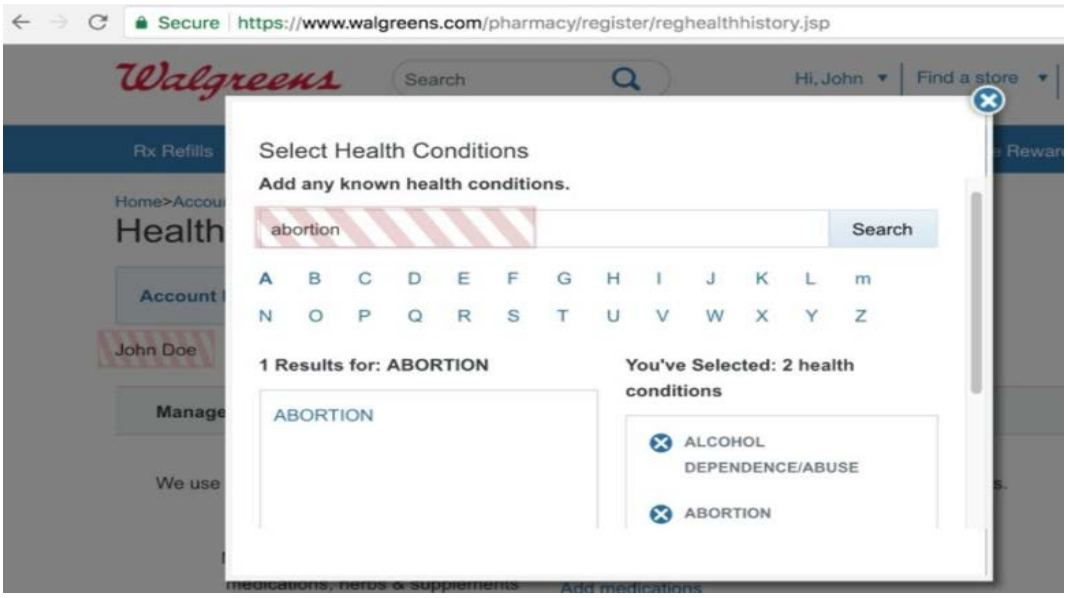
"Session-Replay Scripts"

December 3, 2017

ISSUE

Businesses of all sizes are continuously focusing on the analytics of their website and its performance. Many just rely on Google Analytics to help them understand a visitor's demographic, experience and behavior. However, many website analysts feel that this is simply not enough to understand the user experience and how the site visitor behaves and navigates through the site. To better assist them, a trending analytics technology is emerging – "Session-replay scripting". These scripts are used to gather data on how users are interacting specifically with their sites and records these movements. This allows the website developers and operators to determine areas of the site are serving no relevancy or simply not aligning to the organization's overall sales or marketing strategy.

The recording of the browsing sessions is far exceeding the expectations of the script authors such that they are sending a great deal of information back to the servers that is much more than just the specifics about the surfing session. When users are filling out forms, the entire session, including all of the form content, is captured. This is a problem as even if the user deletes the form data, the original content is still accessible via playback, similar to a video surveillance system.



IMPORTANCE

There are many service providers for session-replay scripting services, but some do not offer options to exclude certain information that would be deemed an invasion of privacy. The providers that do offer exclusion functionality are overall complicated and too labor-intensive to implement. Many website developers would simply add the scripting without programming the exclusions. Most of these exclusions involve diligently reviewing all pages within a site and "scrubbing" them to avoid areas where the site visitor would have to key in confidential information. This is especially challenging when sites are dynamically adding pages – this would involve digging deep into the coding and applying modifications on how the pages are created. Due to the nature of the programming time and the pressures to Go Live with sites, this is often overlooked.

There has been a great deal of research completed on the service providers and use of their scripting products. The products usually have a script that is placed in certain areas of the site or on all pages and then there is a playback dashboard. Statistics are gathered and displayed along with the recordings and content of the mouse and keystrokes involved. The research also included an analysis sample of some of the top providers and test pages were created in testing environments. The test results revealed that confidential information included passwords, medical information, credit card information and personal information such as social security numbers and driver's license information. Also, it was observed that none of these companies used technology to remove the sensitive data being captured – such as automated redaction technologies.

Redacted Field	FullStory	UserReplay	SessionCam	Hotjar	Yandex	Smartlook
Name	○	◐	◐	○	○	○
Email	○	◐	◐	○	○	○
Phone	○	◐	◐	○	○	○
Address	○	◐	◐	○†	○	○
SSN	○	◐	◐	○	○	○
DOB	○	◐	◐	○	○	○
Password	●	◐	●	●	●	◐
CC Number	●	◐*	◐	◐	○	●
CC CVC	●	◐	◐	○	○	●
CC Expiry	●	◐	◐	○	○	●

Summary of the automated redaction features for form inputs enabled by default from each company.
Filled circle: Data is excluded; **Half-filled circle:** equivalent length masking; **Empty circle:** Data is sent in the clear
* UserReplay sends the last 4 digits of the credit card field in plain text
† Hotjar masks the street address portion of the address field.

MY TAKEAWAYS

This will be a compromise in my opinion. The intent to use session-replay scripts to understand how a website is performing is a typical business process. The technology implemented however can have malicious outcomes from misuse of acquired information. This becomes an ethical question about responsible use of technology and addressing privacy concerns. Adware was originally intended to sell products and services on websites in a harmless fashion, but this technology was quickly exploited to allow hackers into these coding systems and it quickly became harmful, or systems became overwhelmed with allocating resources to accommodate the quickly spreading ads.

There are software products available to consumers to block these types of scripts, but they come at a price. Most of these must be purchased and they all can impact the functionality of the site in some fashion. The software will also act against sites that do not use this scripting technology. It seems odd that you have to purchase software so you can be protected from sites that you depend on – such as reordering a prescription from Walgreen's or purchasing a domain from GoDaddy. This is a topic that is emerging quickly with privacy advocates, but exploitation is looming. The session-replay providers must find a way to simplify the exclusion process and make their systems easier to use and administer so that businesses can learn about performance rather than passwords.

REFERENCES

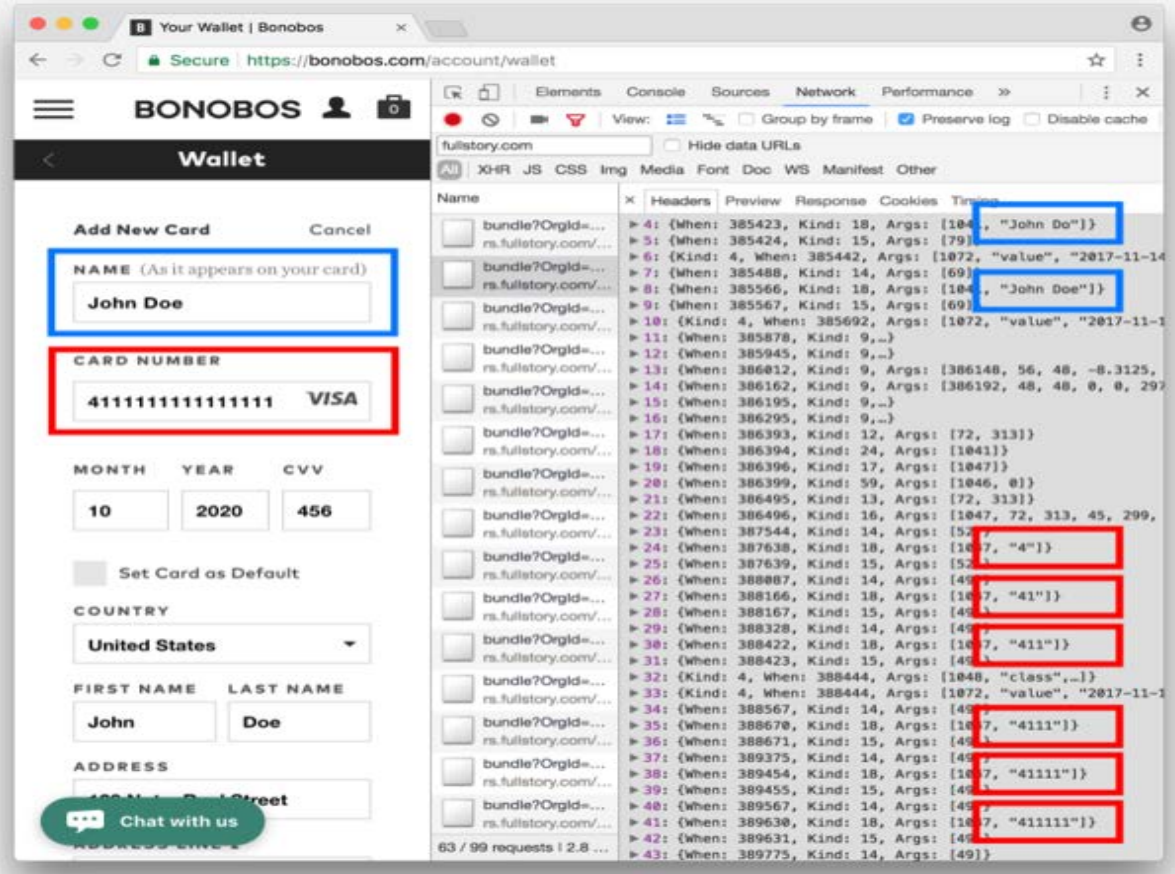
https://www.technewsworld.com/story/84973.html

https://freedom-to-tinker.com/2017/11/15/no-boundaries-exfiltration-of-personal-data-by-session-replay-scripts/

https://webtransparency.cs.princeton.edu/no_boundaries/session_replay_sites.html

https://gist.github.com/gunesacar/0c67b94ad415841cf3be6761714147ca

https://github.com/citp/OpenWPM



The account page of the clothing store Bonobos leaks full credit card details to FullStory. The screenshot of Chrome's network inspector shows the leaked data being sent letter-by-letter as it is typed. The user's full credit card number, expiration, CVV number, name, and billing address are leaked on this page. Email address and gift card numbers are among the other types of data leaked on Bonobos site.

Reference: "Freedom to Tinker" illustration – Nov 15, 2017