
OSINT
COMBINE
Dark Web
The Other Side
www.osintcombine.com | info@osintcombine.com | @osintcombine

Who we are…

Background

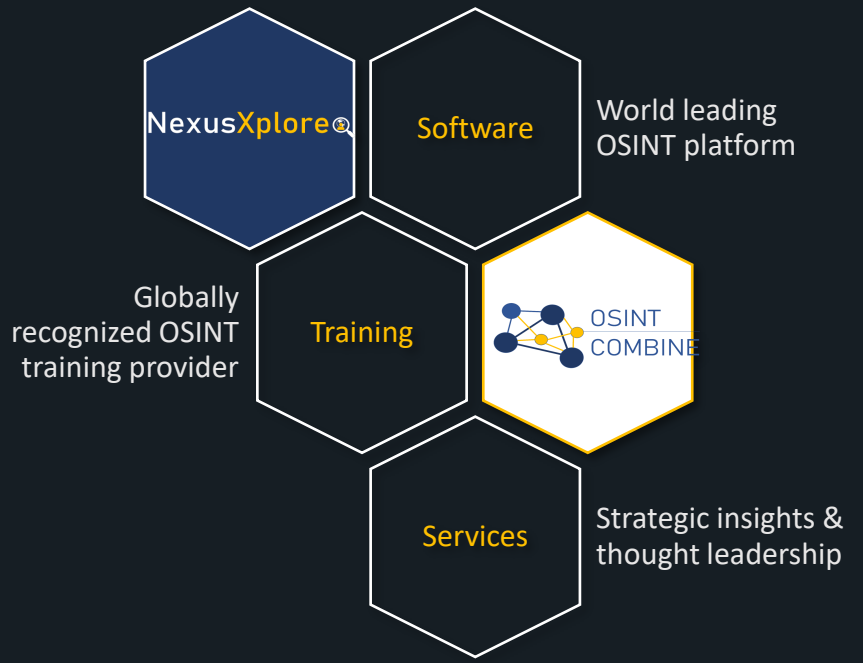
Australian veteran owned & operated business who specialize in open source intelligence (OSINT) training & software & services.

Successful delivery of advanced OSINT training & software globally to government organisations including national intelligence agencies, tri-service military, local & federal law enforcement through to private sector organisations including Fortune 500 & ASX 200 companies

We are the leading OSINT training & software provider in Australia & are internationally recognized as an OSINT leader

Mission

Develop enduring open source intelligence capability to strategically orientated organizations through partnership



Scope

- PART I
  - Dark Nets Overview
  - Understanding TOR
  - Attribution
  - Safe Access
- PART II
  - Exploring the Dark Web
  - Automating Collection with Scraping

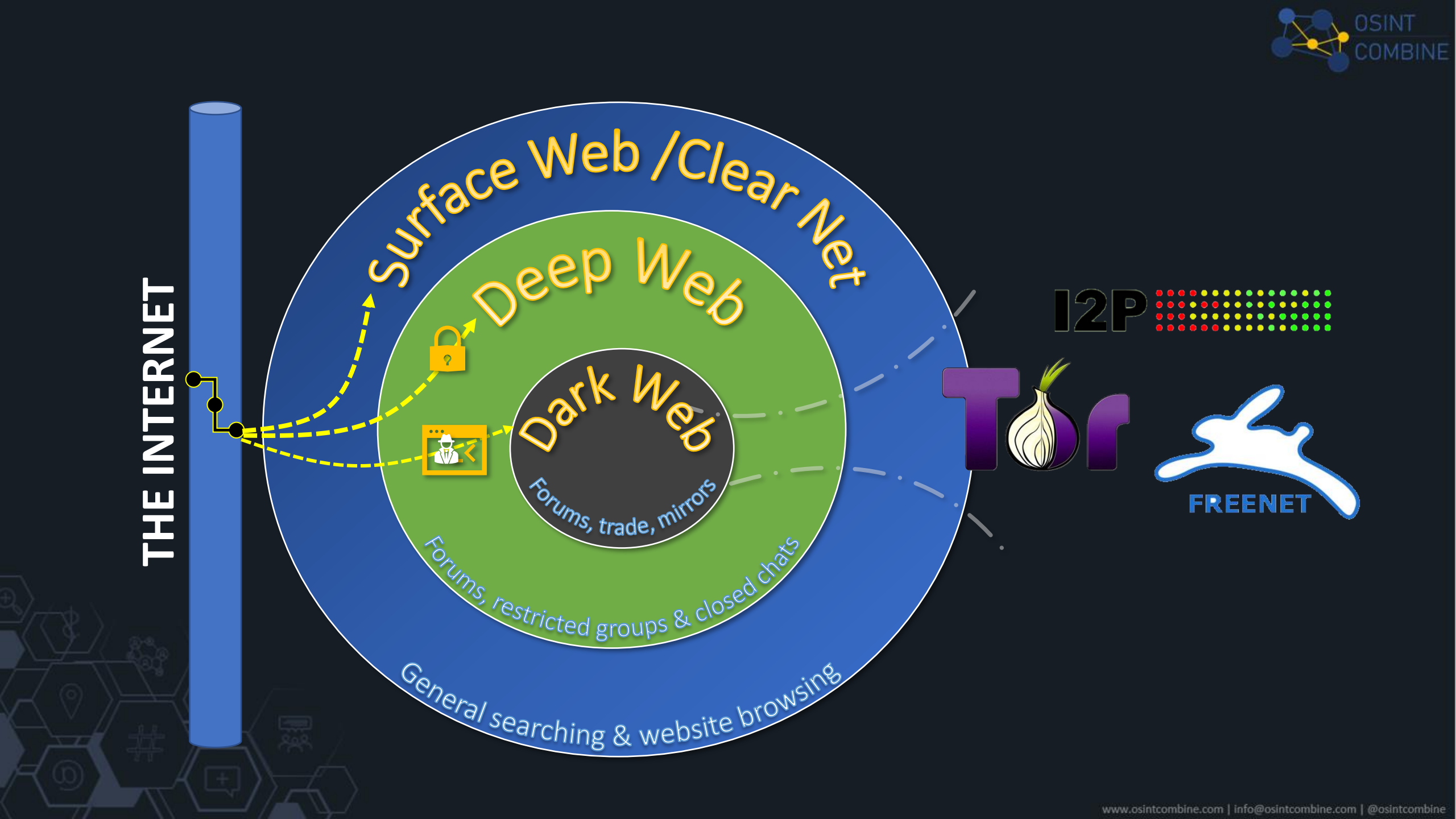
Takeaways

1. Build a deeper understanding of dark nets
2. Learn stronger attribution management techniques
3. Create some automation for TOR collection

Part I

TOR Network & Attribution

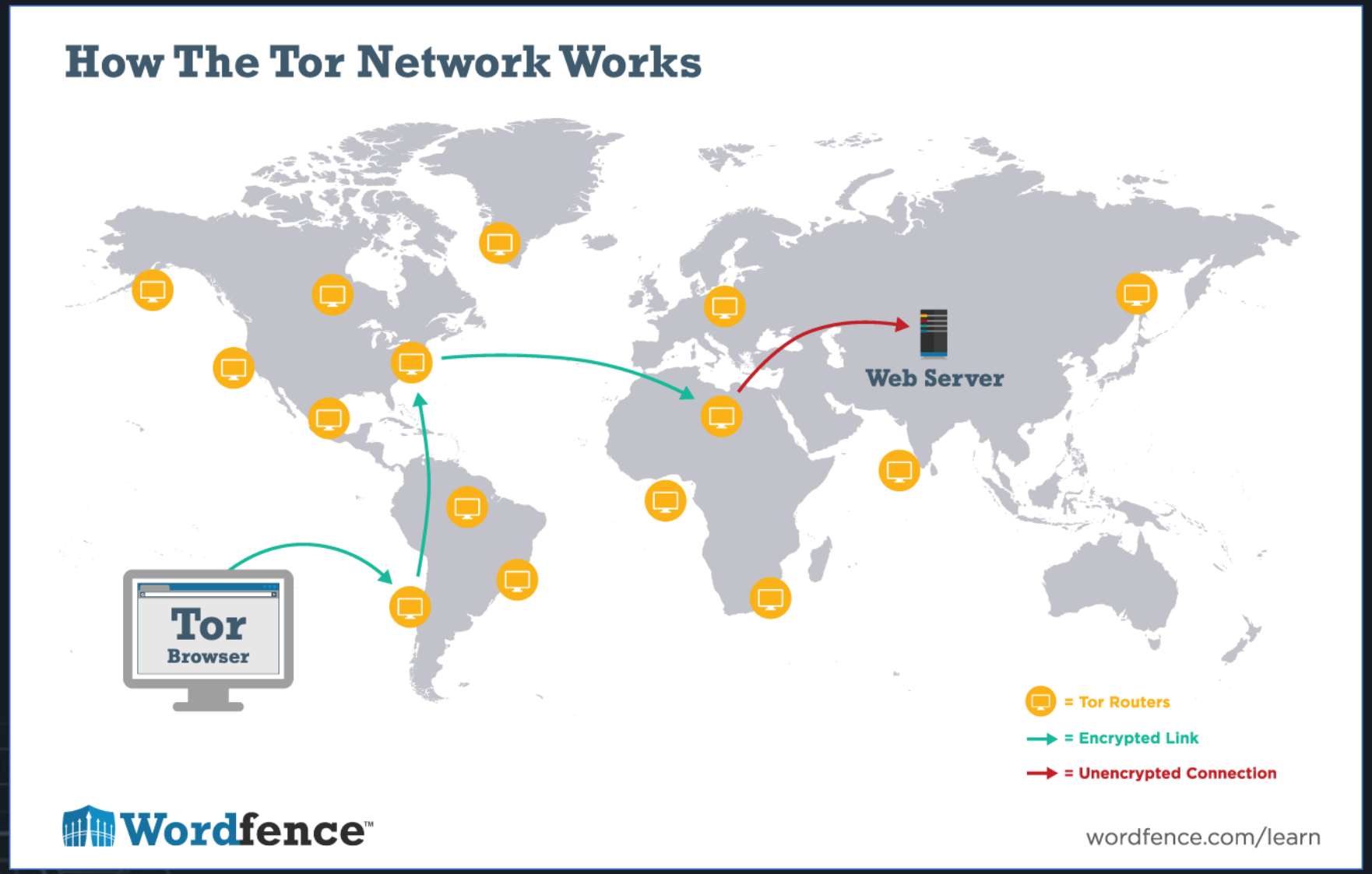
Dark Nets Overview


OSINT COMBINE
THE INTERNET
Surface Web /Clear Net
Deep Web
Dark Web
Forums, trade, mirrors
Forums, restricted groups & closed chats
General searching & website browsing
I2P
Tor
FREENET
www.osintcombine.com | info@osintcombine.com | @osintcombine

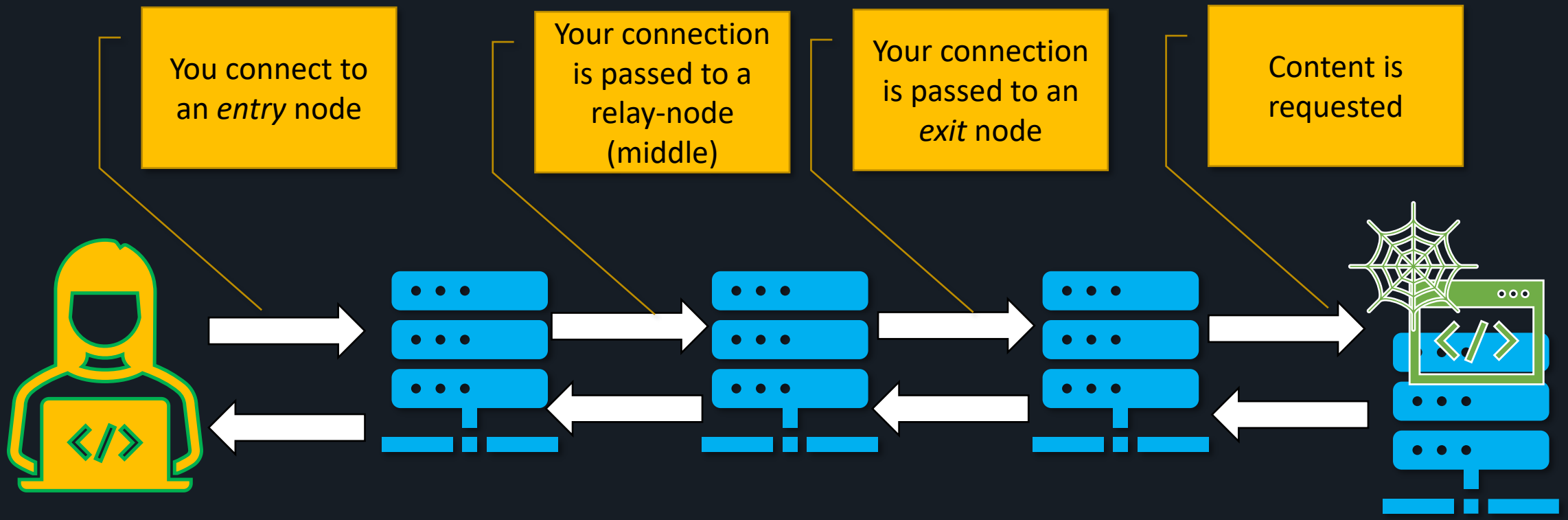
Understanding TOR

Let’s get comfortable…with TOR

TOR Network Design



Your Connection



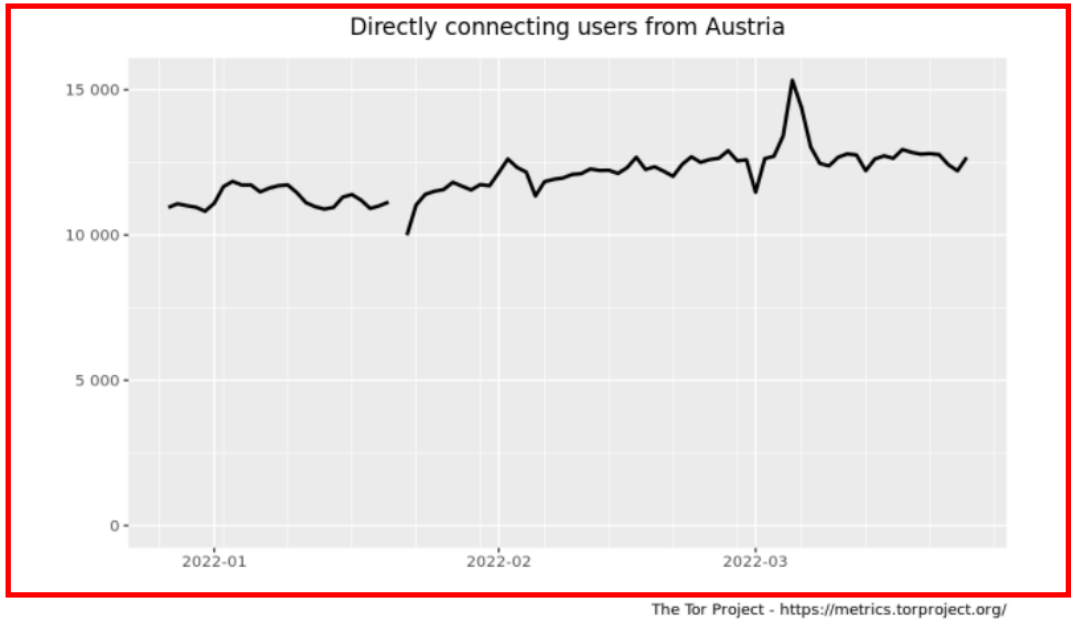
Users

We estimate the number of users by analyzing the requests induced by clients to relays and bridges.

Relay users | Bridge users by country | Bridge users by transport | Bridge users by country and transport | Bridge users by IP version

BridgeDB requests by requested transport | BridgeDB requests by distributor | Top-10 countries by relay users

Top-10 countries by possible censorship events | Top-10 countries by bridge users | "The anonymous Internet"



Start date: 27/12/2021

End date: 27/03/2022

Source: Austria

Show possible censorship events if available: Off

Update graph

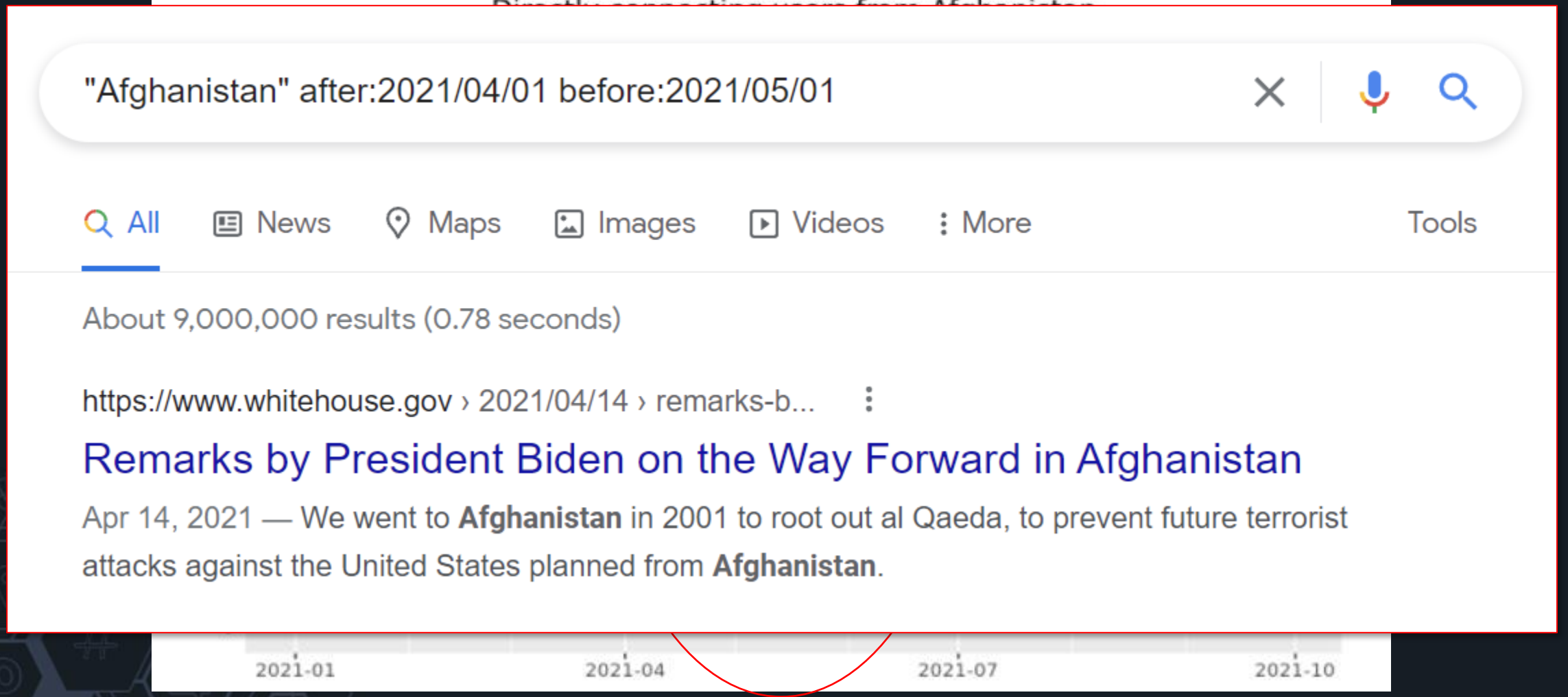
This graph shows the estimated number of directly-connecting clients; that is, it excludes clients connecting via bridges. These estimates are derived from the number of directory requests counted on directory authorities and mirrors. Relays resolve client IP addresses to country codes, so that graphs are available for most countries. Furthermore, it is possible to display indications of censorship events as obtained from an anomaly-based censorship-detection system (for more details, see this technical report). For further details see these questions and answers about user statistics.

Download graph as PNG or PDF.

Download data as CSV.

Learn more about the CSV data format or how to reproduce the graph data.

TOR Activity



TOR

- Sear
  - Fi
    p
  - P
    n
  - U
    in
    in

Relay Search

Simple Search | Aggregated Search | Advanced Search

The advanced search tool allows you to build advanced queries to search for data about single relays and bridges in the Tor network or aggregated data about currently running relays. For single relays, it provides useful information on how relays are configured along with graphs about their past. Aggregated data provides insight into diversity in the network and the probabilities of using relays in a particular country or AS as a guard, middle or exit relay filtered by the search parameters.

Nickname:

Hostname:

Fingerprint:

Include Family

First Seen:

From 0 to 0 days ago

Contact:

Last Seen:

From 0 to 0 days ago

Flag:

Any

Version:

Country:

Any

Type†:

Any

Autonomous System:

AS

Running†:

Any

Advanced Search

Advanced Aggregation | by AS | by CC | by AS+CC | by Version

† These options are ignored when performing an aggregation as the aggregated searches are restricted to only currently running relays.

g/rs.html

Incidents [edit]

Hacking [edit]

In June 2013, Hetzner Online suffered from a security breach where customer information was exposed to attackers who had compromised Hetzner Online's monitoring systems.[28][29]

Russian complaints about Glavcom.ua [edit]

In early August 2014, the Russian Federal Service for Supervision of Communications, Information Technology and Mass Media (Roskomnadzor) sent a demand to many news agencies, including BBC, prohibiting any mention of the demonstration that was being arranged in the Siberian city of Novosibirsk in support of the federalization of Siberia.[30][31] A number of such messages were sent to Ukraine, which was in the midst of undeclared war with Russian paramilitary in Donetsk region.[32] Since the Ukrainian online newspapers did not remove the article, Roskomnadzor sent letters to their Internet providers demanding removal of the news item. Hetzner Online complied with the demands and sent a notice to glavcom.ua, saying "Please solve the problem and reply within the next 24 hours to avoid suspension. This is the final deadline."[33]

This story was widely reprinted in news sources.[34] Ukrainian Ministry of Foreign Affairs issued a statement expressing solidarity with glavcom.ua owners and journalists. Vassily Zvarych, vice-head of Communications Department of Foreign Ministry, gave a press conference saying that he was surprised by Hetzner Online's compliance with the Russian complaint.[35] Also, a German chapter of Reporters Without Borders issued a statement condemning Roskomnadzor.[36]

The notices to suspend Glavcom.ua were issued by Hetzner Online August 6, 2014; on August 10 Hetzner Online issued apologies, denying that any censorship was planned and that their technical support made a wrong decision, which they regret.[37] However, by that time the story was widely published in German mass-media,[38] and Glavcom.ua already migrated from Hetzner Online to another hosting provider.[39][40]

no.spam.ee lawsuit [edit]

In 2013, an Estonian anti-spam activist Tõnu Samuel posted a blog entry about an alleged spammer Silver Teede on his website no.spam.ee. In retaliation, Teede wrote a complaint to the blog's service provider, Hetzner Online, who decided to terminate services for the blog. In an ensuing court case, Estonian courts found the complaints to be baseless and awarded Samuel damages from Silver Teede for the loss of Samuel's servers.[41][42][43]

Duplicate Ed25519 SSH keys [edit]

From April 2015 to December 2015, many of the OS images used by Hetzner's installation program installimage had used duplicate Ed25519 SSH keys. This could potentially mean that an attacker could use a Man-in-the-middle attack to compromise an SSH connection that was using Ed25519 keys. Hetzner sent an email to all affected customers with any potentially affected servers on information about the issue, and how to fix it. [44][45]

Blocking "Novaya Gazeta" [edit]

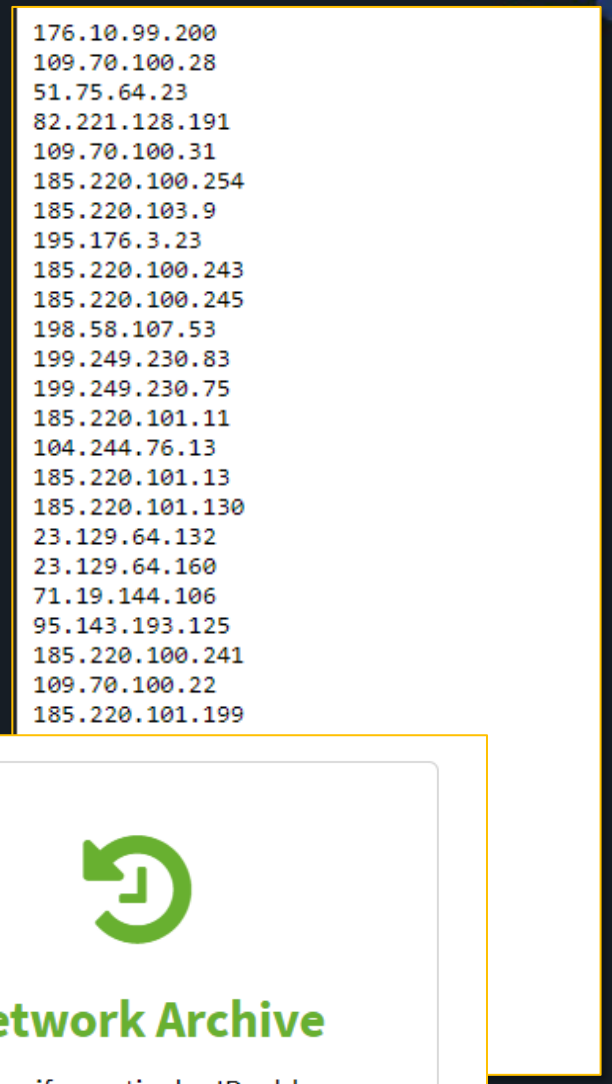
On January 11, 2016, Hetzner blocked the St. Petersburg site of Novaya Gazeta, the leading oppositional, non-governmental newspaper in Russia.[46] The newspaper marked the act as political censorship without any legal procedure.[47]

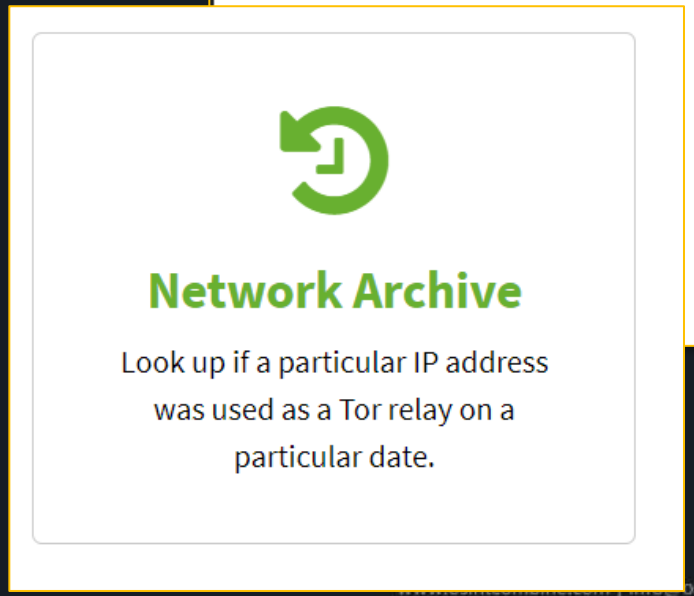
Blocking Ukraine War information [edit]

TOR Exit Nodes

- https://check.torproject.org/torbulkexitlist
- All active TOR exit nodes

- https://metrics.torproject.org/exonerator.html
- See if an IP was used as a TOR relay in the past (historical investigations)





176.10.99.200

21st March 2022

Summary

Result is negative

We did not find IP address 176.10.99.200 on or within a day of 2022-03-21.

21st January 2021

Summary

Result is positive

We found one or more Tor relays on IP address 176.10.99.200 on or within a day of 2021-01-21 that Tor clients were likely to know.

Technical details

Looking up IP address 176.10.99.200 on or within one day of 2021-01-21. Tor clients could have selected this or these Tor relays to build circuits.

Timestamp (UTC)	IP address(es)	Identity fingerprint	Nickname	Exit relay
2021-01-20 00:00:00	176.10.99.200, 176.10.99.207	0516085D6CAC40ED4CDCEFDFC5CCF6B00DE61DED	AccessNow007	Yes
2021-01-20 00:00:00	176.10.99.200, 176.10.99.203	2DFDEA5DD415B95594BFB12D59FE841167F94B5F	AccessNow003	Yes
2021-01-20 00:00:00	176.10.99.200, 176.10.99.201	3C5915348D731505C48112F4F03235FDE7B8C837	AccessNow001	Yes
2021-01-20 00:00:00	176.10.99.200	4273E6D162ED2717A1CF4207A254004CD3F5307B	AccessNow000	Yes
2021-01-20 00:00:00	176.10.99.200, 176.10.99.210	46F90EF3A3628C134DBB4654D0E4FF7EB914B690	AccessNow010	Yes
2021-01-20 00:00:00	176.10.99.200, 176.10.99.202	6290A2D08E5EB89C809223C5C7BF52597690751D	AccessNow002	Yes

Attribution

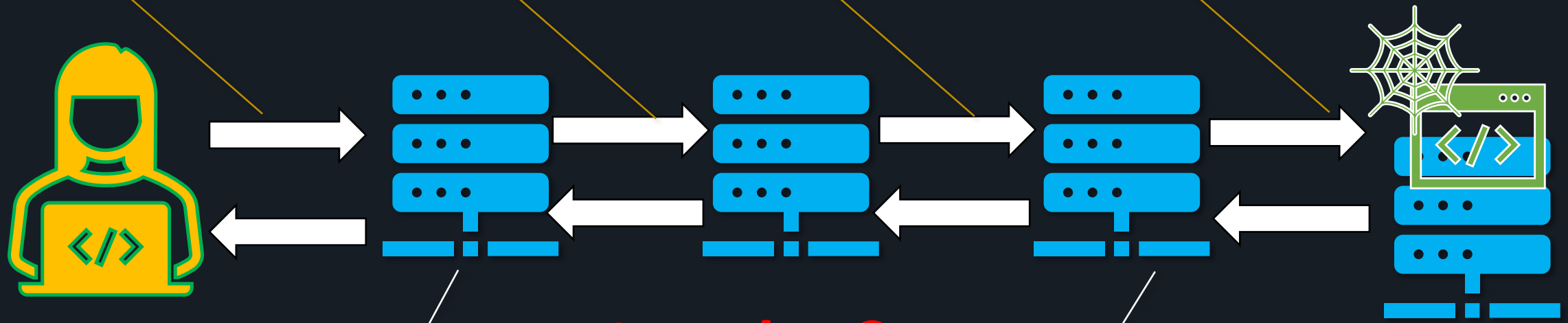
Your Connection

You connect to an *entry* node

Your connection is passed to a relay-node (middle)

Your connection is passed to an *exit* node

Content is requested

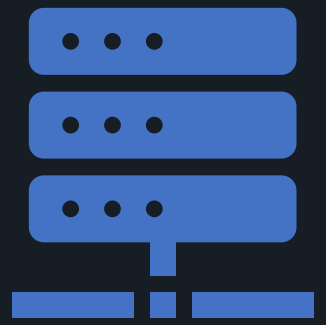


So What?

This server sees your *originating* IP

This server sees *what* your browsing/requesting

Types of Attribution



Network based attribution
(IPs, systems etc)

Client/user based attribution
(tradecraft, OSINT)

Network Attribution

- Attribution is *relative* to your requirements
- Sophisticated actors can run their own TOR entry guards & harvest & profile who connects through to the TOR network
  - *Saturate enough TOR entry nodes & you can potentially profile IP ranges associated with investigative body (government or other)*
- Hiding your connection to the network may be important. *Why?*
  - *Obfuscating that your agency is actively looking for Dark Net targets*
  - *Cursory protection of government network infrastructure addresses from bulk collection*
  - *Communicating in a censored environment & you're worried about state actors persecuting you*

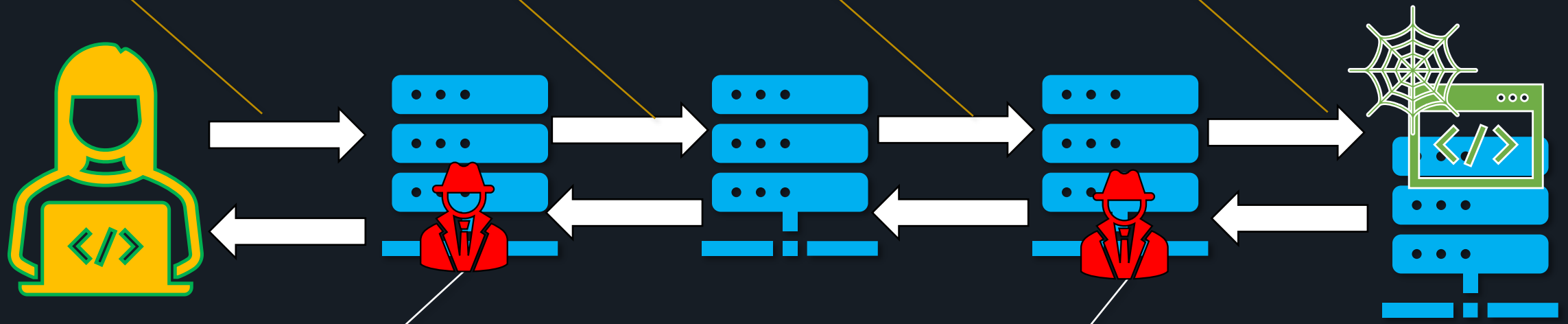
Hypothetical

You connect to an *entry* node

Your connection is passed to a relay-node (middle)

Your connection is passed to an *exit* node

Content is requested



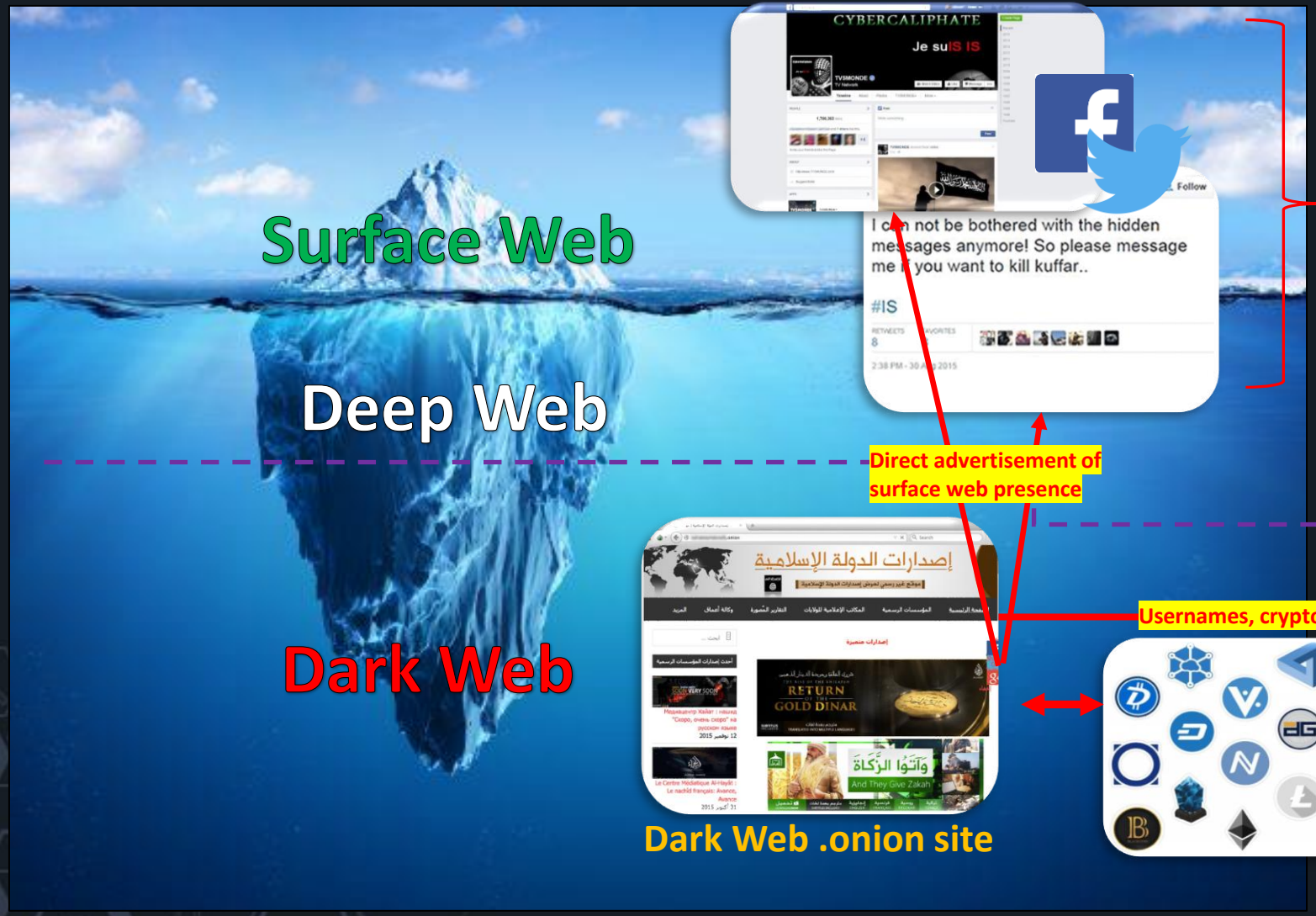
Threat actor runs this relay & harvests your originating IP, reverses it to your organisations infrastructure or you

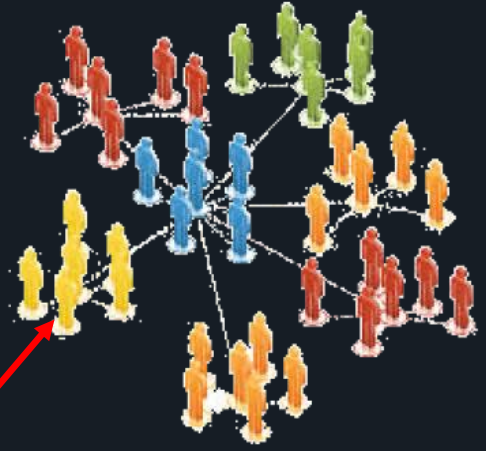
Threat actor runs the exist relay and harvests every request coming through, SSL stripping attacks etc

Client Attribution

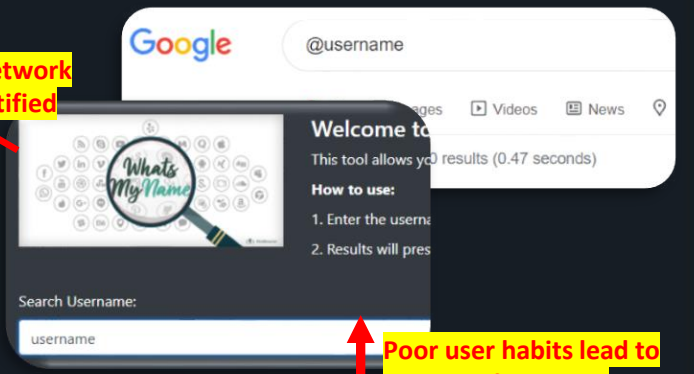
- Client attribution can be difficult but is achievable through:
  - Poorly configured TOR settings
    - HTML5 Canvas fingerprinting
    - Javascript embeds
  - Accessing personal services with an exit node watching
    - Facebook
    - Twitter
    - Email
- More importantly, attribution can occur through information slippage
  - This is where attributable information slips between the surface, deep and dark webs e.g., usernames, email addresses, twitter accounts etc

Information Slippage





Associated social network Presence from identified users

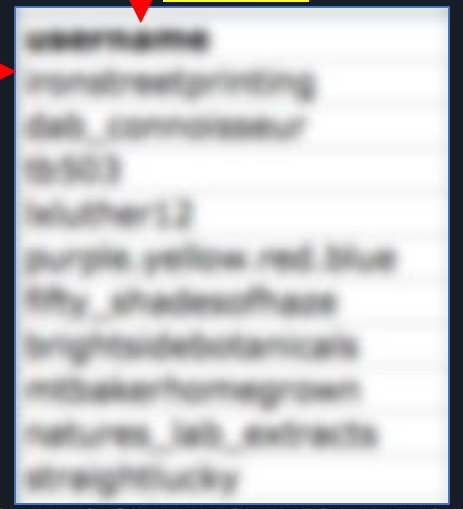


Direct advertisement of surface web presence

Poor user habits lead to shared information across Webs

Dark Web .onion site

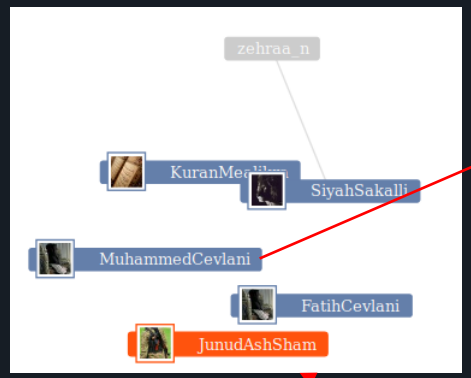
Usernames, crypto addresses & PGP keys

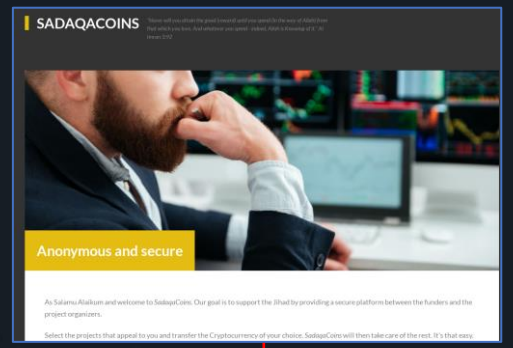


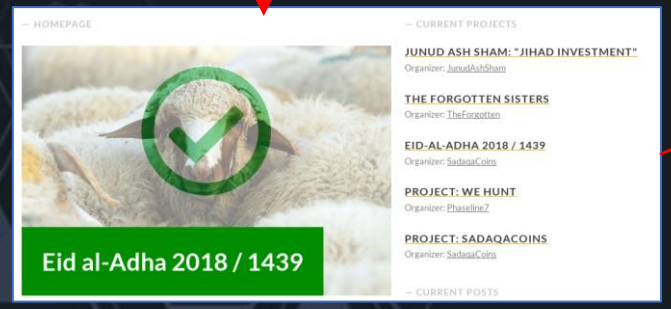
Example

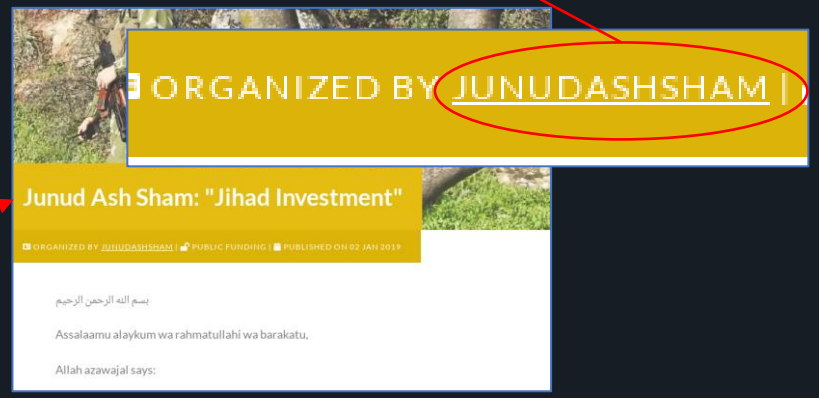
Original Article - Ben Strick:

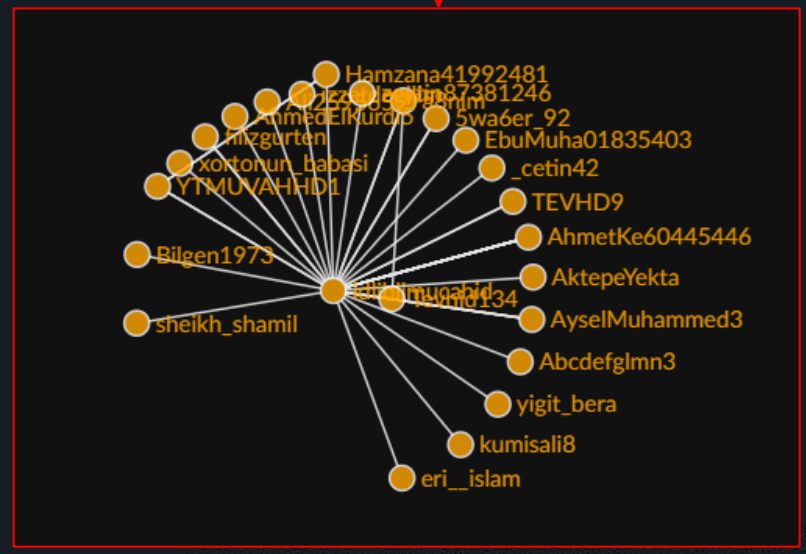
https://medium.com/@bendobrown/first-jihadi-cryptocurrency-crowdsource-platform-on-dark-web-263edf8885b7







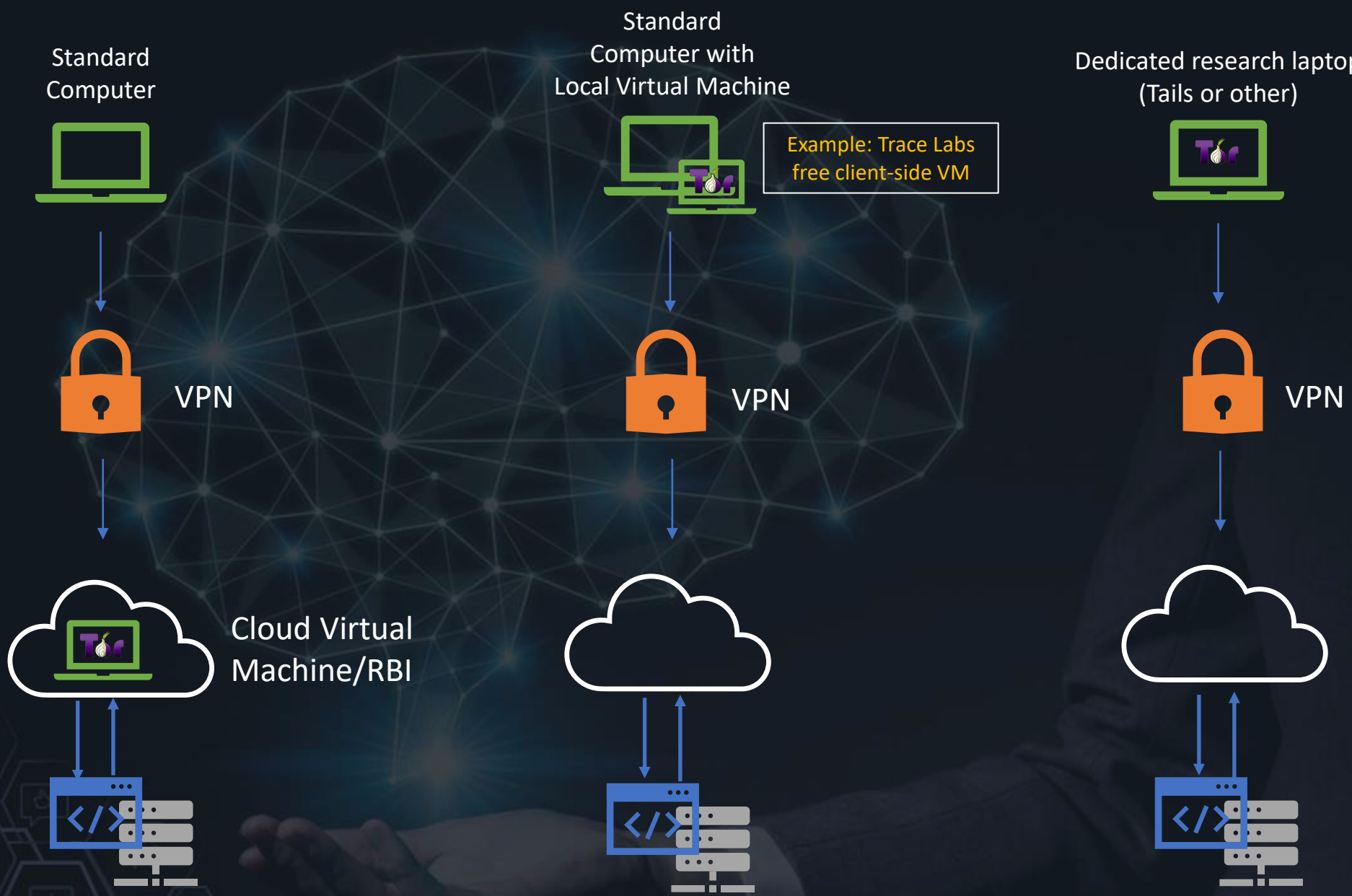





What can we do????

Tiered Protection

BASIC SAFE DARK WEB CONNECTION OPTIONS



ADVANCED DARK WEB CONNECTION OPTIONS

Don't bother as this level of control isn't required for you

Hide in the noise of other users by using a VPN or non-associated IP before connecting to TOR

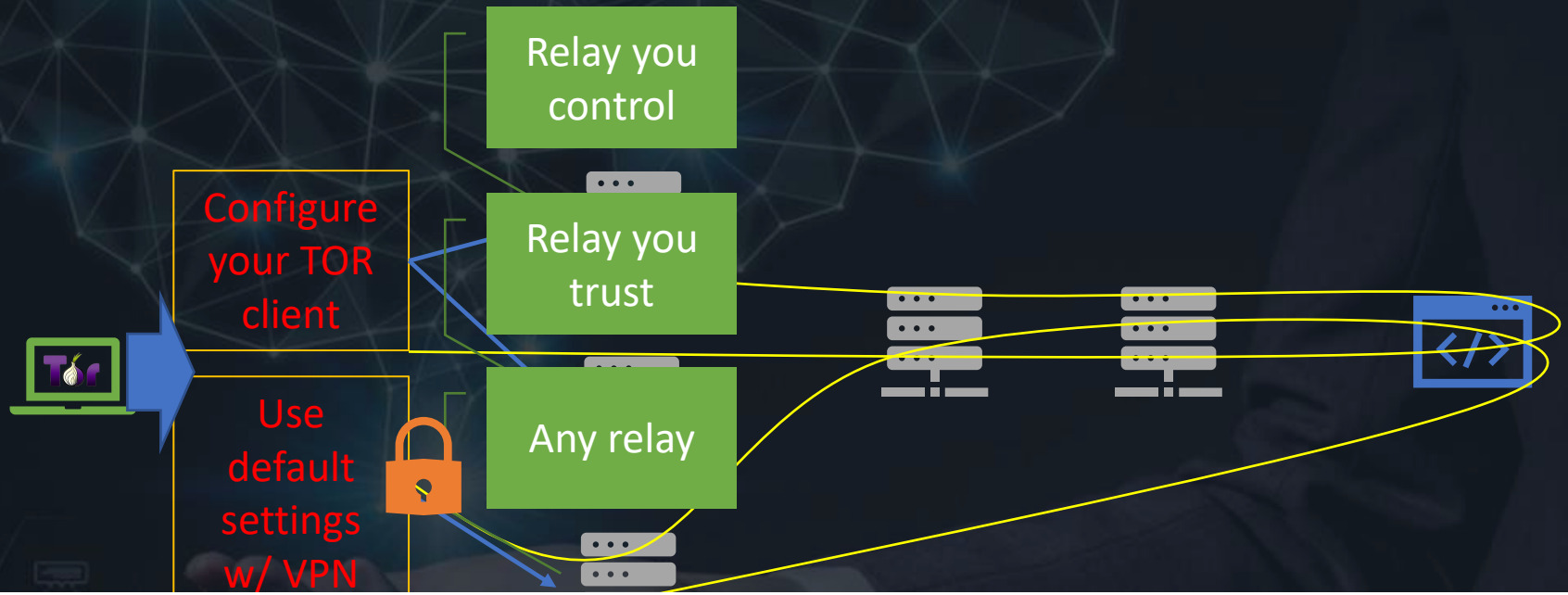
OR

Choose a TOR entry node you are comfortable with

OR

Set up your own TOR entry node (*lesson for another time*)

← EASIEST ACCESS — MOST CONTROL →



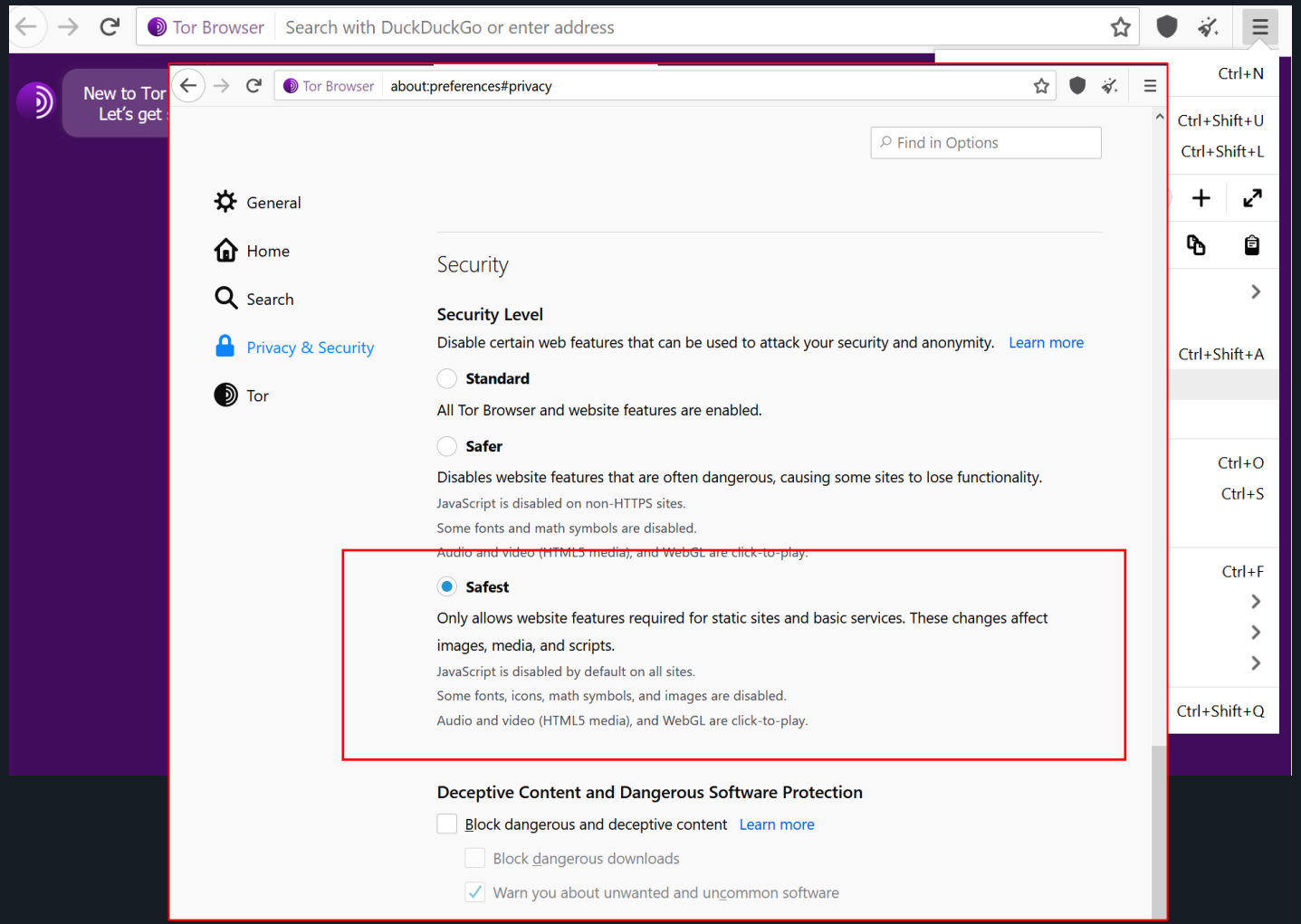
The same applies for exit nodes, but carries additional risk


OSINT
COMBINE
Client Configuration
www.osintcombine.com | info@osintcombine.com | @osintcombine

Installation

- Download TOR Browser
  - Firefox with TOR access pre-configured
- Install & configure client security

https://www.torproject.org/



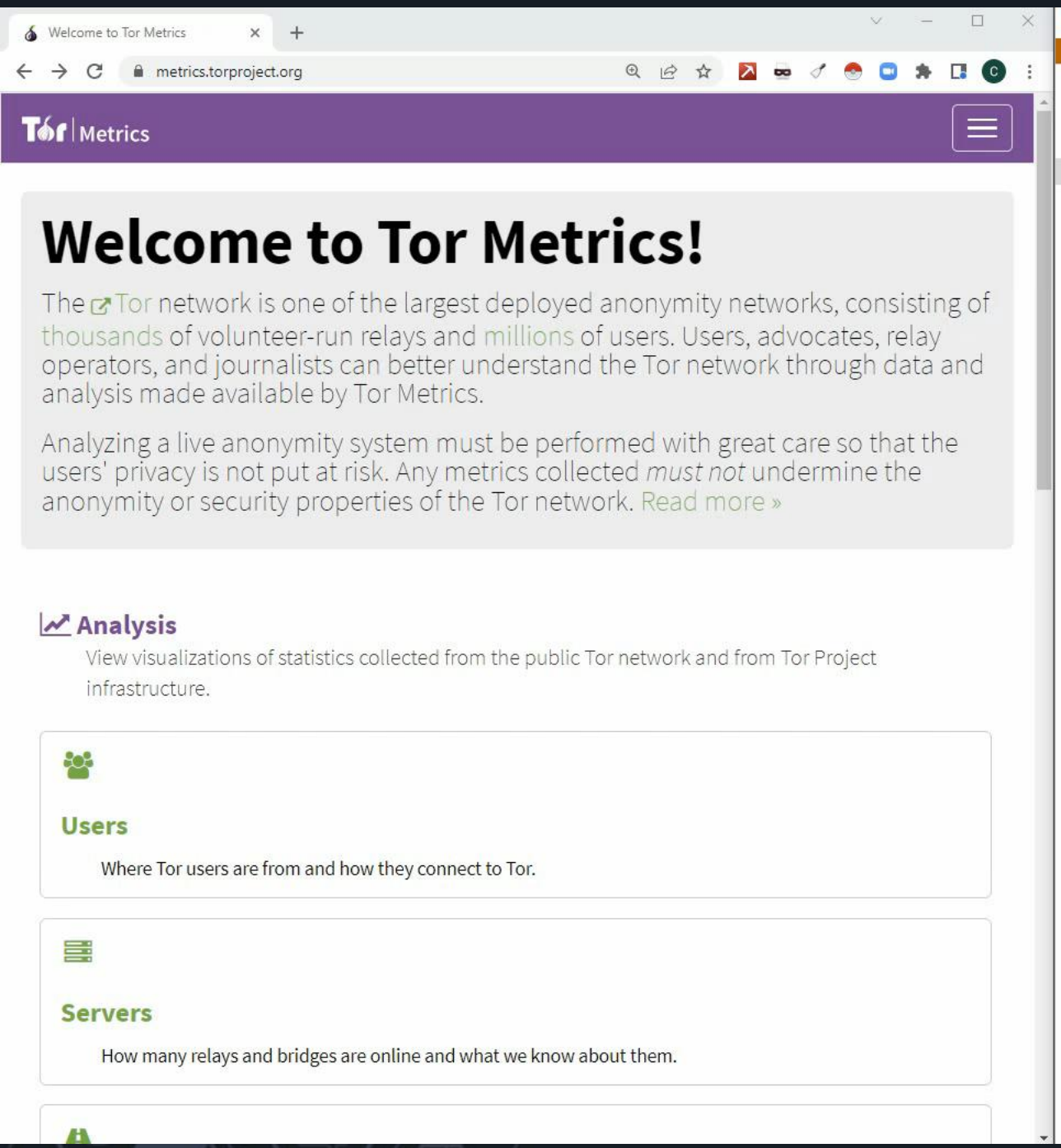
On Windows or Linux:

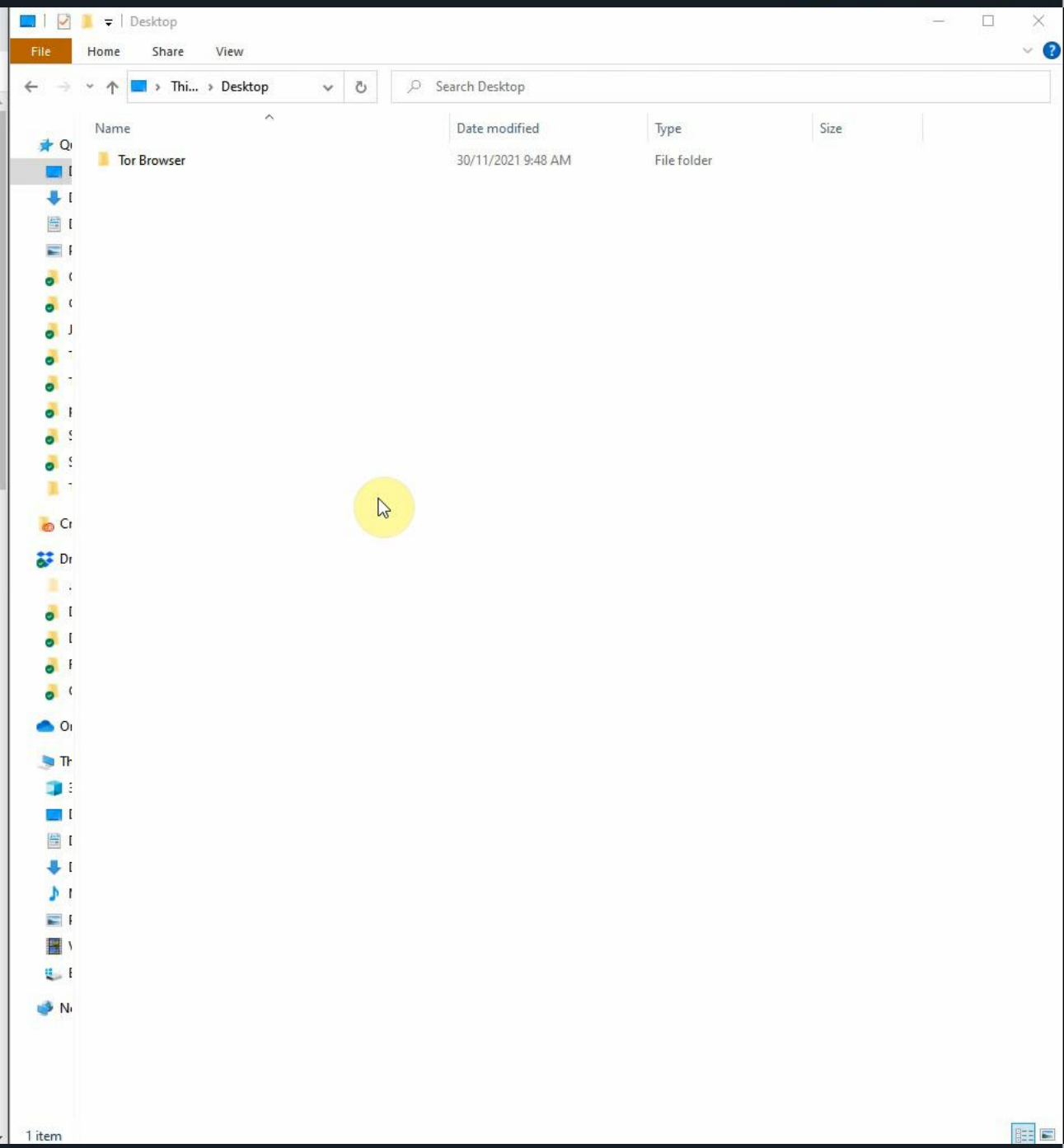
- The `torrc` is in the Tor Browser Data directory at `Browser/TorBrowser/Data/Tor` inside your Tor Browser directory.

On macOS:

- The `torrc` is in the Tor Browser Data directory at `~/Library/Application Support/TorBrowser-Data/Tor` .
- Note the Library folder is hidden on newer versions of macOS. To navigate to this folder in Finder, select "Go to Folder..." in the "Go" menu.
- Then type `~/Library/Application Support/` in the window and click Go.

https://support.torproject.org/tbb/tbb-editing-torrc/


Welcome to Tor Metrics
metrics.torproject.org
Tor Metrics
Welcome to Tor Metrics!
The Tor network is one of the largest deployed anonymity networks, consisting of thousands of volunteer-run relays and millions of users. Users, advocates, relay operators, and journalists can better understand the Tor network through data and analysis made available by Tor Metrics.
Analyzing a live anonymity system must be performed with great care so that the users' privacy is not put at risk. Any metrics collected must not undermine the anonymity or security properties of the Tor network. Read more »
Analysis
View visualizations of statistics collected from the public Tor network and from Tor Project infrastructure.
Users
Where Tor users are from and how they connect to Tor.
Servers
How many relays and bridges are online and what we know about them.


Desktop
File
Home
Share
View
Thi... › Desktop
Search Desktop
Name
Date modified
Type
Size
Tor Browser
30/11/2021 9:48 AM
File folder
1 item

Part II

Scraping & Automation

Exploring the Dark Web

Moving to V3

- You can identify v3 onion addresses by their 56 character length

  Tor Project's v2 address:
  http://expyuzz4wqqyqhjn.onion/,

  Tor Project's v3 address:
  http://2gzyxa5ihm7nsggfxnu52rck2vv4rvmdlkiu3zzui5du4xyclen53wid.onion/

- v2 onion services will not be reachable after September 2021.

Bookmarks are your friend, if you are on the Dark Web regularly. Otherwise, create your own list of .onion links & mirrors for efficiency.

TOR Services Lists

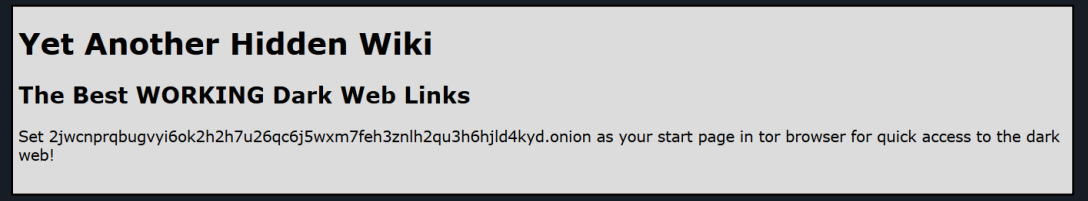
Surface Web

- Real-time mirror checks
  - https://onion.live/
- The Hidden Wiki is a Darknet directory
  - https://thehiddenwiki.org
  - https://thehiddenwiki.com/

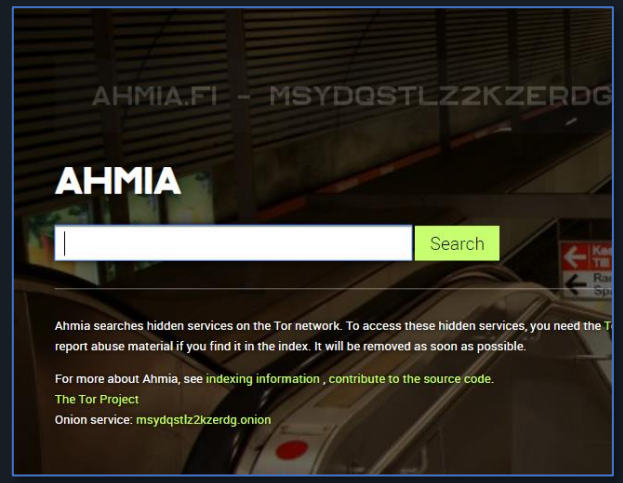
Dark Web

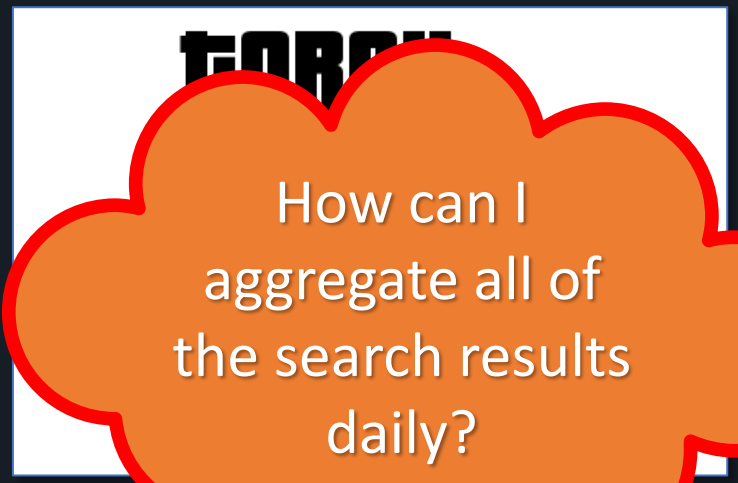
- Onion Links (Hidden n Links, Yet Another Hidden Wik
  - http://s4k4ceiapwwg igecpo7kv dnfr5gg7sph7jjppqkv on/
  - http://paavlaytlfsqyvk mg5jw2jdg2fgk za5ruf6lplwseeqtvyd.

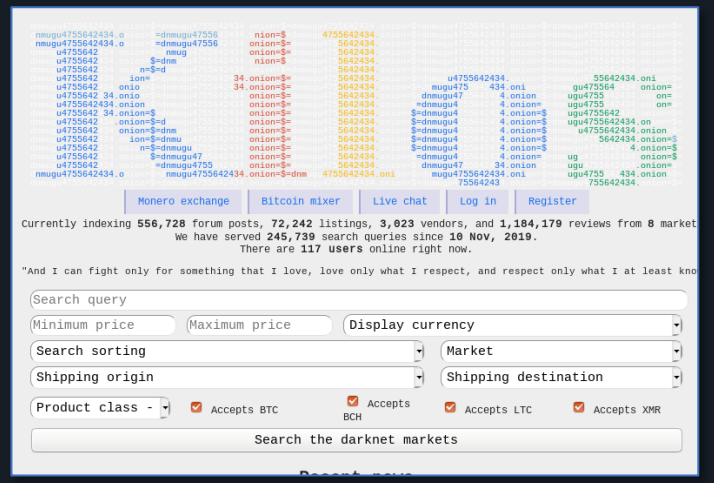
How can I keep an eye on changes to these lists?



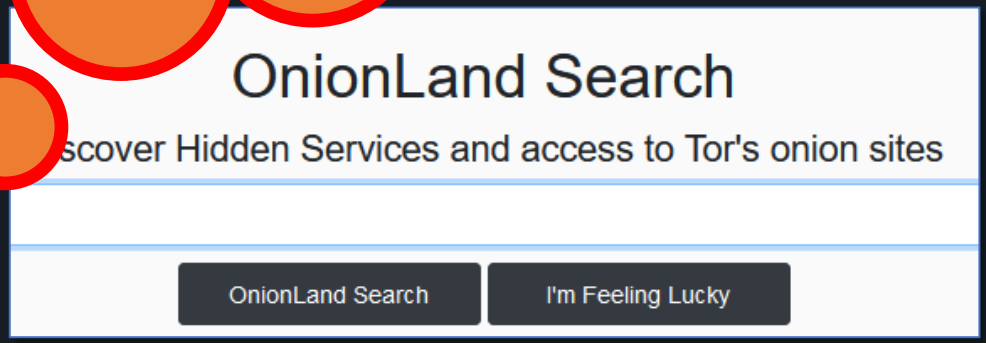
Free TOR Search Engines

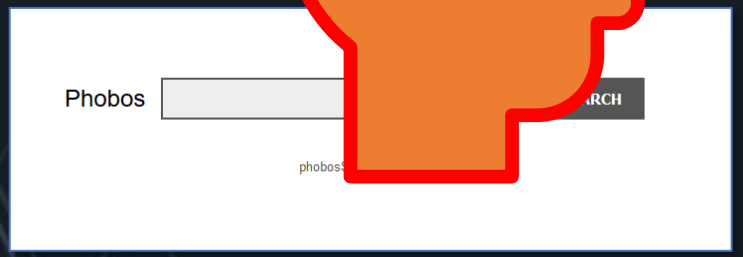






How can I aggregate all of the search results daily?





Monitoring & Scraping

Scraping Approaches

1

Scrape search engine results

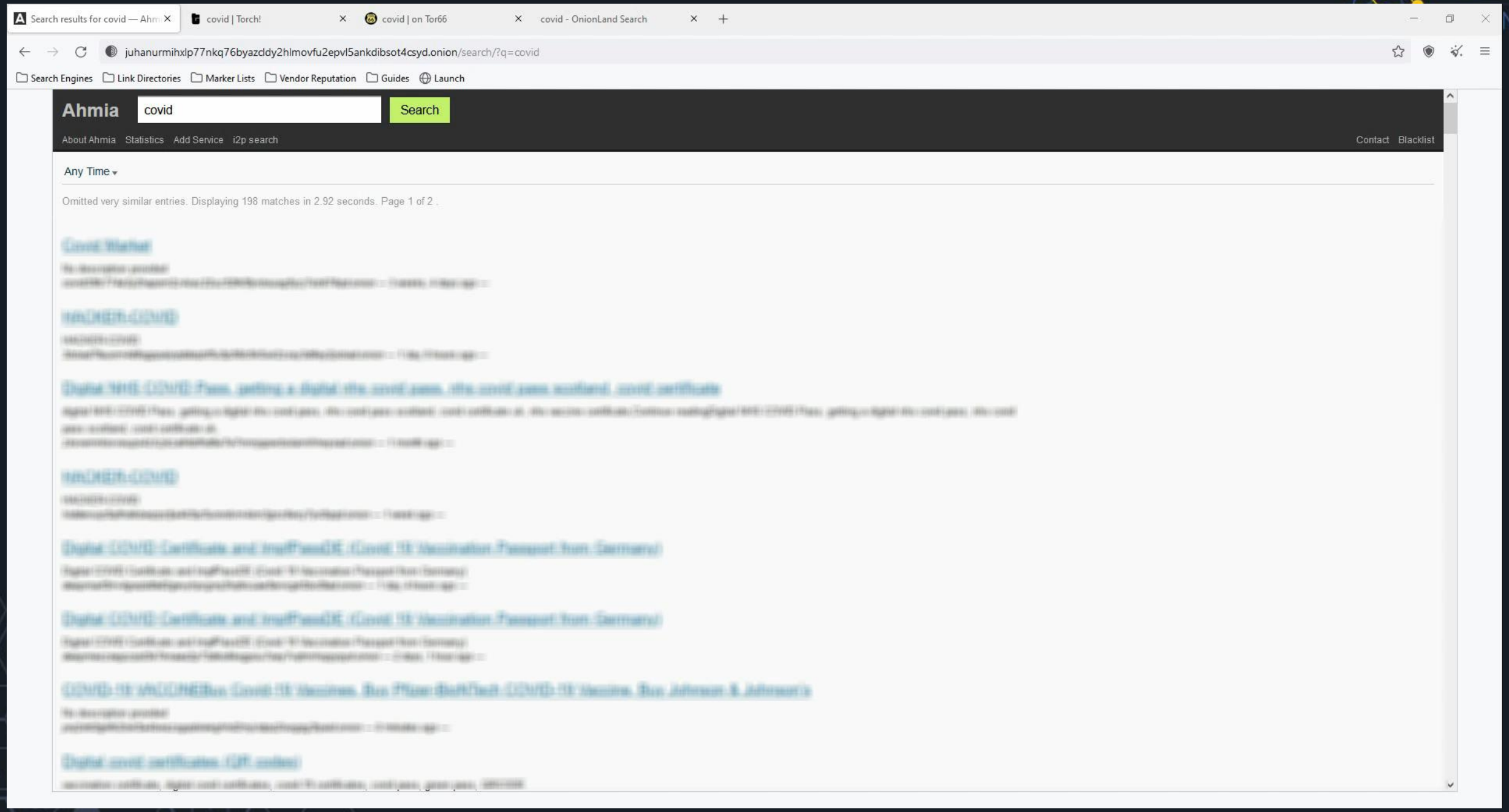
Aggregated & broad monitoring

2

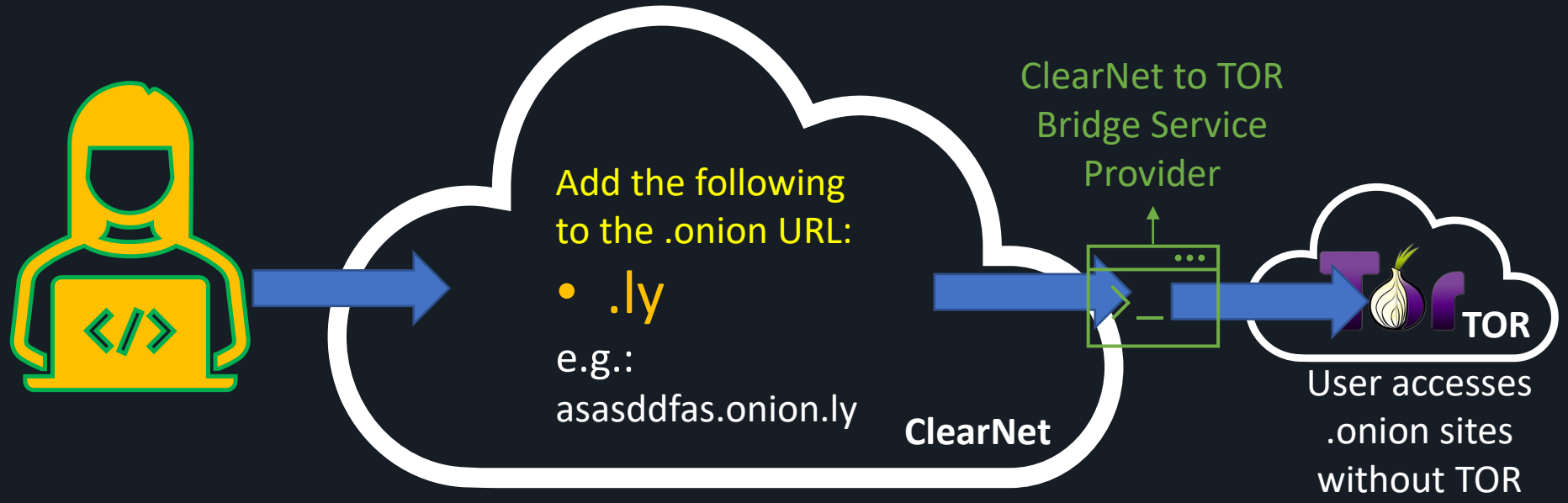
Scrape a specific site

Targeted monitoring

- GUI Scraping Tools
- SaaS Platforms
- Build our own


Search results for covid — Ahm
covid | Torch!
covid | on Tor66
covid - OnionLand Search
juhanurmihxlp77nkq76byazcldy2hlmovfu2epvl5ankdibsot4csyd.onion/search/?q=covid
Search Engines
Link Directories
Marker Lists
Vendor Reputation
Guides
Launch
Ahmia
covid
Search
About Ahmia
Statistics
Add Service
i2p search
Contact
Blacklist
Any Time
Omitted very similar entries. Displaying 198 matches in 2.92 seconds. Page 1 of 2 .

A word on Clear Net to TOR Proxy (Tor2Web)

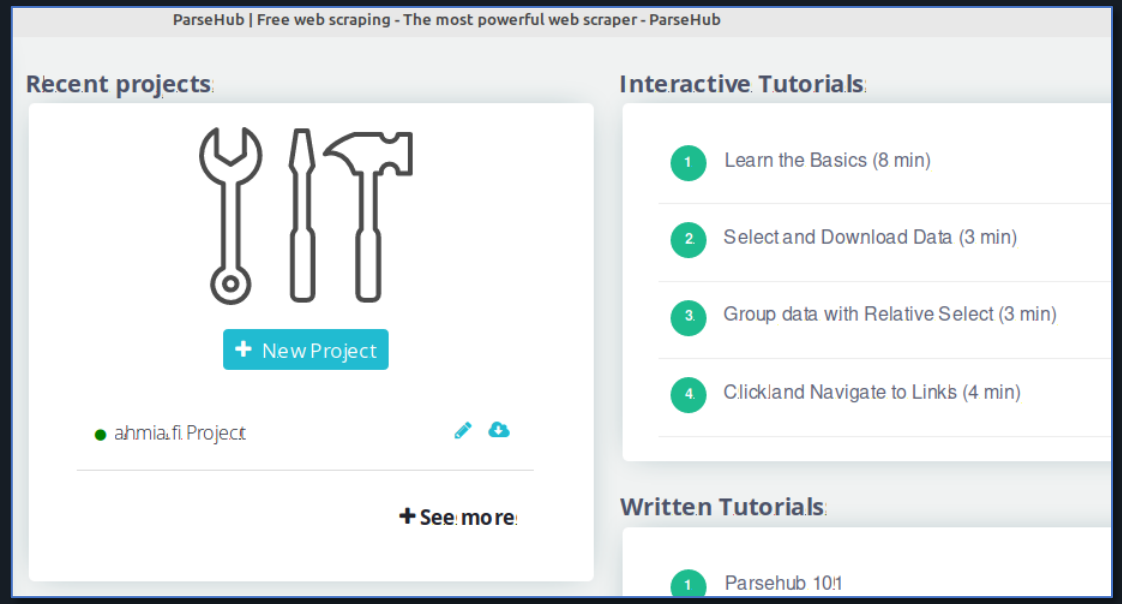


IMPORTANT

- Use only in a clean environment (Virtual Machine)
- Use only with clean browser (clear cookies & cache)

Easy Scraping

1. Get a list of the query URLs for the search engines of choice
    1. Add .ly to URLs if it's an onion site & no TOR infrastructure is set up
2. PraseHub (GUI) – www.parsehub.com
    1. Install in a clean environment
    2. Create your tasks
    3. Schedule scraping



https://help.parsehub.com/hc/en-us/articles/218187697-Enter-a-list-of-URLs-to-crawl


LinuxMint_20.3_VB_LinuxVMImages.COM [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Computer
Home
22:59
Right Ctrl

Advanced Scraping

1. Build your own scraping tools

- https://www.osintcombine.com/thoughtleadership (SANS 2021 talk)
- https://realpython.com/beautiful-soup-web-scraper-python/
- https://oxylabs.io/blog/python-web-scraping

2. Use a command-line tool, e.g. scrapy (www.scrapy.org)

IMPORTANT

- Only scraping & extracting text initially
- Use clean and isolated systems
- Consider your network attribution

Wrap up

What we covered….

- Networks (attribution, events, *informed* access)
- Services & searching
- Scraping & monitoring

Summary

- The Dark Web is a hive of illegal activity. Consider your access approach & misattribution requirements

- Automation can be fairly simple, just plan your requirements & find the most suitable approach

Chris Poulter

https://www.linkedin.com/in/chris-poulter/

chris@osintcombine.com

Dark Web focused class coming up....

https://www.osintcombine.com/darkweb

Tuesday, 10th May 2022
9:00am - 5:00pm
Australian Eastern Daylight Time (AEDT)

OSINT COMBINE

www.osintcombine.com

academy.osintcombine.com

@OSINTCOMBINE

NexusXplore

www.nexusxplore.com