credit card information. Some experts believe that if a website using Hotjar is actively recording keystrokes and other inputs without telling users, much less without getting their permission, this would be illegal, and in violation of the GDPR. [5] Other major sites that use Hotjar are as diverse as Walmart, Pearson, and Patagonia [6].

The second type of third party user analytics that Grammarly uses is tracking pixels. On the homepage multiple different companies have pixels, there is a Facebook pixel, one from bat.bing, a site used by Microsoft's Ad division[7], and one from Grammarly itself. When a user loads the page, this pixel can collect data such as what kind of browser the user has, or even their IP address. A company like Facebook or Microsoft can have pixels all over the web, and they can build up a good model of a single user's online behavior. Like Hotjar's recording in the previous paragraph, tracking pixels are a point of contention with privacy advocates because it allows companies to collect data on customers without them realizing.

Grammarly uses https, meaning that the traffic between the user and the website is encrypted. This is beneficial to the user because eavesdroppers are not able to gain access to anything the user sends to Grammarly, such as their password or payment information, because all of this information is encrypted.[8] Also, there is no mixed https/http content on Grammarly. If Grammarly did have mixed content, then a man-in-the-middle could potentially intercept the http content and modify it. For example, say an iframe loaded content from a http page. The attacker could then modify the http content and use it to gain sensitive information on the user, or they could install malware on the user's machine.[9] Grammarly does not load any http content, so the site is safe from attacks like this. Also, Grammarly severely limits the amount of content that they load from domains that are not hosted by Grammarly. The vast majority of images all come from the domain, https://static.grammarly.com/assets. This provides more security to the site, as Grammarly is mainly using content from their own domain, so the only way for this content to be compromised is if Grammarly itself is compromised. An example of images that are not from the Grammarly domain that are a part of the website are tracking pixels. Grammarly is secure because it uses only https content, and it limits the content that it loads from other domains.

One unique page on the Grammarly website is its demo. Here, users can see how the service works on a provided demo document. Despite its appearance, the webpage is not a word processor. Rather, upon typing most keystrokes (numbers, letters, spacebar, etc.), a popup is triggered for users to create an account. At the same time, a few requests are sent. There are two called "logv2" and one called "events". These requests are consistently made every time the user closes the signup prompt and types a key. This suggests the possibility of a side channel being present. Side channels are exploitable when a computer provides an observable pattern. The pattern itself can vary, including electrical emissions, power consumption, or in this case, response timing when communicating with other devices [10]. Since this prompt only occurs when a user is not signed in, it doesn't appear to pose a threat. But one can be signed in while using the demo document. In this case they are able to type and edit the document without any popups. However, moving the type cursor over any underlined errors that Grammarly has detected results in requests similar to before being sent. The "events" request is important in this case, because it contains the "csrf-token" cookie, which is meant to protect the user from CSRF attacks. If an attacker were able to intercept the "events" request as well as break the cryptography of this token via a side channel exploit such as a timing attack [11], they would have access to it, allowing cross-site request forgery. All of this assumes that the previously mentioned requests are indeed a side channel, and while it can't be decided for sure, is at least worth noting.

According to Grammarly's privacy policy, they "do not and will not sell your information". Grammarly collects two types of information from the user. One is "Personal Data", which is information that identifies the user, and the other is "Non-Personal Data", which is not connected to the user. In the privacy policy, Grammarly is ambiguous about the distinction between these two types of data, and they do not explicitly state what is considered Personal Data and what is considered Non-Personal Data. Grammarly will share Personal Data with third parties when Grammarly is in need of a service from the third party. Grammarly states that the "service providers may only access, process, or store Personal Data pursuant to our instructions and to perform their duties to us". Third parties are unable to buy user data from Grammarly, and can only gain access when Grammarly deems it is necessary, as described above. Grammarly will also disclose your Personal Data due to a merger, or if they are bought by a different company. In this case, "some or all of your Personal Data may be shared with or transferred to another entity". A third party could gain access to Grammarly's user data, but only if they buy out the company. Also, Grammarly states that users may exercise the rights given to them by the GDPR. User's are able to "request a Personal Data report", and "have [their] Personal Data corrected or deleted". User's can also ask Grammarly to stop using their Personal Data. Grammarly's privacy policy states that they only share User data when they deem it necessary, and they are compliant with the GDPR.[12]

Over the course of this audit, we were not able to find any security vulnerabilities. What we did find was evidence of Grammarly collecting lots of personal data, and not being explicit about how that data is shared with other companies.

Link to Presentation:
https://drive.google.com/file/d/1o7hBVu4l0YRTRGcJ96gNs7vE2uKLbMHO/view?usp=sharing

Sources

[1] owasp.org/www-community/HttpOnly
[2] https://www.grammarly.com/keyboard
[3] https://better.fyi/trackers/adsymptotic.com/
[4]
https://www.theguardian.com/technology/2012/apr/23/doubleclick-tracking-trackers-cookies-web-monitoring#:~:text=Doubleclick%20is%20a%20business%20owned,Ad%20delivery
[5] www.infosecurity-magazine.com/news/top-sites-expose-breaches-tracking/
[6] trends.builtwith.com/websitelist/Hotjar/$1m-Sales-Revenue
[7]answers.microsoft.com/en-us/bing/forum/all/does-batbing-track-your-browser-searches-and-sites/0a402f00-60c2-4d54-bd7d-81b67ccc7f13
[8]https://en.wikipedia.org/wiki/HTTPS
[9]https://developer.mozilla.org/en-US/docs/Web/Security/Mixed_content
[10] https://www.wired.com/story/what-is-side-channel-attack/
[11] http://gauss.ececs.uc.edu/Courses/c653/lectures/SideC/intro.pdf
[12]https://www.grammarly.com/privacy-policy