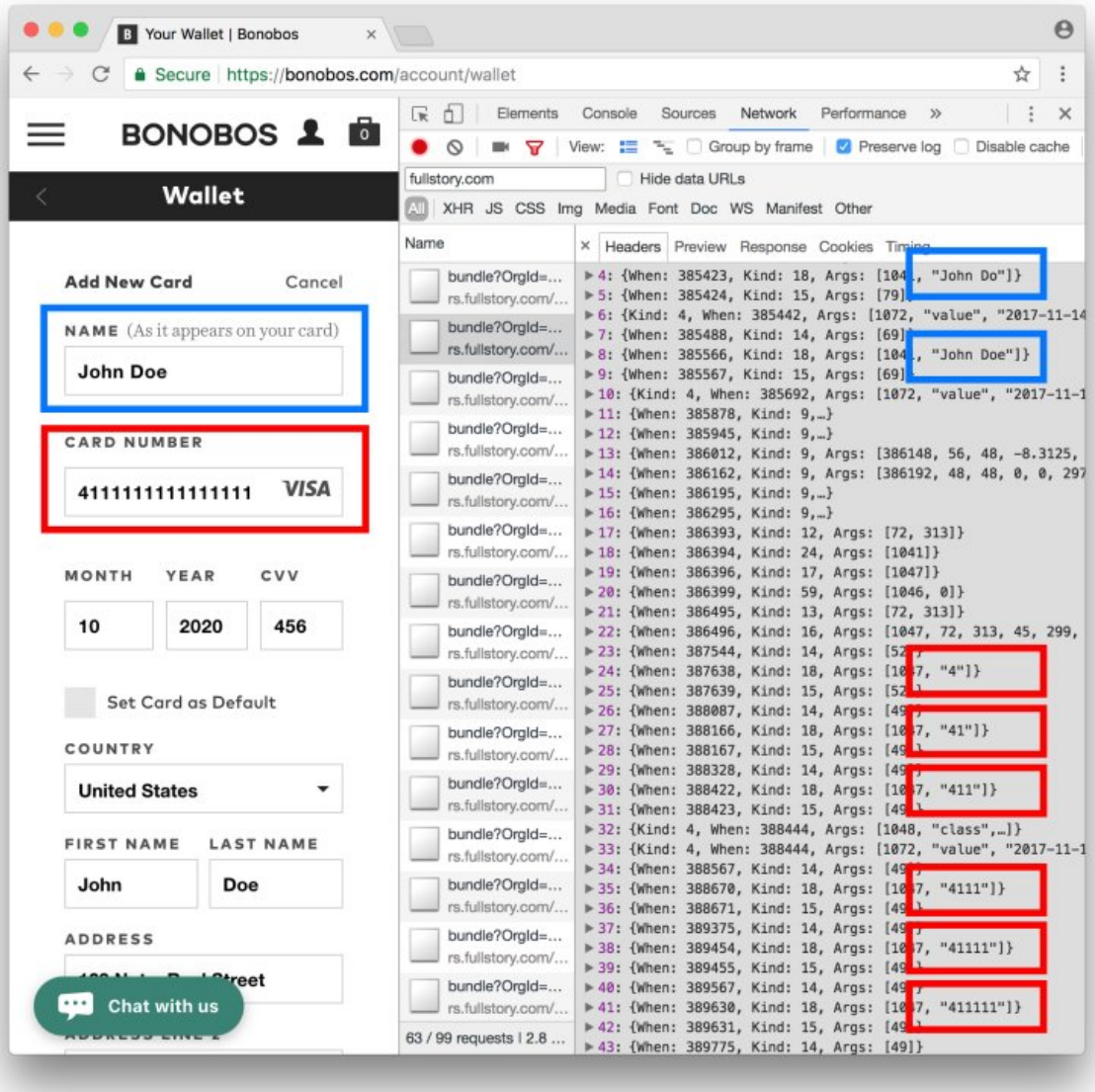
Princeton researchers spot website visits being watched

November 22 2017, by Nancy Owano



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. Credit: freedom-to-tinker.com

(Tech Xplore)—File under Uncomfortable. A study by a Princeton team finds you may be watched (and watched and watched) as you go on about your business of using the Internet and visiting sites.

This is the imposing New World of something called "scripts and session replay." In *New Scientist*, Abigail Beall said that "A website you visit might have hundreds of scripts running in the background; some deposit cookies, others track you to other websites."

Researchers at Princeton University, combing through hundreds of websites to examine the scripts they were running, see "use of a type of script, called a session replay, that logs everything you do on a website, including what you type," she wrote.

The study authors explored "session-replay scripts," third-party scripts on websites.

Wake-up call: Your every keystroke in such circumstances could be recorded.

Researchers at Princeton University have found that 482 globally popular websites are keylogging data and sending it to third-party servers. These are 482 of the world's top 50,000 sites, based on Alexa's ranking.

Findings of the study did not sit easily with Dan Goodin, security editor at *Ars Technica*.

"No, you're not being paranoid. Sites really are watching your every move" was the *Ars Technica* headline on Monday.

Who is providing these session replay scripts and what is their goal?

Goodin in *Ars Technica*: "Session replay scripts are provided by third-party analytics services that are designed to help site operators better understand how visitors interact with their Web properties and identify specific pages that are confusing or broken."

Steven Englehardt, Gunes Acar, and Arvind Narayanan are the study team and they even released data showing the list of sites with third-party session-replay-scripts.

Lately, there have been more sites using such scripts. Keystrokes, mouse movements, and scrolling behavior are recorded, along with contents of the pages visited, and sent to third-party servers. This is not a typical analytics services providing aggregate statistics.

They said the scripts "are intended for the recording and playback of individual browsing sessions, as if someone is looking over your shoulder."

Goodin said the sites discovered in the study use scripts "that record visitors' keystrokes, mouse movements, and scrolling behavior in real time, even before the input is submitted or is later deleted."

All in all, clicks, inputs and scrolls can be recorded and played back later. "Checking the 'do not track' option built into some browsers also failed to stop the logging," Goodin said. And keystrokes typed into a

field may be logged even if the visitor later deletes the field and does not press the submit button.

"Session replay scripts are used by companies to gain insight into how their customers are using their sites and to identify confusing webpages," said Louise Matsakis, *Motherboard*. "But the scripts don't just aggregate general statistics, they record and are capable of playing back individual browsing sessions. The scripts don't run on every page, but are often placed on pages where users input sensitive information, like passwords and medical conditions."

The researchers analyzed seven of the top session replay companies. They found these services in use on 482 of 50,000 websites, "usually with no clear disclosure," said Goodin. They said they found the services in use on 482 of the Alexa top 50,000 sites.

The authors set up test pages and installed replay scripts from six of the seven companies. Based on test results plus analyzing several live sites, the team highlighted four types of vulnerabilities:

1. Passwords. Yes, services they studied try to prevent password leaks by excluding password input fields from recordings. Mobile-friendly login boxes using text inputs to store unmasked passwords are not redacted by this rule, though, unless the publisher manually adds redaction tags to exclude them. 2. Sensitive user inputs are redacted in a partial, imperfect way. 3. Manual redaction of personally identifying information displayed on a page is an insecure model. 4. Recording services may fail to protect user data.

Arvind Narayanan, a team member, said on Monday that "We are publishing this study as a series of blog posts in which we reveal our findings, and then plan to compile them into a paper where we will go into detail on our methods and innovations."

So now for the big question. What can you do about this? One of the researchers said in *Motherboard* that it was difficult for the user to understand what was happening unless one dug deep into privacy policies

Goodin said it was unclear what recourse a user can have that would be meaningful to prevent data collection, as ad-blockers can filter out some, but not all of the replay scripts.

"Until more robust protections are available, people should remember that just about anything they do while visiting a website can be logged," said Goodin.

More information: freedom-to-tinker.com/2017/11/ … sion-replay-scripts/
webtransparency.cs.princeton.e … on_replay_sites.html