A Large-scale Analysis of Content Modification by Open HTTP Proxies

open HTTP proxies are an attractive option for bypassing IP-based filters and geo-location restrictions, circumventing content blocking and censorship, and in general, hiding the client's IP address when accessing a web server.

But trusting a third party can have consequences, Including collecting user information or directing the user to pages that contain malware and ...
5.15% of the tested proxies were found to perform modification or injection that can be considered as malicious or unwanted.
One solution is to use end-to-end encryption, However, the problem is not alleviated even when end-to-end encryption is used, as man-in-the-middle attacks are still possible using fake or even valid.
By identifying and analyzing multiple cases of content injection and modification, this study provides insights about the behavior of rogue web proxies and reveals many patterns that exist between these modifications. This enabled us to build a web service for assessing web proxies on a daily basis, and making publicly available a list of proxies that did not perform any modification during our tests.
To understand and measure the extent of content modification by rogue HTTP proxies, in this work, we have designed a methodology for detecting and analyzing content alteration and code injection attempts. Specifically, we have built a framework that regularly collects HTTP proxies from several "proxy list" websites, and tests them using a novel technique based on decoy websites (dubbed honeysites) under our control. Furthermore, we have implemented a content modification detection approach that operates at the level of a page's DOM tree, which can detect even slight object modifications. To facilitate the analysis of content modification incidents, we have implemented a clustering technique for grouping together cases of content modification that follow similar patterns.

:Malicious proxy behaviors fall into the following categories

Injection ad.
Tracking users
Fingerprinting
Privacy leakage.
Malware.
Unclassified behavior.

We opted for a simple but effective scheme that relies on decoy websites under our control to detect content modification without false positives. This approach, however, is not robust against determined rogue proxy operators who may anticipate our attempts, and refrain for performing any suspicious activity.

Our service is available at http://proxyscan.ics.forth.gr