

Study reveals misuse of archive services by fringe communities on the web

June 26 2018, by Tiffany Westry Womack



Credit: CC0 Public Domain

In a large-scale analysis, researchers at the University of Alabama at Birmingham, Cyprus University of Technology and University College London reveal fringe communities within Reddit and 4chan push the use

of URLs from archive services to avoid censorship and undercut advertising revenue of new sources with contrasting ideologies.

"Web archiving services play an increasingly important role in today's information ecosystem by preserving online content," said Jeremy Blackburn, Ph.D., assistant professor of computer science in the UAB College of Arts and Sciences. "News and [social media posts](#) have been found to be the most common types of content archived. URLs of archiving services are extensively shared on 'fringe' communities within Reddit and 4chan to preserve possibly contentious content."

Researchers analyzed millions of URLs from archive.is and Wayback Machine shared on four social networks: Reddit, Twitter, Gab and 4chan's politically incorrect board (/pol/). The results of the study were published this week in a paper at the 12th International Conference on Web and Social Media in Stanford, California.

The social-network-specific analysis shows, among other things, that moderators leverage web archiving services to ensure content shared on their community persists. In particular, they found that 44 percent of URLs from archive.is and 85 percent of URLs from Wayback Machine URLs are shared by Reddit moderation bots. Web archiving services were also found to be used extensively for the archival and dissemination of content related to conspiracy theories and world events related to politics, suggesting these services play an important role in the alternative news ecosystem.

Additional evidence shows moderators from specific subreddits force users to misuse web archiving services so as to ideologically target certain news sources by depriving them of traffic and potential ad revenues. Links from unwanted news websites shared are deleted, and users are prompted to utilize a cached link, screenshot or archive.is.

"For example, we observed that 'The Donald' subreddit systematically targets ad revenue of news sources with conflicting ideologies," Blackburn said. "Moderation bots block URLs from those sites and prompt users to post archived URLs. According to our conservative estimates, popular news site like the Washington Post lose approximately \$70,000 worth of ad revenue annually due to the use of archiving services on Reddit."

The analysis reveals that out of 3,800 submissions made to Reddit using links from the Washington Post and 3,300 submissions with links from CNN, 44 percent and 39 percent were removed.

"These findings highlight the importance of archiving services in the web's information and ad ecosystems, the need to carefully consider them when studying [social media](#) and when designing systems to detect and contain the cascade of misinformation on the web," Blackburn said.

Blackburn is a co-founder of the International Data-driven Research for Advanced Modeling and Analysis Lab, or iDRAMA Lab, an international group of scientists focusing on modern socio-technical issues with expertise ranging from low-level cryptography to video games.

More information: Understanding Web Archiving Services and Their (Mis)Use on Social Media. arxiv.org/abs/1801.10396

Provided by University of Alabama at Birmingham

Citation: Study reveals misuse of archive services by fringe communities on the web (2018, June 26) retrieved 10 April 2024 from <https://techxplore.com/news/2018-06-reveals-misuse-archive-fringe-web.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.