Thousands of Popular Websites See What You Type—Before You Hit Submit

A surprising number of the top 100,000 websites effectively include keyloggers that covertly snag everything you type into a form.

Lily Hay Newman May 11, 2022 7:00 AM

Photograph: Shana Novak/Getty Images

When you sign up for a newsletter, make a hotel reservation, or check out online, you probably take for granted that if you mistype your email address three times or change your mind and X out of the page, it doesn't matter. Nothing actually happens until you hit the Submit button, right? Well, maybe not. As with so many assumptions about the web, this isn't always the case, according to new research: A surprising number of websites are collecting some or all of your data as you type it into a digital form.

Researchers from KU Leuven, Radboud University, and University of Lausanne crawled and analyzed the top 100,000 websites, looking at scenarios in which a user is visiting a site while in the European Union and visiting a site from the United States. They found that 1,844 websites gathered an EU user's email address without their consent, and a staggering 2,950 logged a US user's email in some form. Many of the sites seemingly do not intend to conduct the data-logging but incorporate third-party marketing and analytics services that cause the behavior.

After specifically crawling sites for password leaks in May 2021, the

researchers also found 52 websites in which third parties, including the Russian tech giant Yandex, were incidentally collecting password data before submission. The group disclosed their findings to these sites, and all 52 instances have since been resolved.

“If there’s a Submit button on a form, the reasonable expectation is that it does something—that it will submit your data when you click it,” says Güneş Acar, a professor and researcher in Radboud University's digital security group and one of the leaders of the study. “We were super surprised by these results. We thought maybe we were going to find a few hundred websites where your email is collected before you submit, but this exceeded our expectations by far.”

The researchers, who will present their findings at the Usenix security conference in August, say they were inspired to investigate what they call “leaky forms” by media reports, particularly from Gizmodo, about third parties collecting form data regardless of submission status. They point out that, at its core, the behavior is similar to so-called key loggers, which are typically malicious programs that log everything a target types. But on a mainstream top-1,000 site, users probably won't expect to have their information keylogged. And in practice, the researchers saw a few variations of the behavior. Some sites logged data keystroke by keystroke, but many grabbed complete submissions from one field when users clicked to the next.

“In some cases, when you click the next field, they collect the previous one, like you click the password field and they collect the email, or you just click anywhere and they collect all the information immediately," says Asuman Senol, a privacy and identity researcher at KU Leuven and one of the study coauthors. "We didn’t expect to find thousands of websites; and in the US, the numbers are really high, which is interesting,”

The researchers say that the regional differences may be related to companies being more cautious about user tracking, and even potentially integrating with fewer third parties, because of the EU's General Data Protection Regulation. But they emphasize that this is just one possibility, and the study didn't examine explanations for the disparity.

Through a substantial effort to notify websites and third parties collecting data in this way, the researchers found that one explanation for some of the unexpected data collection may have to do with the challenge of differentiating a “submit” action from other user actions on certain web pages. But the researchers emphasize that from a privacy perspective, this is not an adequate justification.

Since completing their paper, the group also had a discovery about Meta Pixel and TikTok Pixel, invisible marketing trackers that services embed on their websites to track users across the web and show them ads. Both claimed in their documentation that a customers could turn on “automatic advanced matching,” which would trigger data collection when a user submitted a form. In practice, though, the researchers found that these tracking pixels were grabbing hashed email addresses, an obscured version of email addresses used to identify web users across platforms, before submission. For US users, 8,438 sites may have been leaking data to Meta, Facebook’s parent company, through pixels, and 7,379 sites may be impacted for EU users. For TikTok Pixel, the group found 154 sites for US users and 147 for EU users.

The researchers filed a bug report with Meta on March 25, and the company quickly assigned an engineer to the case, but the group has not heard an update since. The researchers notified TikTok on April 21—they discovered the TikTok behavior more recently—and have not heard back. Meta and TikTok did not immediately return WIRED's request for comment about the

findings.

“The privacy risks for users are that they will be tracked even more efficiently; they can be tracked across different websites, across different sessions, across mobile and desktop,” Acar says. “An email address is such a useful identifier for tracking, because it’s global, it’s unique, it’s constant. You can’t clear it like you clear your cookies. It's a very powerful identifier.”

Acar also points out that, as tech companies look to phase out cookie-based tracking in a nod to privacy concerns, marketers and other analysts will rely more and more heavily on static IDs like phone numbers and email addresses.

Since the findings indicate that deleting data in a form before submitting it may not be enough to protect yourself from all collection, the researchers created a Firefox extension called LeakInspector to detect rogue form collection. And they say they hope their findings will raise awareness about the issue, not only for regular web users but for website developers and administrators who can proactively check whether their own systems or any of the third parties they're using are collecting data from forms without consent.

Leaky forms are just one more type of data collection to be wary of in an already extremely crowded online field.