Splunking for Dorks

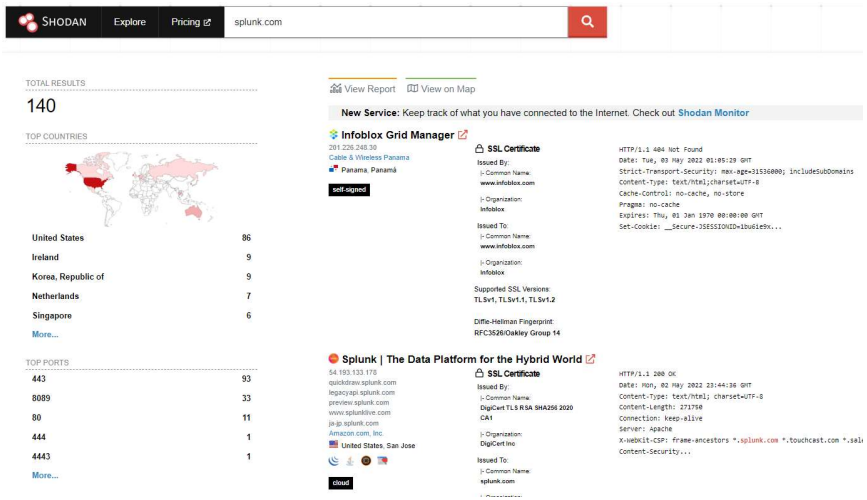
Greg Ford

github.com/gf13579

Overview

Attack surface monitoring

- What do you have exposed?
- What can people find out about you?



As defenders

- How can we identify changes in our surface?
- How can we find out what's publicly exposed?

How can we use Splunk for yet another thing?

splunk>

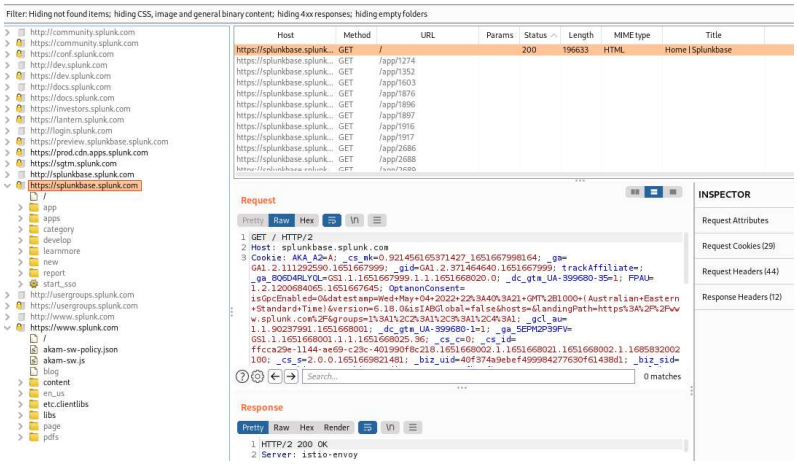
Background

Detection engineering + application security

AppSec: SAST, DAST, but also…

- Web app pen tests
- Bug bounty
- Real attacks

Narrow scope – great



Wide scope – will start with **recon**: what can we attack, and how?

splunk>

Defense

Reducing your attack surface is more effective than detection (but you need both)

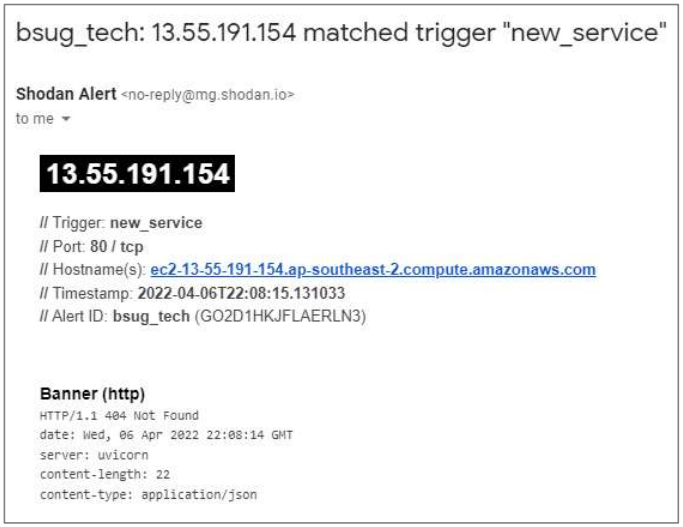
Knowing your attack surface – and changes to it – is hard

Be alerted to changes

- Close the port
- Harden the service
- Take down the content from a third party

Services exist – Shodan, Censys etc.

You can also DIY – port scan… or just google it



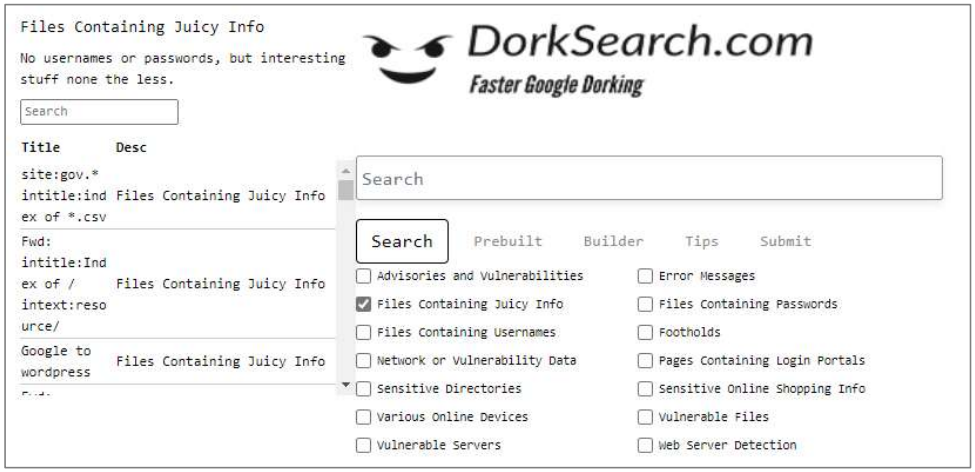
splunk>

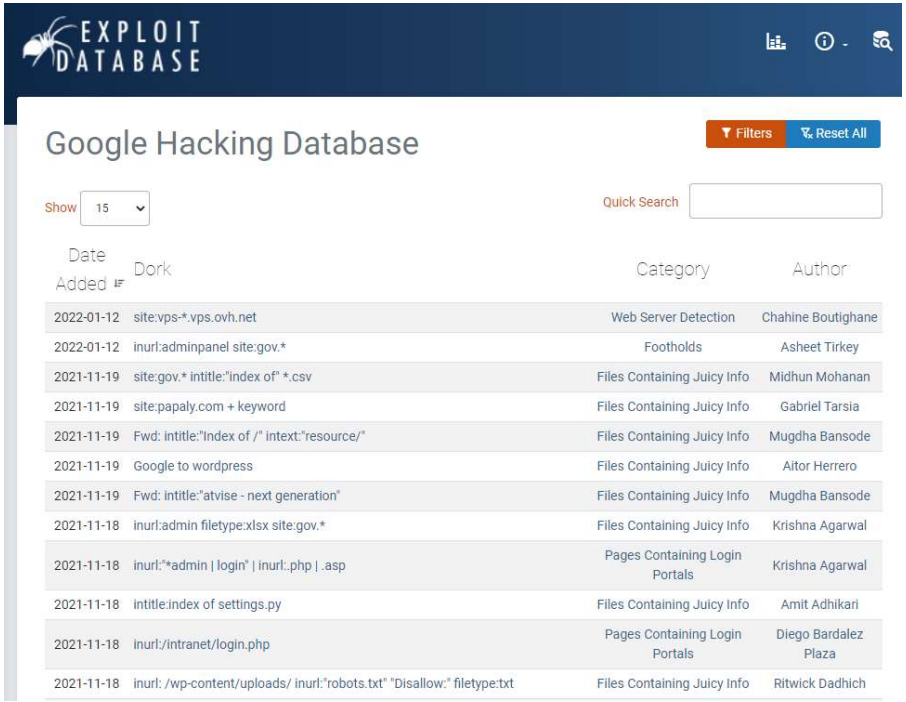
Google Dorks

```
site:example.com inurl:login
```

Widely known and documented:

- https://www.exploit-db.com/google-hacking-database
- https://dorksearch.com…





splunk>

Examples found during development

40k results for **`intitle:"index of /" site:*.gov.au "index of“`**

Customer code posted to stackoverflow and private GitHub repos by devs

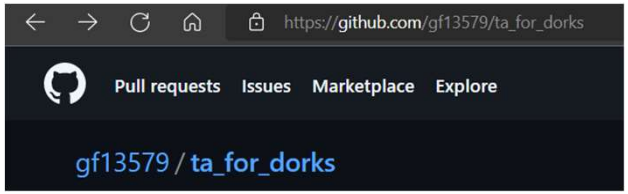
Customer mobile API Swagger definitions posted to codebeautify.org

Exposed **`hxxps://phpmyadmin.internal.redacted.qld.gov.au/`** referencing an open source VPN - doesn't seem very internal, despite subdomain

`test2.home.bigbank.com.au`, issuer name = Amazon, not-before = 2022-03-08

`hxxps://www.redacted.gov.au/DisplayAWebsite.aspx?WebsiteURL=hxxps://naughtywebsite.redacted/`

TA For Dorks



TA to run a bunch of google dorks (searches)

Support for limiting by date indexed i.e. only what's new

Extensible – add queries, extend templating schema

Subdomain discovery using crt.sh

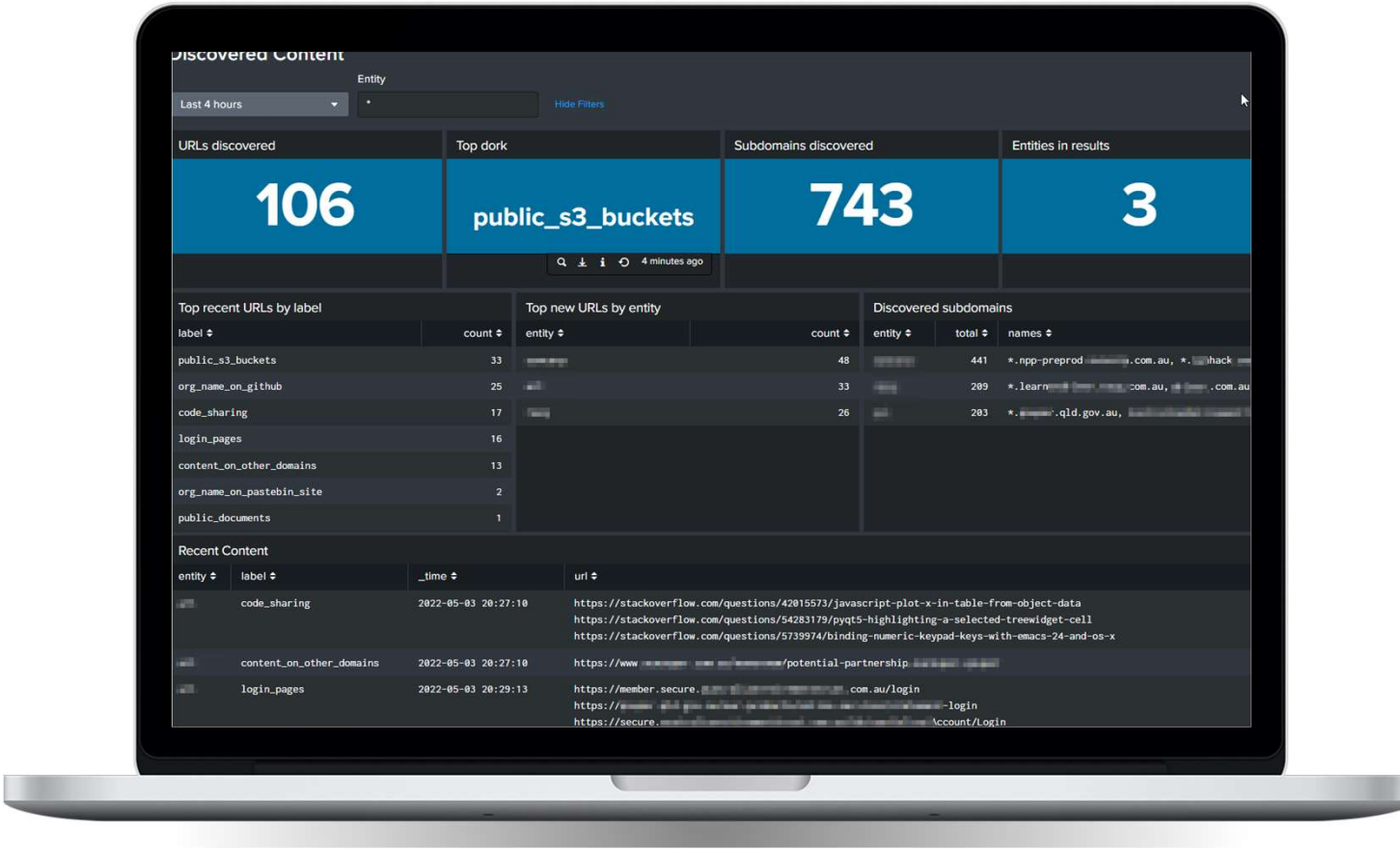
Dashboard and alerts – inc. correlation searches

AOB-free: mod input, simple setup page for keys, lookups for configuration

Demo time…

splunk>

Demo


Discovered Content
Entity
Last 4 hours
*
Hide Filters
URLs discovered
106
Top dork
public_s3_buckets
4 minutes ago
Subdomains discovered
743
Entities in results
3
Top recent URLs by label
label
count
public_s3_buckets 33
org_name_on_github 25
code_sharing 17
login_pages 16
content_on_other_domains 13
org_name_on_pastebin_site 2
public_documents 1
Top new URLs by entity
entity
count
48
33
26
Discovered subdomains
entity
total
names
441
209
203
Recent Content
entity
label
_time
url
code_sharing 2022-05-03 20:27:10
https://stackoverflow.com/questions/42015573/javascript-plot-x-in-table-from-object-data
https://stackoverflow.com/questions/54283179/pyqt5-highlighting-a-selected-treewidget-cell
https://stackoverflow.com/questions/5739974/binding-numeric-keypad-keys-with-emacs-24-and-os-x
content_on_other_domains 2022-05-03 20:27:10
https://www /potential-partnership
login_pages 2022-05-03 20:29:13
https://member.secure. .com.au/login
https:// -login
https://secure. \ccount/Login

splunk>

Configuration

Lookups / dork_entities.csv

Right-click the table for editing options

	entity	site	disabled	org_name	copyright_statement
1	example	example.com	0	Example Co.	Copyright Example Ltd ABN 66 010 123 456
2	big_bank	somebigbank.com.au	0	Intalock	Intalock Technologies Pty Ltd. All rights reserved
3	some_big_insurer	whatcouldgowrong.com.au	1	Big Insurers United	Copyright BBU trading as Some Big Insurer

Lookups / dork_queries.csv

Right-click the table for editing options

	service	query	label	disabled
1	crt.sh	[host]	subdomains	1
2	google_cse	site:{site} intitle:"index.of.\/"	open_folders	1
3	google_cse	site:{site} inurl:login	login_pages	1
4	google_cse	site:github.com "{org_name}"	org_name_on_github	1
5	google_cse	"{org_name}" (site:justpaste.it | site:heypasteit.com | site:pastebin.com)	org_name_on_pastebin_site	1
6	google_cse	site:{site} intitle:"default page"	default_pages	1
7	google_cse	site:.s3.amazonaws.com {org_name}	public_s3_buckets	1
8	google_cse	site:{site} -site:www.{site}	subdomain_content	1
9	google_cse	"{copyright_statement}" -site:{site}	content_on_other_domains	1
10	google_cse	"{org_name}" (site:trello.com | site:docs.google.com)	public_documents	1
11	google_cse	site:sharecode.io | site:controlc.com | site:codepad.co |site:ideone.com | site:codebeautify.org | site:jsdelivr.com | site:codeshare.io | site:codepen.io | site:repl.it | site:jsfiddle.net | site:stackoverflow.com "{org_name}"	code_sharing	1

splunk>

Engines

Google CSE

- Richest results, though we only really need title + description
- 100 queries/day then $5/1000, up to 10k/day
- Not as good as a normal Google search – missing some results

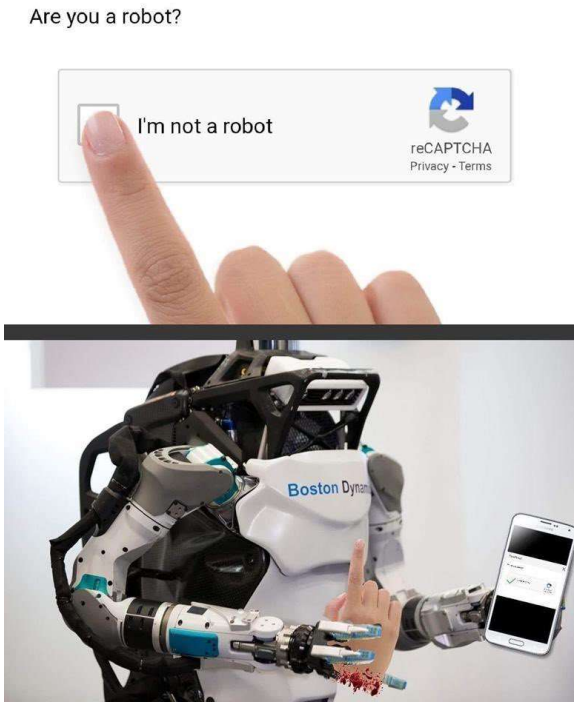
Google GET

- Best results
- Can rapidly hit Captcha tests, limiting reliability

DuckDuckGo

- Weaker results
- Lacking some search operators inc. **`inurl`**
- No Captchas, yet

GitHub…



splunk>

Closing Thoughts

Alert on changes in your externally-facing attack surface and public info

- From open RDP
- To insecure dev/test sites
- To a developer posting your code on stackoverflow

If nothing else, get Shodan/similar to alert you to newly exposed ports, services and subdomains. Investigate each one.

Consider dorking

- Tailor to your organisation
- Use robots.txt to limit exposure (with care)

Stop putting all the things on the Internet

splunk>

Questions

splunk> turn data into doing™