

An Efficient and Usable Client-Side Cross Platform Compatible Phishing Prevention Application

FINAL YEAR PROJECT REPORT

Guide

Dr. Angelin Gladston
Associate Professor
Department of CSE

Submitted by

N. Dhanush	2016103021
G. Santhosh	2016103057
S. Ben Stewart	2016103513

OUTLINE

1. INTRODUCTION
2. OVERALL OBJECTIVE
3. LITERATURE SURVEY
4. PROPOSED SYSTEM
5. HIGH LEVEL BLOCK DIAGRAM
6. IMPLEMENTATION
7. EVALUATION METRICS
8. REFERENCES

INTRODUCTION

- Phishing
- Lists of such sites
- Time constraints
- Computational resources
- Vulnerabilities
- Cross platform

OVERALL OBJECTIVE

- Create a phishing list
- Cross Platform application
- Web browser add-on
- Provide temporal resilience
- Remove false positives from list

LITERATURE SURVEY

- Previous work by Samuel Marchal, Giovanni Armano, Tommi Grondahl, Kalle Saari, Nidhi Singh, and N. Asokan
- Implemented a client-side phishing prevention application.
- Had background tasks communicate with a browser add-on.
- Not platform independent.

AUTOMATIC PHISHING CLASSIFICATION

- Colin Whittaker, Brian Ryner and Marria Nazif for Google
- Proc. Netw. Distrib. Syst. Security Symp., 2010
- Features used
 1. The URL of the page
 2. The HTML page contents
 3. The host server details
- Needs blacklist updating.

CANTINA

- Guang Xiang, Jason Hong, Carolyn P. Rose and Lorrie Cranor
- ACM Trans. Inf. Syst. Secur., 2011
- Page similarity
- SHA 1 algorithm
- Easy to break
- Performance gains

AUTO UPDATED WHITELIST

- Ankit Kumar Jain and B. B. Gupta
- EURASIP J. Inf. Secur., vol. 2016, no. 1, Dec. 2016
- Whitelist
 - a. the domain name
 - b. the IP address
- Reverts to old system if not in whitelist

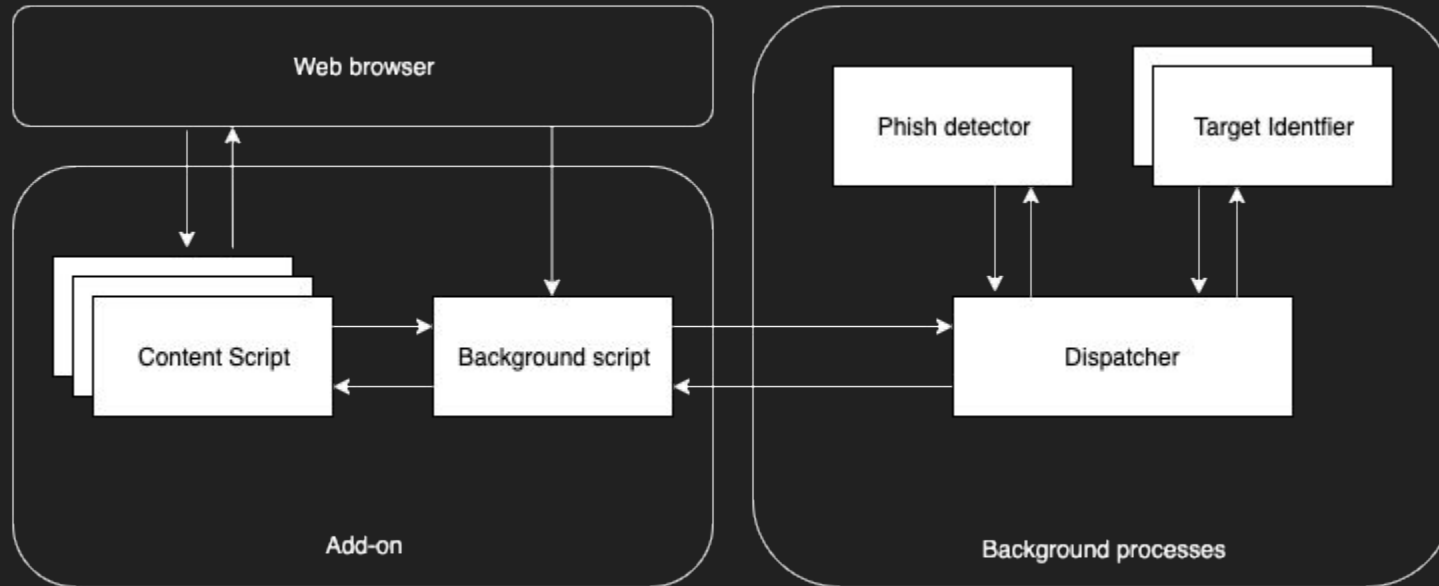
COMPARISON

Paper	Publication	Solved	Limitations
Large-Scale Automatic Classification of Phishing Pages	Proc. Netw. Distrib. Syst. Security Symp., 2010	Machine learning model can be used with reliable accuracy.	Needs blacklist for updating.
CANTINA: A feature-rich machine learning framework for detecting phishing Web sites	ACM Trans. Inf. Syst. Secur., 2011	SHA1 based similarity check for similar looking sites.	SHA1 could be manipulated.
A novel approach to protect against phishing attacks at client side using auto-updated white-list	EURASIP J. Inf. Secur., vol. 2016, no. 1, Dec. 2016	Auto-updated whitelist for faster detection of sites on average.	Not temporally resilient.
Off-the-hook: An efficient and usable client-side phishing prevention application	IEEE Trans. Comput., vol. 66, no. 10, pp. 1717-1733, Oct. 2017	A user oriented implementation.	Imperfect domain name string splitting.

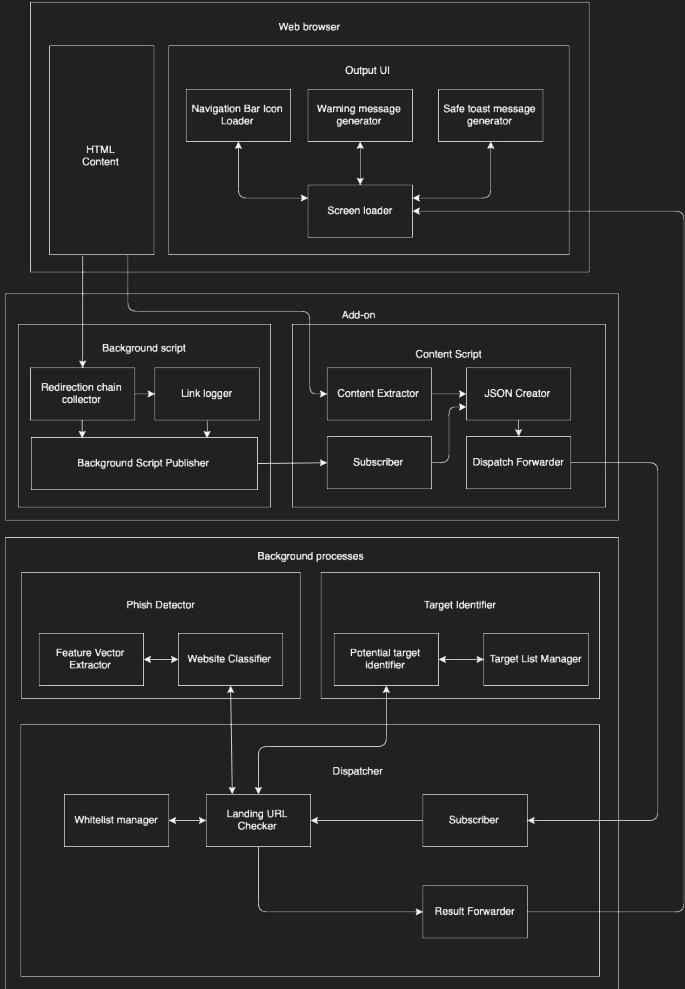
PROPOSED SYSTEM

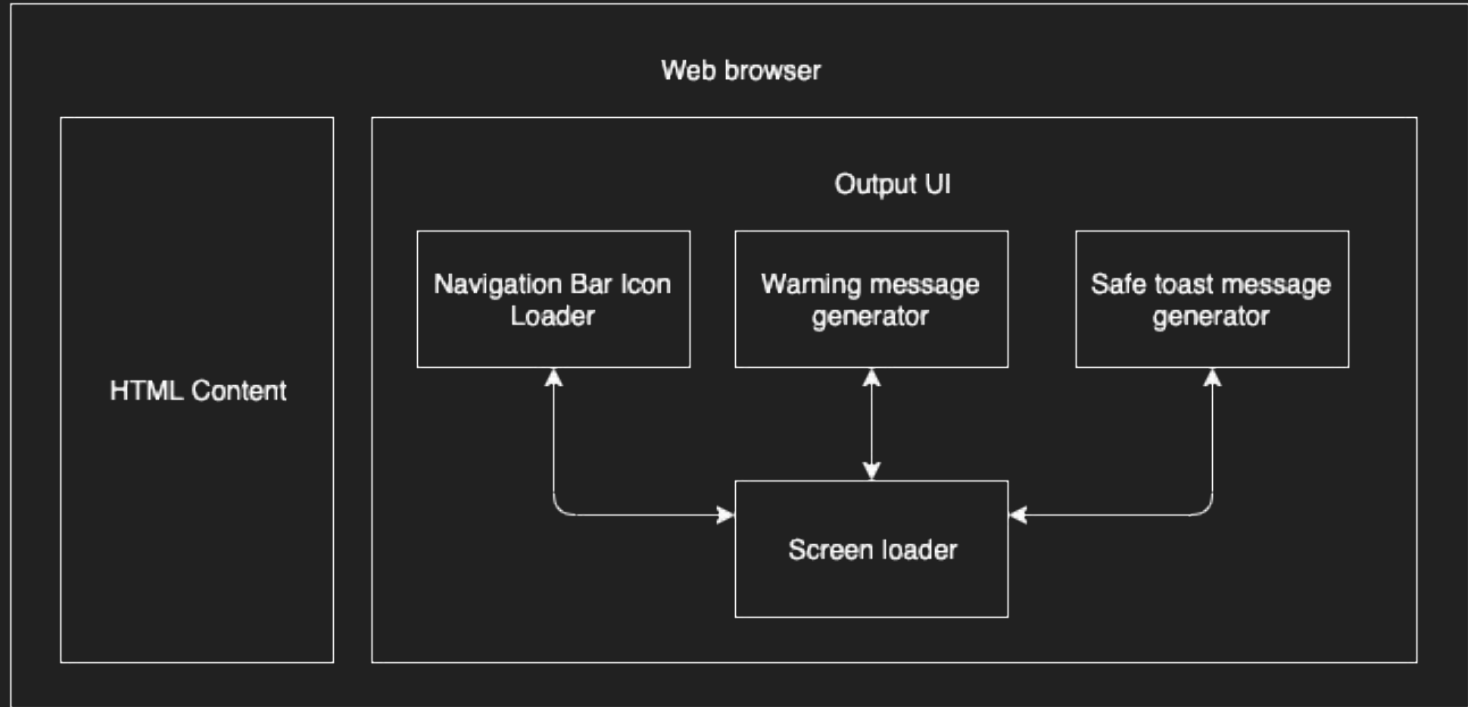
- Platform independent
- Browser add-on
- Reduce false warnings
- Context independent detection
- Static observations

SYSTEM ARCHITECTURE



HIGH LEVEL BLOCK DIAGRAM





Add-on

Background script

Redirection chain
collector

Link logger

Background Script Publisher

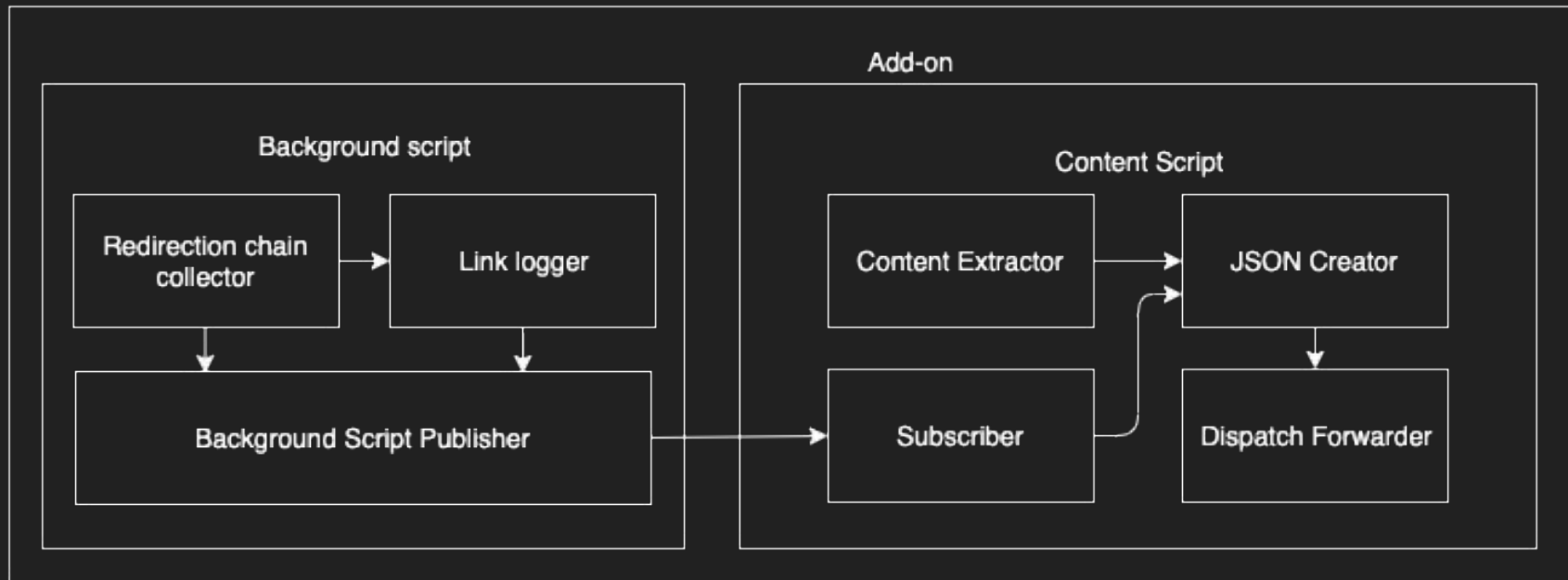
Content Script

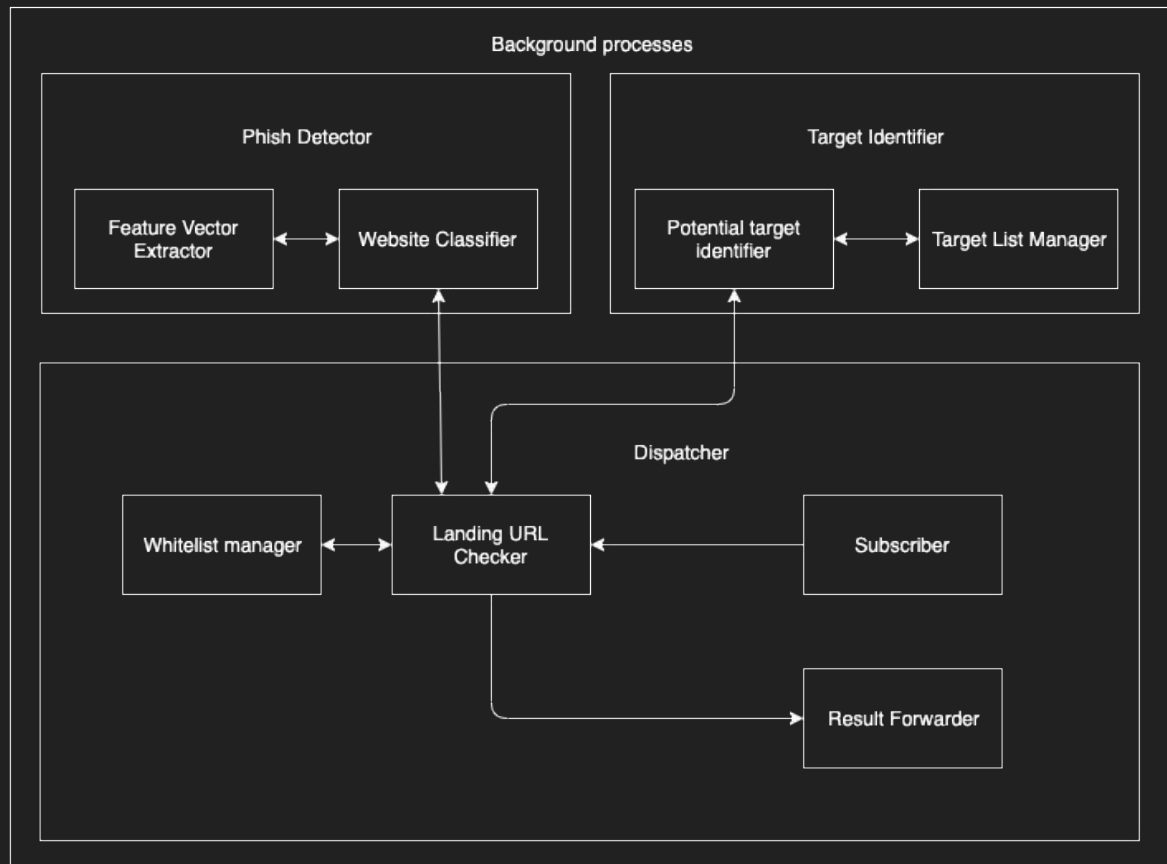
Content Extractor

JSON Creator

Subscriber

Dispatch Forwarder





MODULE LIST

- Add on
 - a. Background script
 - b. Content script
- Background process
 - a. Dispatcher
 - b. Phish Detector
 - c. Target Identifier
- Web Browser
 - a. HTML content
 - b. Output UI

BACKGROUND SCRIPT

Begin

For each page load redirect

Add listener to that event

Get the list of redirects from listener

If page is fully loaded

Send the list of redirects to content script

Done

End

CONTENT SCRIPT

Begin

For each page load redirect

If page is fully loaded

Get the URL from the tab

Get the HTML content from innerHTML tag

Get redirection list from background script

Send them to the background process

Done

End

DISPATCHER

Begin

If page address is in whitelist

Send the GREEN signal

Else

Send content to phish detector

Get results from phish detector

If phish is FALSE

Send the GREEN signal

Else

Send the RED signal

Send content to target identifier

If target is found

Publish target

Else

No target matched

End

PHISH DETECTOR

Begin

For each page URL

Get the feature values for the URL

Load the saved model

Publish the result

Done

End

TARGET IDENTIFIER

Begin

Get the hash value for page content

Compare with values in hash list

If match

Display target

Else

No target found

End

IMPLEMENTATION

- CONTENT SCRIPT
- BACKGROUND SCRIPT

Chrome

File Edit View History Bookmarks People Tab Window Help

Incognito

Check your WebPage

Find out now...

SAFE OR NOT? ✓

Content Script

```
https://www.google.com/<br><doctype html><html
itemscope="" itemType="http://schema.org/WebPage"
lang="en-IN"><head><meta content="text/html;
charset=UTF-8" http-equiv="Content-Type"><meta
content="/images/branding/google/1x/
googlep_standard_color_128dp.png"
itemprop="image"><title>Google</title><script
nonce="9nVJqg1H+VucdU6aFmhAQ==">(function()
(window.google=(kEI:5JUHxp3HcyH4-
EPsv2PwAM,kEXPI:0,1353746,5663,730,224,4727,37
8,206,467,1947,540,250,10,168,122,423,338,175,364,
671,483,3,278,4,60,315,427,208,10,1129468,143,11977
09,4,40,38,329080,1294,12383,4855,32691,15248,86
7,1744,11240,383,3320,5505,8384,1700,3158,1582,4
323,4968,3029,3888,850,3118,6196,1714,1,1812,1976,
2044,5762,1,3146,5297,2054,920,873,1217,1336,4375,
1558,1720,1,415,1141,7509,2,1782,252,620,2884,20,31
8,689,2483,975,1,368,2778,519,402,990,1285,10,279
4,967,612,14,667,612,2212,202,328,149,1103,840,517,
317,1157,50,156,662,3438,108,152,52,1132,4,3,2063,6
06,1839,184,1777,143,377,886,1261,747,429,44,1009,9
3,339,1283,16,8,...antfont-weight:bold).gbm1(border-
top:1px solid #bebebe;font-size:0;margin:10px 0)
#gbd4 .gbmc(background:#f5f5f5;padding-top:0)
#gbd4 .gbsbic::-webkit-scrollbar-
track:vertical(background-color:#f5f5f5;margin-
top:2px)#gbmpdv(background:#fff;border-bottom:1px
solid #bebebe;-moz-box-shadow:0 2px 4px
rgba(0,0,0,12);-o-box-shadow:0 2px 4px
rgba(0,0,0,12);-webkit-box-shadow:0 2px 4px
rgba(0,0,0,12);box-shadow:0 2px 4px
rgba(0,0,0,12);position:relative;z-index:1)
#gbd4 .gbmh(margin:0).gbmc(padding:0;margin:0;lin
e-height:27px).GBMCC:last-child:after,.GBMPAL:las
t-child:after(content:"\0A\0A";white-
space:pre;position:absolute)#gbmps("zoom:1)
#gbd4 .gbpc.#gbmps .gbml(line-height:17px)
#gbd4 .gbps .gbmc(line-height:27px)
#gbd4 .gbmtc(border-bottom:1px solid #bebebe)
#gbd4 .gbpc(display:inline-block;margin:16px 0
10px;padding-right:50px;vertical-align:top)
#gbd4 .gbpc("display:inline).gbpc .gbps .gbpc .gbps2
(display:block;margin:0 20px)#gbmplp.gbps(margin:0
10px).gbpc .gbps(color:#000;font-weigh
```

OK

India

Advertising Business About How Search works

CONTENT SCRIPT

//Retrieve URL JS

tablink = tab.url;

//Retrieve Page content PHP

\$site=\$_POST['url'];

\$html = file_get_contents(\$site);

BACKGROUND SCRIPT

Chrome File Edit View History Bookmarks People Tab Window Help

Amazon.com: Kindle Keyboard x +

amazon.com/Kindle-Wireless-Reading-Display-Globally/dp/B003FSUDM4/ref=amb_link_353259562_2?pf_rd_m=ATVPDKIKX...

amazon

Deliver to India

Today's Deals Help Registry Gift Cards Sell Your Account

We ship internationally
We're showing you items that ship to India. To see items that ship to a different country, change your delivery address.
Additional language and currency settings are available. [Learn more](#)

Don't Change Change Address

Kindle Wireless Reading Display

Shop year-end deals Dec. 24 to Jan. 1

Past Is Prologue
Everything begins with a story
— Joseph Campbell

I was born in Toronto, one month early and during a blizzard that covered the city in snow and silence. The surprise and the low-mortality conditions that accompanied my arrival were patterns, though they went unrecognized at the time. My mother, as a recent immigrant from India, was of two worlds, and she would pass that multiple identity on to me. My father was making his way to Canada, but had not yet arrived; his absence at my birth was a sign of the deeper absence yet to come. Looking back, I see all the ways in which my life was set the moment I was born into it. Whether in the street or in silence, whether by the hand of God or some unnameable force, it was already written, and every action of mine would serve to confirm the text.

22%
Launches 10-48 23%

Background Script

- <http://tinyurl.com/KindleWireless>
307: Internal (browser cached) redirect to https://tinyurl.com/KindleWire
- The server has previously indicated this domain should always be accessed via HTTPS (HSTS Protocol). Chrome has cached this internally, and did not connect to any server for this redirect. Chrome reports this redirect as a "307 Internal Redirect" however this probably would have been a "301 Permanent redirect" originally. You can verify this by clearing your browser cache and visiting the original URL again.
- <https://tinyurl.com/KindleWireless>
301: Permanent redirect to http://www.amazon.com/Kindle-Wireless-Res
- <https://r.eablink.com/?key=a982cfabb5482be54a4d3a52b21>
302: Temporary redirect to http://www.amazon.com/Kindle-Wireless-Res
- <http://www.amazon.com/Kindle-Wireless-Reading-Display-G>
307: Internal (browser cached) redirect to https://www.amazon.com/Kind
- The server has previously indicated this domain should always be accessed via HTTPS (HSTS Protocol). Chrome has cached this internally, and did not connect to any server for this redirect. Chrome reports this redirect as a "307 Internal Redirect" however this probably would have been a "301 Permanent redirect" originally. You can verify this by clearing your browser cache and visiting the original URL again.
- <https://www.amazon.com/Kindle-Wireless-Reading-Display-G>
200: HTTP/1.1 200

BACKGROUND SCRIPT

```
//URL path item  
url: pathItem.url,  
status: pathItem.status_line,  
redirect_type: pathItem.redirect_type,  
redirect_url: pathItem.redirect_url,  
meta_timer: pathItem.meta_timer
```

EVALUATION METRICS

1. Phish detection accuracy
2. Target detection ratio
3. Memory usage profiling
4. Addon rendering time
5. Temporal resilience accuracy

REFERENCES

1. S. Marchal, G. Armano, T. Gröndahl, K. Saari, N. Singh, N. Asokan, "Off-the-hook: An efficient and usable client-side phishing prevention application", IEEE Trans. Comput., vol. 66, no. 10, pp. 1717-1733, Oct. 2017.
2. A. K. Jain, B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list", EURASIP J. Inf. Secur., vol. 2016, no. 1, Dec. 2016.
3. G. Xiang, J. Hong, C. P. Rosé, L. Cranor, "CANTINA: A feature-rich machine learning framework for detecting phishing Web sites", ACM Trans. Inf. Syst. Secur., vol. 14, no. 2, 2011.
4. Implementation for the Usage of Google Safe Browsing APIs (v4), 2019, [online] Available: <https://github.com/google/safebrowsing>.
5. C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," in Proc. Netw. Distrib. Syst. Security Symp., 2010, pp. 1–14.