PriSEC: A Privacy Settings Enforcement Controller

R. Khandelwal, T. Linden, H. Harkous, K. Fawaz

wiscprivacy.com
Privacy settings are difficult to find.

- Users need to navigate to ten URLs to change a privacy setting
- Privacy choices are inconsistent across websites
A lot of settings, even when you know how to find them
Over 60 privacy settings to choose from
Even if you know where the privacy settings are, it’s cumbersome to find the one you want.
Related Works

Android Privacy
- Identifies “hard-to-find” privacy settings via the analysis of app source code

Web Settings
- Presents the automatically-detected opt-out options to the users via a web extension

Limited to Android apps
Limited to opt-out settings

Chen et al., Demystifying Hidden Privacy Settings in Mobile Apps, IEEE S&P 2019
Kumar et al., Finding a Choice in a Haystack, WWW ‘20
Research Objective

Address reachability and usability issues of web privacy settings

- Automatically find privacy control pages on the web
- Extract fine-grained options for privacy settings
- Design usable interface for easy access and enforcement of settings
PriSEC Overview

Client

User Interface

Enforcer

Domain Crawler

Backend

Recipe Generator

Control recipes

<query>

Control recipe

Client Service

Recipe Database

Control recipes

HTML Control pages

JSON
PriSEC Overview

Client

User Interface

Enforcer

Backend

Domain Crawler

Recipe Generator

Control pages

Recipe Database

Client Service
Domain Crawler

1. **Web Search**
   - Domain

2. **Landing URL**
   - BFS + Pruning

3. **IsControl Classifier**
   - Control pages
Recipes consist of:

- **Group Text**
- **Option Text**
- **Type of Settings**
- **XPath to the element**
Challenges in Recipe Generation

- Need to identify the control elements
- Cannot do static analysis due to dynamic element loading
- Need to extract text; difficult due to non-standard HTML

Can be implemented using `<div>` tag

Or can be implemented using `<input>` tag
Recipe Generation

Identify Focusable Elements

Push notifications
Get push notifications to find out what's going on when you're not on Twitter. You can turn them off anytime.

Related to you and your Tweets
When you turn on Tweet notifications from people you follow, you'll get notifications about their Tweets or live videos. These will be push or SMS notifications, depending on your settings. View users

Generate Dependency Graph

Classify UI Elements

Push notifications
Get push notifications to find out what's going on when you're not on Twitter. You can turn them off anytime.

Related to you and your Tweets
When you turn on Tweet notifications from people you follow, you'll get notifications about their Tweets or live videos. These will be push or SMS notifications, depending on your settings. View users

Group Elements

Push notifications
Get push notifications to find out what's going on when you're not on Twitter. You can turn them off anytime.

Related to you and your Tweets
When you turn on Tweet notifications from people you follow, you'll get notifications about their Tweets or live videos. These will be push or SMS notifications, depending on your settings. View users
PriSEC Overview

Client

User Interface

Enforcer

Backend

Domain Crawler

Recipe Generator

Control pages

Client Service

Recipe Database
Workflow of Client Side

1. Receive input from PriSEC Extension
2. Execute script for privacy recipe
3. Correct privacy setting enforced
PriSEC Demo

Coming Soon!
Evaluation Datasets

PCP Dataset
(Privacy Control Page Dataset)

Privacy control page annotations for Alexa top-500 domains

100 pages from 58 domains

NLQ Dataset
(Natural Language Query Dataset)

Free form user queries from Twitter and Reddit

135 queries from 15 domains
End-to-End Evaluation

**PCP Set:** 100 control pages from 58 domains

- Candidates: 1400
- Pages Missed: 5
- Control Pages: 323
- Pages Missed: 0
- 94.2% accuracy
- Second stage filter
I do not want to allow search engines to index my reddit user profile. How can I do that?
## Semantic Matching

**NLQ Dataset:** 135 queries from 15 domains

Top-k accuracy for different encoders on NLQ dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>Top-1</th>
<th>Top-3</th>
<th>Top-5</th>
<th>Top-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>USE</td>
<td>66.7</td>
<td>89.6</td>
<td>95.6</td>
<td>100</td>
</tr>
<tr>
<td>USE-Lite</td>
<td>62.9</td>
<td>84.4</td>
<td>91.8</td>
<td>97.8</td>
</tr>
<tr>
<td>SBERT</td>
<td>48.2</td>
<td>76.3</td>
<td>84.4</td>
<td>92.6</td>
</tr>
<tr>
<td>SRoBERTa</td>
<td>45.9</td>
<td>69.6</td>
<td>85.2</td>
<td>90.4</td>
</tr>
</tbody>
</table>

- Universal Sentence Encoder outperforms other encoders
- USE-Lite can be deployed locally at a small cost in accuracy
User Based Evaluation

- 148 participants
- Within subject study
- 6 popular domains
- 2 privacy tasks/user
- Task completion time
- SUS usability score
User Based Evaluation

Participants took 3.75x more time when using the baseline method
**Takeaways**

**PriSEC**: Automated framework for extraction, annotation and enforcement of privacy choices

- Annotates and extracts recipes for 94.3% of the control pages
- Makes it easier to find settings via query interface
- Reduces user burden and achieves better usability scores than baseline approach

[link to wiscprivacy.com](http://wiscprivacy.com)  [email to rkhandelwal3@wisc.edu](mailto:rkhandelwal3@wisc.edu)