Automating Threat Hunting on the Dark Web and other nitty-gritty things

\$whoami

- Apurv Singh Gautam (@ASG_Sc0rpi0n)
- Security Researcher, Threat Intel/Hunting
- Cybersecurity @ Georgia Tech
- Prior: Research Intern at ICSI, UC Berkeley
- O Hobbies
 - Contributing to the security community
 - Gaming/Streaming (Rainbow 6 Siege), Hiking, Lockpicking, etc.
- Social
 - Twitter @ASG_Sc0rpi0n
 - Website https://apurvsinghgautam.me

Agenda

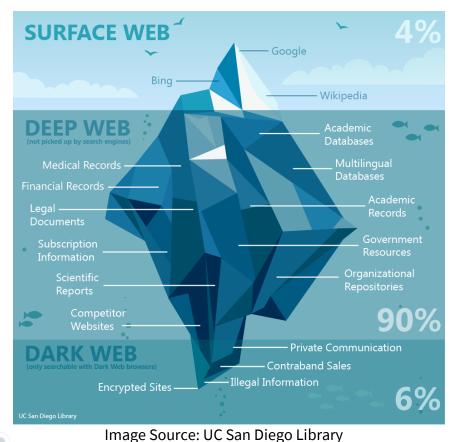
- Introduction to the Dark Web
- Why hunting on the Dark Web?
- Methods to hunt on the Dark Web
- Can the Dark Web hunting be automated?
- Process after hunting?
- OpSec? What's that?
- Conclusion





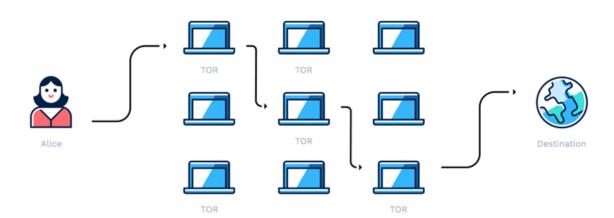


Clear Web? Deep Web? Dark Web?



Accessing the Dark Web

- Tor /I2P/ZeroNet
- onion domains/.i2p domains
- Traffic through relays







What's all the Hype?

- O Hype
 - Vast and mysterious part of the Internet
 - Place for cybercriminals only
 - Illegal to access the Dark Web
- Reality
 - Few reachable onion domains
 - Uptime isn't ideal
 - Useful for free expression in few countries
 - Popular sites like Facebook, NYTimes, etc.
 - Legal to access the Dark Web



Relevant sites?

- General Markets
- O PII & PHI
- Credit Cards
- Digital identities
- Information Trading
- Remote Access
- Personal Documents
- Electronic Wallets
- Insider Threats



Image Source: Intsights

Cost of products?

- O SSN \$1
- Fake FB with 15 friends \$1
- DDoS Service \$7/hr
- Rent a Hacker \$12/hr
- Credit Card \$20+
- Mobile Malware \$150
- Bank Details \$1000+
- Exploits or 0-days \$150,000+
- Critical databases \$300,000+

Product Listings







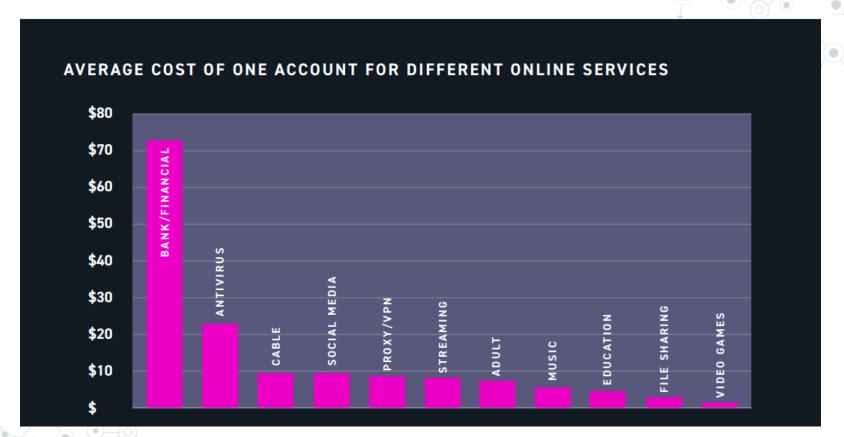


Image Source: Digital Shadows

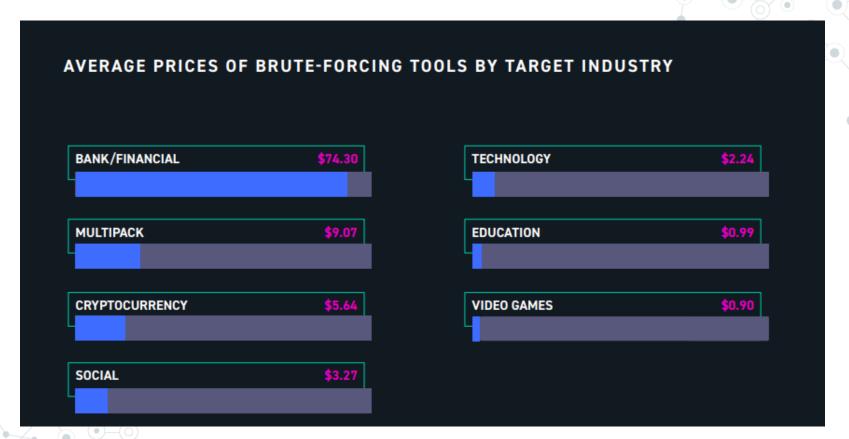


Image Source: Digital Shadows

Why hunting on the **Dark Web?**

What is Threat Hunting?

- Practice of proactively searching for cyber threats
- O Hypothesis-based approach
- Uses advanced analytics and machine learning investigations
- Proactive and iterative search



Why So Serious (Eh! Important)?

- Macker forums, darknet markets, dump shops, etc.
- Criminals can learn, monetize, trade, and communicate
- Oldentification of compromised assets
- Can potentially identify attacks in earlier stages
- Direct impacts PII (Personal Info), financial, EHRs (healthcare records), trade secrets
- Indirect impacts reputation, revenue loss, legal penalties

Benefits of Threat Hunting

- Keep up with the latest trends of attacks
- Prepare SOCs/Incident Responders
- Get knowledge of TTPs (Tactics, Techniques, Procedures) to be used
- Reduce damage and risks to the organization



the Dark Web

Methods to hunt on

Tools

- Open Python
- Scrapy
- Tor
- OnionScan
- Privoxy
- and many more...





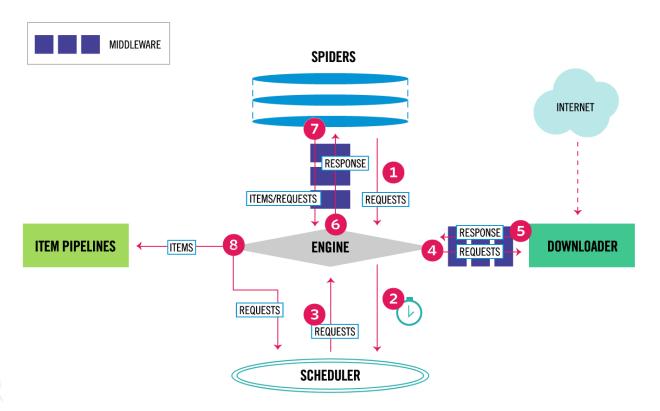








How Scrapy Works?



HUMINT

- O Human Intelligence
- Most dangerous and difficult form
- Most valuable source
- Infiltrating forums, markets, etc.
- Become one of them
- How threat actors think
- Can be very risky
- Time consuming

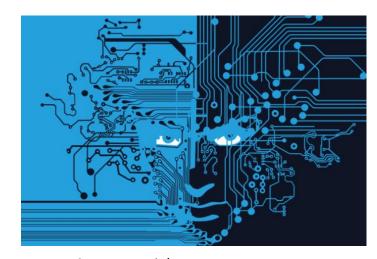


Image Source: Intsights

Can dark web hunting be automated?

Setting up TH Lab

- Cab/VM
- Physical or Cloud
- Isolate the network
- Install relevant tools
 - Scrapy
 - Privoxy
 - Tor
 - ELK
 - Python libraries



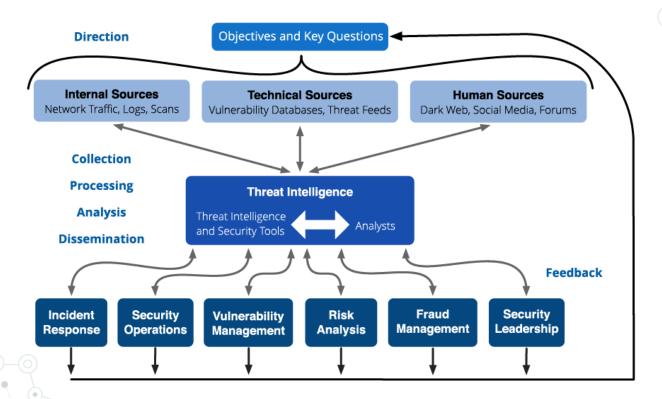
Image Source: Hayden James

Automated Hunting Architecture DESIGN/TRAIN DARK WEB LINKS SCRAPY SETUP CRAWLER PARSER ANALYZER SOCKS PROXIES ELK

AUTOMATIC

Process after hunting

Let's talk about TI Lifecycle



Threat Modelling

- "works to identify, communicate, and understand threats and mitigations within the context of protecting something of value" OWASP
- Define critical assets
- Understand what attackers want
- Threat actor capability and intent
- Sources to target

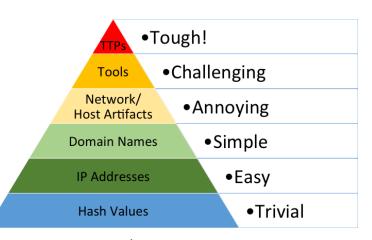


Image Source: David Bianco

Data Collection/Processing

- Collecting data from clear web
 - Pastebin
 - Twitter
 - Reddit
- Collecting data from dark web
 - Forums
 - Markets

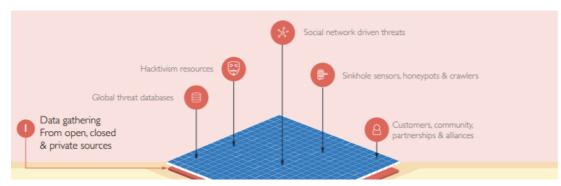


Image Source: Blueliv

Data Analysis

- NLP/ML/DL techniques
- Social network analysis
- Classification
- O Clustering
- MITRE ATT&CK





Image Sources: DataCamp, MITRE ATT&CK

OpSec? What is that?

What is OpSec?

- Actions taken to ensure that information leakage doesn't compromise you or your operations
- Derived from US military Operational Security
- PII Personally Identifiable Information
- Not just a process a mindset
- OpSec is Hard



Maintaining OpSec in your lifestyle

- Hide your real identity
- Use VM/Lab or an isolated system
- Use Tor or Tor over VPN always
- Change Time zones
- Never talk about your work
- Maintain different persona
- Take extensive notes
- Use password manager



Conclusion



What we discussed so far?

- Little about the Dark Web
- Dark Web forums/marketplaces
- Dark Web threat hunting
- Scrapy
- O HUMINT
- Automating the Dark Web hunting
- Little about threat intelligence lifecycle
- A little about OpSec

I don't know how to conclude but...

- Dark Web threat hunting is hard but worth the effort
- Keep OpSec in mind
- Look at more than one resource
- Takes a lot of resources and team effort
- Usage of MITRE ATT&CK framework



Resources

- Blogs & White papers by Recorded Future
- White papers by IntSights
- O Blogs by Palo Alto's Unit 42
- O Blogs by CrowdStrike
- Blogs by CloudSEK
- White papers by Digital Shadows
- Darkweb Cyber Threat Intelligence Mining by Cambridge University Press

Thanks!

Any questions?

You can contact me at:

Twitter: @ASG_Sc0rpi0n

LinkedIn: /in/apurvsinghgautam

