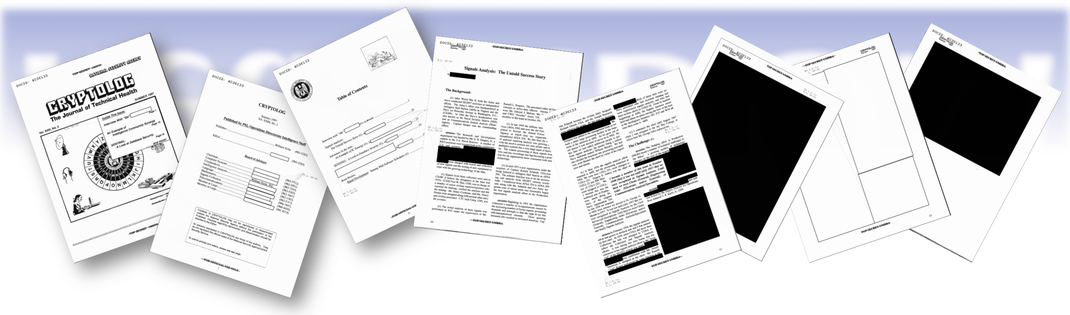
RedactDOM

Preventing Sensitive Data Leaking through Embedded Scripts



Longze Chen, Yuchen Zhou and David Evans
University of Virginia

RedactDOM is an egress-based multi-execution approach that prevents embedded scripts from leaking sensitive information to external servers.

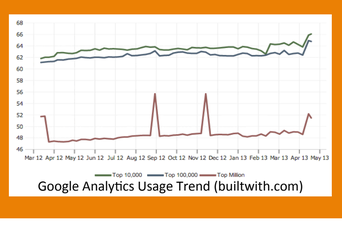
RedactDOM duplicates pages into two synchronized but isolated executions: a **real page** that contains full content and is visible in the browser, and a hidden **redacted page** that only contains insensitive content.

We rely on the **assumption** that all sensitive information on the page is in the form of text content or attributes. Correct execution of (non-malicious) scripts does not normally depend on those values. The host scripts themselves do not contain private information.

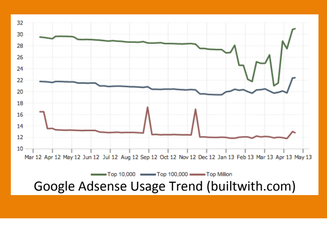
The pages **multi-execute** at the same time. Network traffic from both pages is redirected to a **proxy monitor**, which enforces that no sensitive information is leaked to external servers.

Risks of Embedded Scripts

Modern websites employ dozens of third-party scripts for advertising, analytics, social networking, and other services. These scripts can access any sensitive web content and leak it to external servers.



Google Analytics Usage Trend (builtwith.com)



Google Adsense Usage Trend (builtwith.com)

Dealing with Divergence

The behavior of the redacted page and the normal page may diverge, either because scripts depend on redacted content, user events are not delivered to the redacted page, or other causes. This might further break inter-request correspondence on which our design relies. When this happens, we plan to explore regenerating a new redacted page by re-redacting the page when necessary.