



# Alpha Presentation

Automotive Specific Dark Web Threat Intelligence

The Capstone Experience

Team GM

Anthony Ajlouny

Mitch Neelis

Kevin Kimmel

Ben Kelenski

Yuxin Li



*From Students...  
...to Professionals*

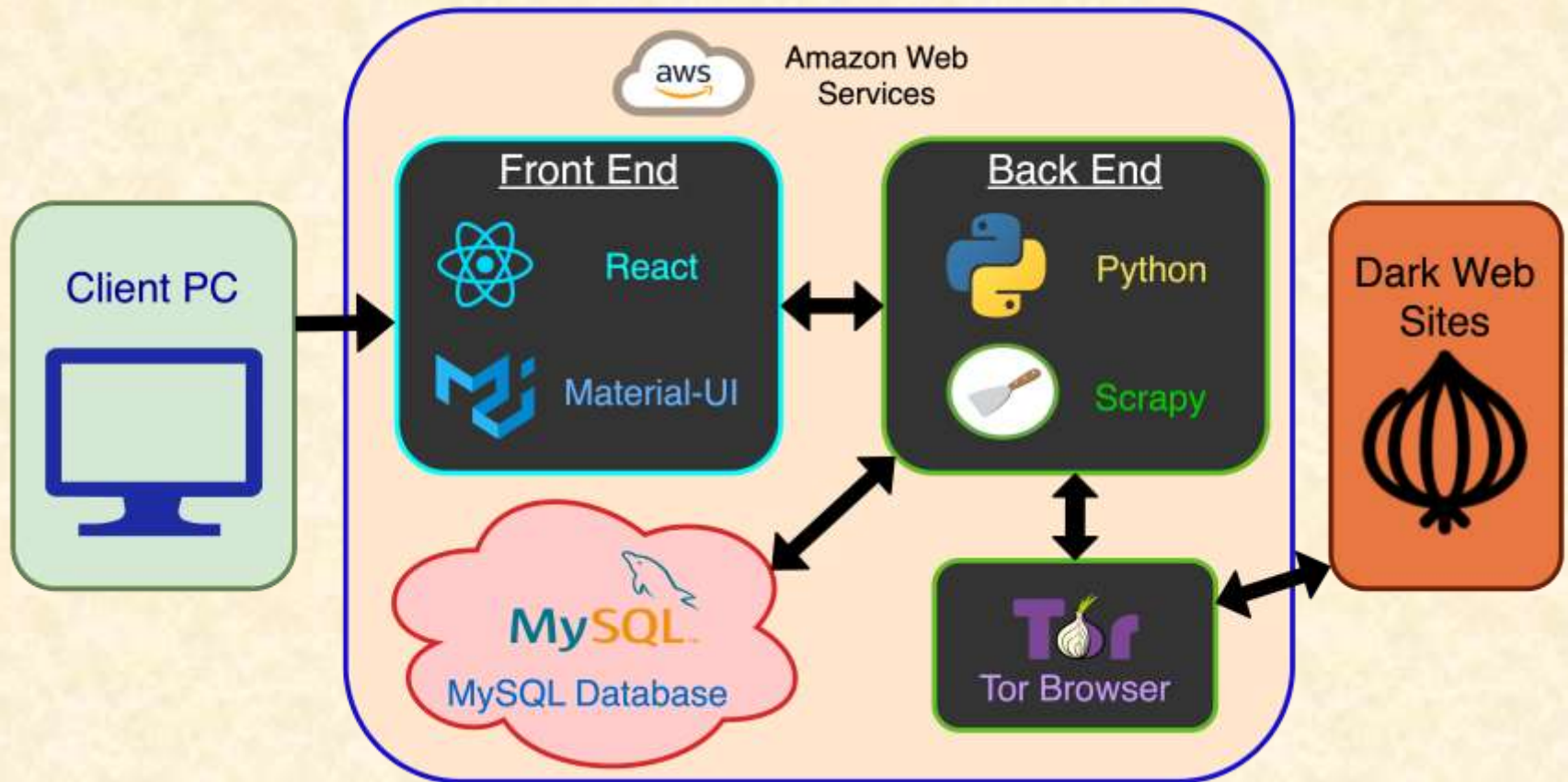
Department of Computer Science and Engineering  
Michigan State University  
Fall 2020

# Project Overview

---

- The program scrapes through the dark web looking for defined key words
- If found, the program takes a screenshot and saves page contents
- The contents run through our scoring system to check for severity of information
- This info is pushed to the database, where can be viewed and edited in the web app

# System Architecture



# Main Page

The screenshot shows a web browser window displaying the 'GM Dark Web Scraper' application. The application has a header with the GM logo and 'GENERAL MOTORS' text, and a navigation bar with 'HOME', 'DETAILS', and 'HEALTH' links. The main content area is titled 'GM Dark Web Scraper' and features a search bar and a table of results. The table has columns for 'Favorite', 'Site Info', 'Score', 'Hits', 'Date Found', and 'Details'. There are five rows of data, each with a star icon in the 'Favorite' column. The 'Details' column contains onion addresses. At the bottom of the table, there is a pagination control showing 'Rows per page: 5' and '1-5 of 5'. Below the table is an orange bar with a 'SIGN OUT' button.

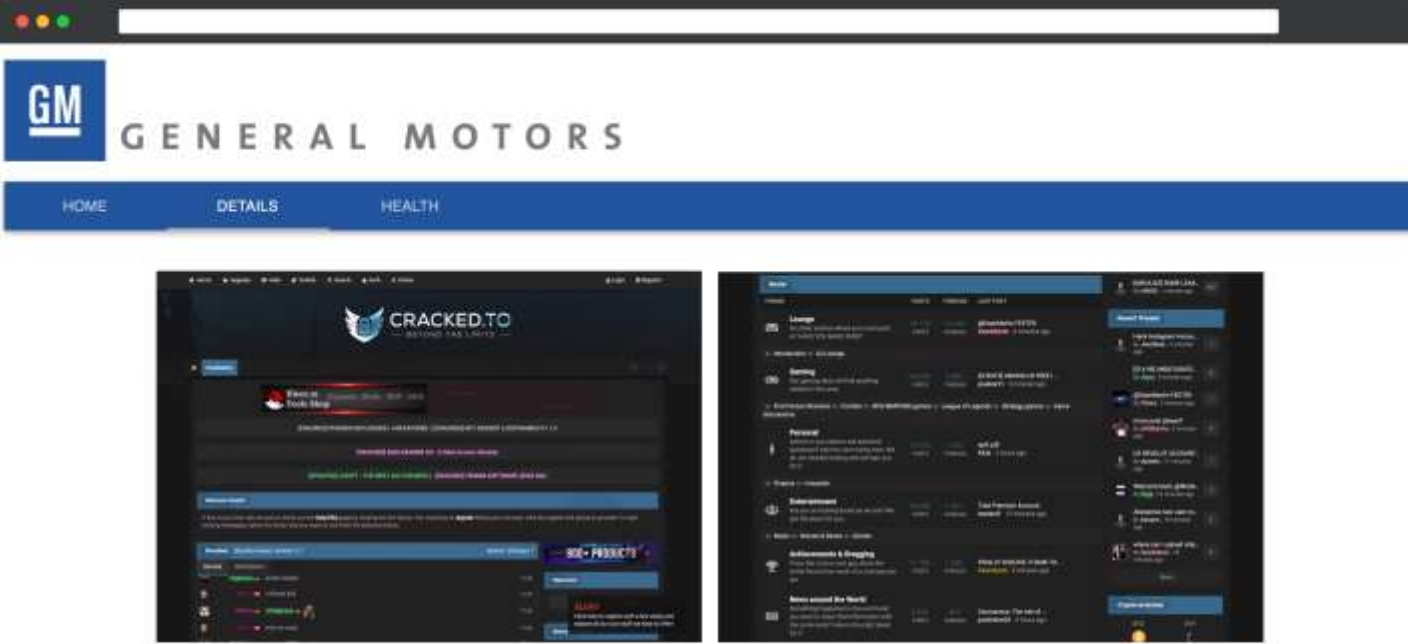
Favorite	Site Info	Score	Hits	Date Found	Details
☆	How to hack into GM?	10	100	2020-10-05 14:13:48	7Essfn22svo4vyf.onion
★	Database for General Motors	1	5	2020-10-05 14:13:48	ptcv1A75KjaOaBh.onion
☆	Job opportunities at General Motors	8	500	2020-10-05 14:13:48	nCSNB0o3r38qYESjO.onion
★	Nothing to report	0	0	2020-10-05 14:13:48	1BRgz9wbtHVpIBlu.onion
☆	Leaked footage of GM vehicle plans 2022?	8	75	2020-10-05 14:13:48	rsTB2djJBtY11BVBf.onion

Rows per page: 5 1-5 of 5

SIGN OUT



# Details Page



REMOVE IMAGE

URL	Hits	Score	Date Found
https://cracked.to	86	5	October 8, 2020 5:24:50 EST

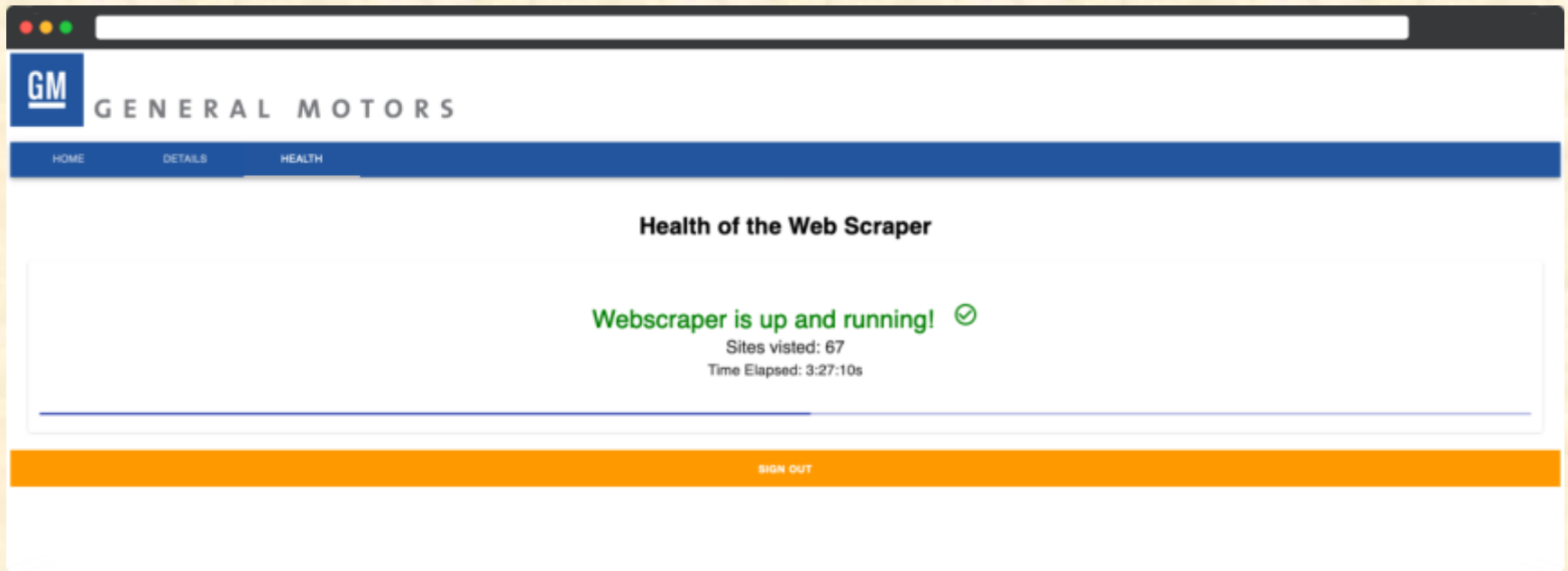
REMOVE ENTRY

current  
page: <https://cracked.to/Thread-Supreme-%E2%AD%90-2020-HOW-TO-HACK-ANY-ACCOUNT-Instagram-Facebook-YouTube-100-Working-%E2%AD%90?action=lastpost> ALERT  
here to register with a few steps and explore all our cool stuff we have to offer! Login Register Cracked.to Hacking & Exploits Hacking Tutorials SUPREME [2020] HOW TO HACK ANY

SAVE



# Scraper Health





# Collected Data Table

	ID	Score	URL	Terms found	Context	Date Found	Favorite
0	73	3	hxt254aygrsziejn.onion/_page_id=6	[breached, leak, leaked]	[from corporate networks that were breached an...	2020-10-18 13:20:05.186433	False
1	12	2	hxt254aygrsziejn.onion/_p=382	[breached, leak]	[Kong and Colombo. MAS holding was breached...	2020-10-18 13:20:05.186433	False
2	1	2	hxt254aygrsziejn.onion/_p=303	[breached, leak]	[Kong and Colombo. MAS holding was breached...	2020-10-18 13:20:05.186433	False
3	45	2	hxt254aygrsziejn.onion/_p=205	[breached, leak]	[Kong and Colombo. MAS holding was breached...	2020-10-18 13:20:05.186433	False
4	4	2	hxt254aygrsziejn.onion/_cat=9	[breached, leak]	[Part 1-2. MAS holding was breached and 300 gb...	2020-10-18 13:20:05.186433	False
5	53	2	hxt254aygrsziejn.onion/_p=166	[breached, leak]	[Part 1-2. MAS holding was breached and 300 gb...	2020-10-18 13:20:05.186433	False
6	109	2	hxt254aygrsziejn.onion/_cat=7	[leak, leaks]	[is the seventh part of leak on COSAN filelist...	2020-10-18 13:20:05.186433	False
7	55	2	hxt254aygrsziejn.onion/_p=427	[breach, leak]	[They allowed us to easily breach their networ...	2020-10-18 13:20:05.186433	False
8	65	2	hxt254aygrsziejn.onion/_p=136	[leak, leaks]	[is the third part of leak on COSAN  <a href...	2020-10-18 13:20:05.186433	False
9	75	1	hxt254aygrsziejn.onion/_author=1	[leak]	[--></header>\n\n</div><div class="entry-s...	2020-10-18 13:20:05.186433	False
10	70	1	hxt254aygrsziejn.onion/_p=346	[leak]	[the third part of the leak on Aliansce Sonae....	2020-10-18 13:20:05.186433	False
11	69	1	hxt254aygrsziejn.onion/_cat=26	[leak]	[.entry-meta --></div>\n</div><div class="entry-head...	2020-10-18 13:20:05.186433	False
12	67	1	hxt254aygrsziejn.onion/_p=142	[breached]	[Part 1-2. MAS holding was breached and 300 gb...	2020-10-18 13:20:05.186433	False
13	63	1	hxt254aygrsziejn.onion/_p=392	[leak]	[is a teaser for the leak on Indoco Remedies L...	2020-10-18 13:20:05.186433	False
14	0	1	hxt254aygrsziejn.onion/_p=362	[leak]	[leak. Second part of the leak includes compli...	2020-10-18 13:20:05.186433	False



# What's left to do?

---

- Implement web crawling code in AWS and connect it with the S3 bucket
- Format the incoming data to be neatly displayed on main page and details page
- Display stats on health screen
- Filtering by "favorite" option on web app





# Questions?

---

?

?

?

?

?

?

?

?

?

