Taking a Look into the Cookie Jar

[1]Sean Chen, [2]Jaelyn McCracken, [3]Tao Hou

[1]Texas A&M University, seanchen25@tamu.edu [2]Towson University, jmccra7@student.towson.edu [3]Texas State University, taohou@txstate.edu

Introduction

Cookies play a crucial role in authenticating users and enabling various functionalities on websites. However, they can also be vulnerable to cyber-attacks, particularly cross-site scripting, which poses a serious risk to user privacy. To address these concerns and enhance overall web security, security header flags such as Samesite, HttpOnly, and Secure were introduced. These protections aim to deter attackers from exploiting cookies and safeguard user privacy.

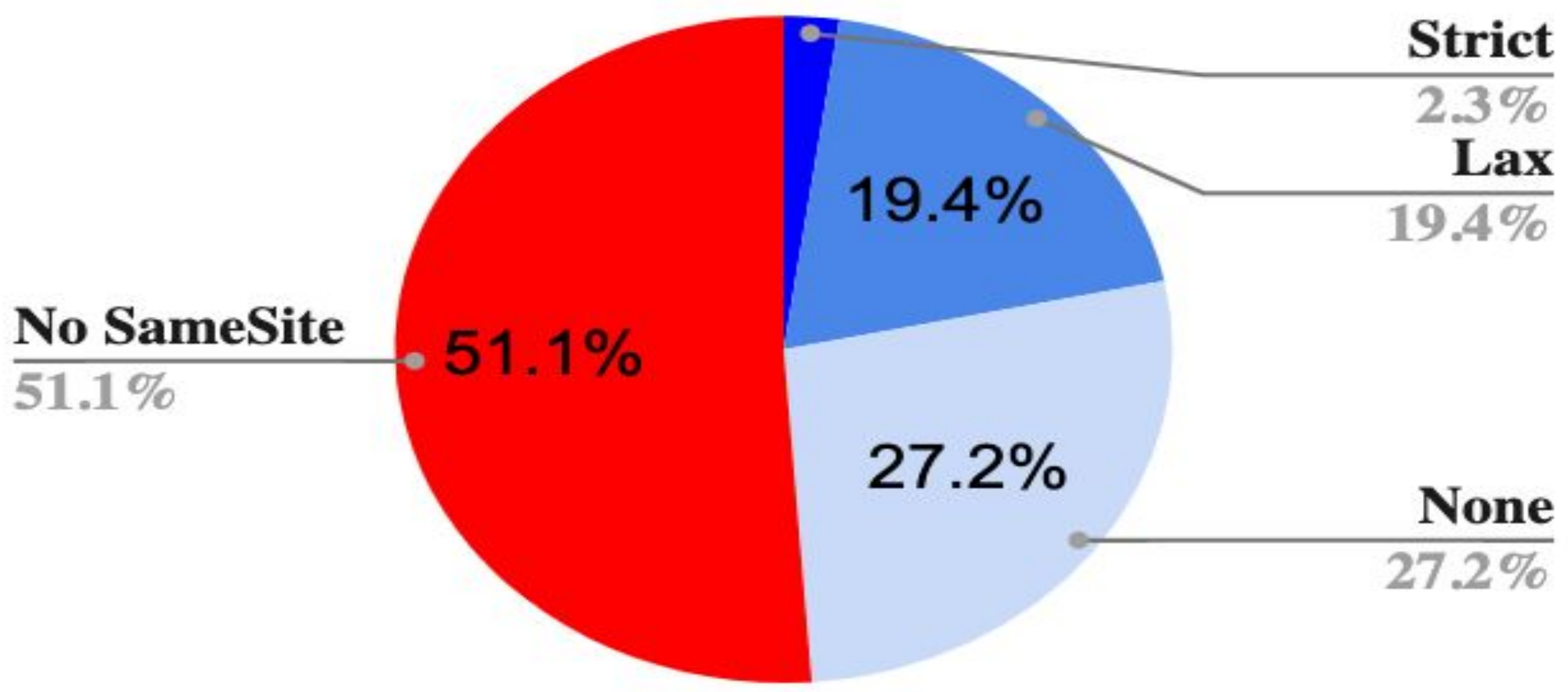
In this project, we aim to gain insights into the deployment situation of these security flags on the Internet. Towards this, we developed a tool to scrape the top, middle, and bottom 1000 websites out of the 1 million most visited websites and conducted a thorough and comprehensive analysis of the usage of these security flags.

Methodology

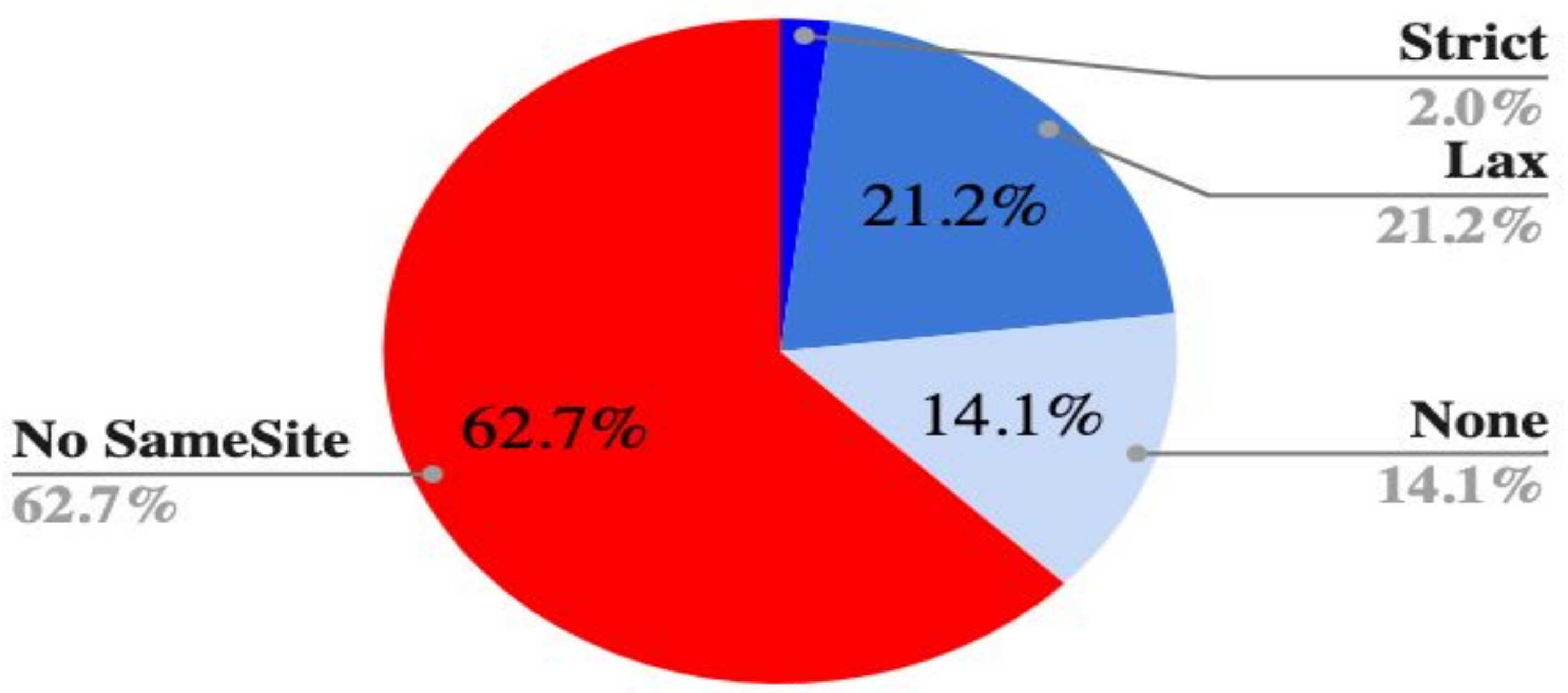
- Libraries:
  - Urllib3
  - Requests
- Data: obtained a list of top 1 million websites (ranked from Alexa)
- Scraper:
  - Request library allowed for sending HTTP requests w/ User-Agent
  - Returns header: 'Set-Cookie'
  - Runs through text file w/ all websites to visit (sleep function for delay)
  - Parses & prints cookie attributes in separate files for each website (only valid visits recorded)
- Merged all 1000 files onto one file for each category
  - Makes parsing easier
- Parsing (with additional python scripts)
  - Extract/print names of websites that have specific attributes
  - Separate those into files to better organize for analysis

Results

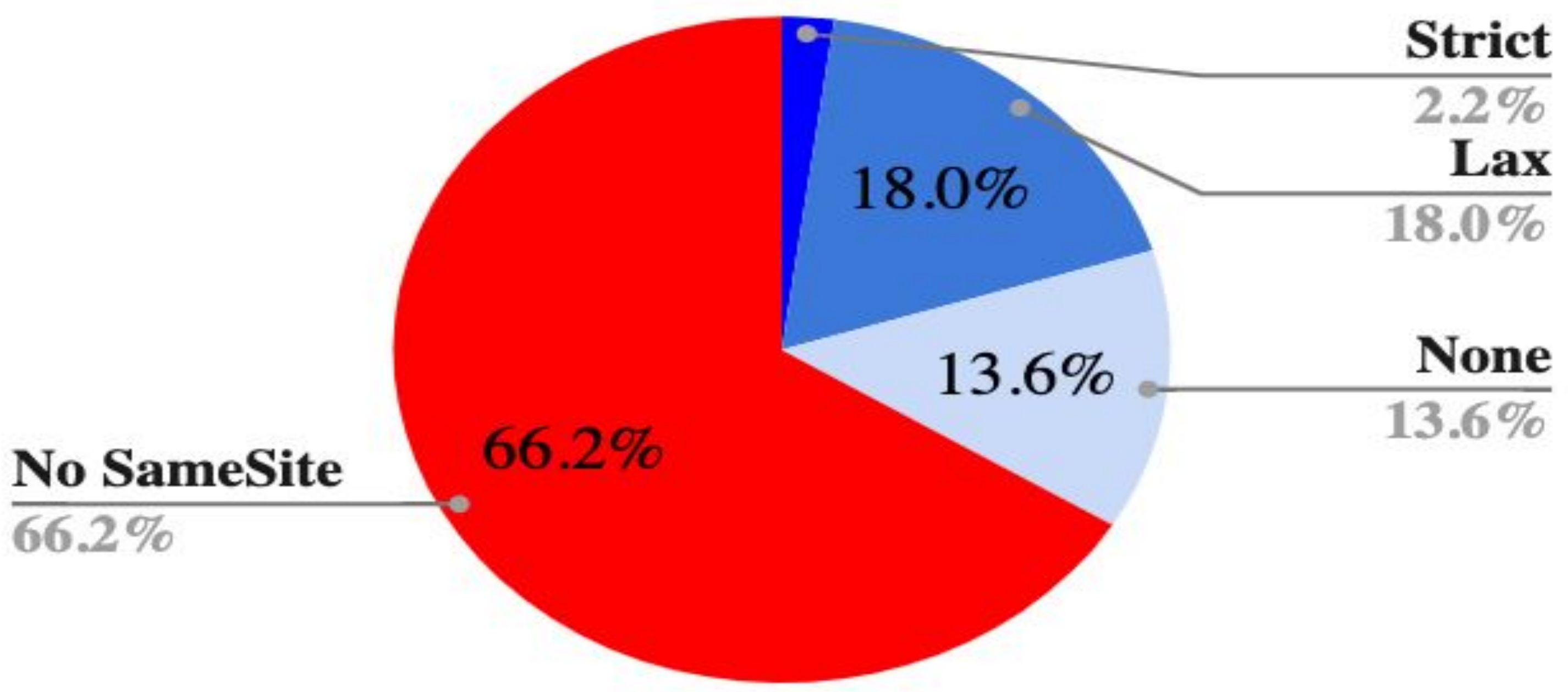
Top 1000 Websites



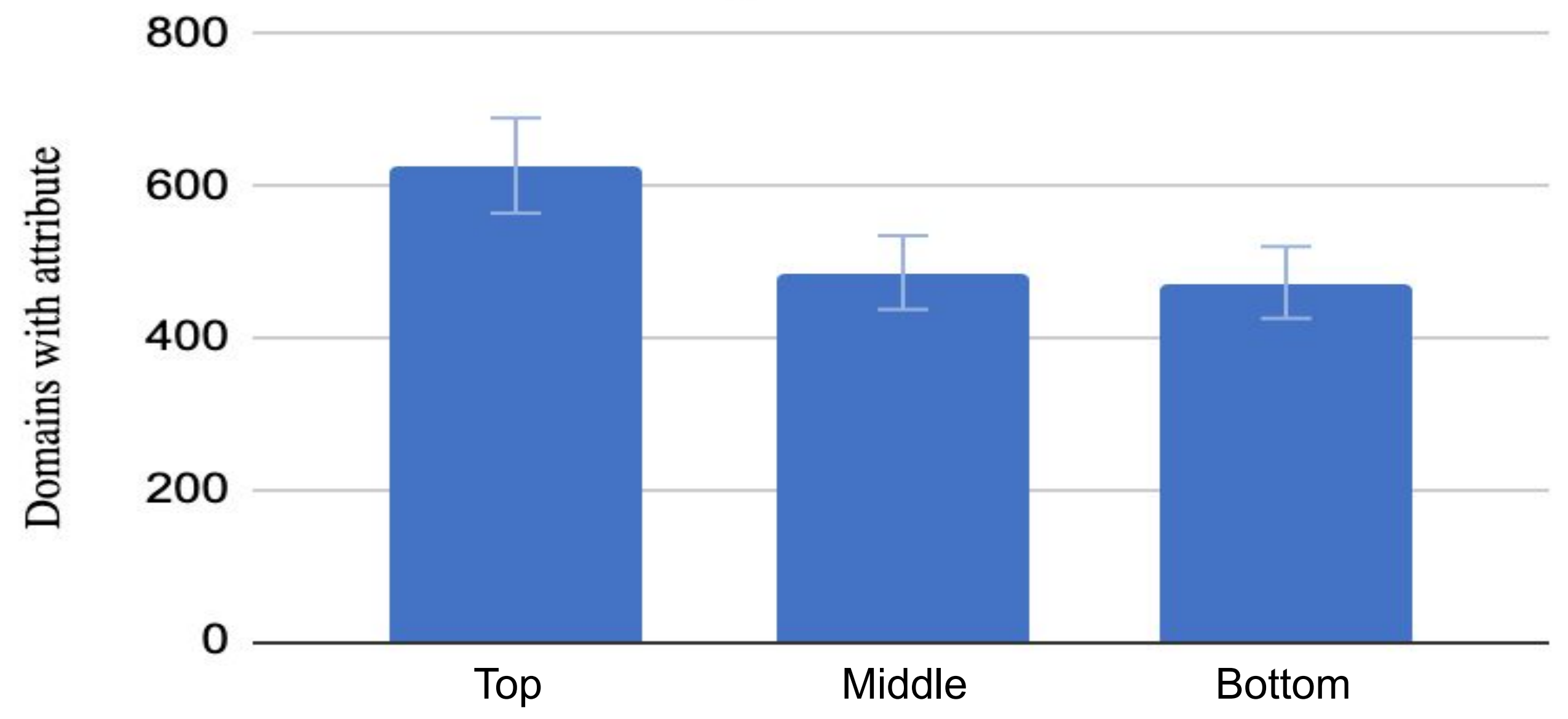
Middle 1000 Websites



Bottom 1000 Websites



Secure Attribute Usage



Background

- SameSite=Strict: only 1st party cookies can be sent/accessed, but not if incoming link is from external site
- SameSite=Lax: enables only 1st party cookies to be sent/accessed
- SameSite=None: cookie data can be shared with 3rd parties & external sites, also specify a 'Secure' attribute
- Secure: cookie only sent when connecting through SSL
- HttpOnly: prevents client-side scripts from accessing cookie data

Discussion

The top 1000 websites show a significant increase in the usage of SameSite attributes when compared to the middle and bottom 1000. This could be due to the fact that the top websites have more reputable adversitors. On the other hand, the middle and bottom sites may need to give greater access to 3rd parties and external sites to increase traffic and revenue. Even though the top sites had a large 'no SameSite' percentage, they also had more secure attributes as well. In addition, the middle and bottom websites differed slightly in their cookie attributes, and this is because they are not as popular as the top websites, so their security is similar too.

Future Work

- Scrape more websites on the bottom and compare their results to our current findings
- Set up selenium & test w/ proxy to automate different browsers to test/analyze their cookie defense
- Extrapolate to find more vulnerabilities and potential exploits

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CNS-2149950. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.