Most websites do not publish privacy policies, researchers say

October 27 2023, by Mary Fetzer

Credit: Markus Winkler from Pexels

Online privacy policies may not only be difficult to find but nonexistent, according to Penn State researchers who crawled millions of websites and found that only one-third of online organizations made their privacy

policy available for review.

"Privacy Lost and Found: An Investigation at Scale of Web Privacy Policy Availability," a paper authored by students and faculty from the Penn State College of Information Sciences and Technology (IST), detailed an analysis of the online privacy policy landscape and studied the unavailability of privacy policies on company domains. It received the Best Student Paper Award at the 23rd Association for Computing Machinery Symposium on Document Engineering, also known as DocEng'23, held Aug. 23–25 at the University of Limerick in Ireland.

"Privacy policies are legal documents that organizations use to disclose how they collect, analyze, share and secure their online users' personal data," said Mukund Srinath, doctoral student in the College of IST and lead author of the paper. "Privacy policies are often the only source of information regarding what happens to users' personal information online. The availability of privacy policies and the ability of users to understand them are fundamental to ensuring that individuals can make informed decisions about their personal information."

Legal jurisdictions around the world require organizations to post privacy policies on their websites. The European Union, for example, regulates this disclosure through laws such as the General Data Protection Regulation (GDPR). In the United States, privacy policy regulations are set at the state level, such as the California Privacy Rights Act (CPRA).

These laws work under the principle of notice and choice, according to the researchers. Notice is a presentation of terms—in this case, the privacy policy—and choice is an action signifying the acceptance of those terms, such as clicking an "Accept" link or simply continuing to use the site.

Despite regulations such as GDPR and CPRA, most organizations are not in compliance, according to the researchers. That could mean that a company does not post its privacy policy or that it does so ineffectively, such as with a broken link, a blank page or unreadable content.

"Not many websites have privacy policies," Srinath said. "For a user landing on a random website, there is only a 34% chance that a privacy policy exists. Among them, there is a 2% to 3% chance that the link is broken. And 5% of the links that do work will lead to a page that contains irrelevant information, such as placeholder text or documents in a language that doesn't match the website's landing page."

The researchers conducted a large-scale investigation of the availability of privacy policies by crawling millions of English-language websites to identify when privacy policies were unavailable. They used the capture-recapture technique to estimate the frequencies of the failure modes and the overall unavailability of privacy policies on the web.

"We borrowed the technique that ecologists might use for animals in the wild," said Pranav Venkit, doctoral student in the College of IST and co-author of the paper. "They go into a forest of bears, capture a small sample, tag them and send them back into the wild. They go back the next day and capture another set. The unseen versus the previously seen bears enable the ecologists to estimate the bear population."

According to the researchers, proper resources are needed to support efforts to improve privacy policy availability rates.

"Regulators cannot keep up," Srinath said. "They are often overwhelmed by the numbers of privacy policies on the web and forced to rely on user complaints or compliance self-certification to prompt investigations of missing or ineffective privacy policies."

Promoting transparency and accountability in online data privacy practices is critical to the continued growth and development of the digital economy, according to co-author Shomir Wilson, assistant professor of IST, director of the Human Language Technologies (HLT) Lab at Penn State and adviser to Srinath and Venkit.

"This research provides important insights into the current state of privacy policy practices on the web that can inform efforts to develop more effective privacy policy standards and best practices as well as to improve the accessibility and comprehensibility of existing policies for users," Wilson said.

Provided by Pennsylvania State University