HACKHALT
CYBER INTELLIGENCE COUNCIL

Report on Tools & Techniques
in Dark Web Investigation

DIVYANG SARKARI
SECURITY RESEARCHER

Introduction

The dark web is a part of the internet that is not accessible through our daily life browsers or search engines. The part that we normally use in our daily life is called surface web. Then there is a part of the internet that is used for encrypted databases, private networks, internal servers and much more, which is called deep web. It can be accessed via normal browsers and search engines but we need proper authentication allowance to access this part of the internet or illegally break into these infrastructures. Then comes the part of the internet that is basically not hidden, but is not accessible to the majority of the population from available normal browsers and search engines. This part of the internet is known as The Dark Web and the most interesting thing about the dark web itself is that it is not regulated by any authority which makes it a perfect place for all sorts of illegal digital activities. You can find almost every information which is out there on the internet or at least connected to the internet in some sort.

To access the dark web it is not a very big problem, all you have to do is get a Tor browser then connect to the tor network. But, this only gives you the ability to connect to the internet without any restrictions. Your ISP would know that you just connected to the Internet via Tor Network but beyond that your activity is kind of hidden from the ISP. Still, to access the dark web after connecting to Tor, you must look for a significant amount of security protocols so that you don't mistakenly put yourself in any kind of danger that can affect you in your life (mentally, physically, emotionally and financially). There is nothing harmful in connecting through Tor, you can still access all the surface web from there but because there are no regulations for the safety of an individual it is not recommended to do it often.

This report describes the tools and techniques which are widely used for Dark Web investigations and the safety-first setup which is required in order to keep the

investigator and the investigative body safe from any kind of data leakage. These security measures are very important from the investigation point of view because a little carelessness can lead to exposure which can lead to life threatening situations.

Background

The dark web operates primarily on onion routing, where traffic is encrypted and also routed through many relays to ensure anonymity. Each relay only decrypts a single layer which prevents the full path from source to destination invisible from any kind of exposure. Even if there is a MITM, he will only know about the layer which the relay got its traffic from. Previous research shows that the Tor's architecture makes attribution lot complicated due to:

1. Multi-hop routing
2. End-to-end Encryption
3. Decentralized node architecture

It wasn't created by the hackers, it was actually created by the U.S. government for high-stakes intelligence. It was started as a military project only to become a global enclave of anonymity.

The Military Origin (1990s)

The researchers at the U.S. Naval Research Laboratory were looking for ways to protect the dignity of the intelligence communications online because they had already realized that if a spy sent an encrypted message, an observer can see where that data is going.

To resolve this issue, they developed Onion Routing. The idea was to wrap the data in multiple layers of encryption (like layering of an onion) and bounce it through several nodes or servers before it finally reaches the destination. Each node only knew the previous and the next hop, making the source and destination impossible to trace.

The Birth of Tor (2002)

In the year 2002, this project evolved into The Tor Project. While it started as a government tool for safe and secure intelligence communication, its creators realized a sort of paradox: if only the government agents used the network, then any traffic on the network would be identified as government activity.

So to create a camouflage, the government released the whole project as open-source software which allowed activists, journalists, and everyday citizens to operate it which provided a perfect crowd for intelligence agents to blend in.

The Hidden Services (2004)

A vital moment occurred when Tor added the facility for users to not just browse the internet anonymously, but to host websites anonymously too. These sites used the .onion suffix and didn't show up on google. This functionality created what we know today as The Dark Web - a place where both visitor and the host were invisible to each other.

The Crypto and Commerce boom (2011)

Dark web was a niche tool for privacy advocates until two major catalysts arrived:

1. Bitcoin: provided a decentralized, semi-anonymous way to move money.
2. The Silk Road: It was launched by Ross Ulbricht in 2011. This was the first modern “Darknet Market”.

These two technologies turned the Dark Web into a functional shadow economy because there was a way to buy and sell anything without a bank or a government as a middleman. Today, the Dark Web is a complex duality. It is a crucial sanctuary for whistleblowers and people who are living under an oppressive regime and simultaneously it still remains the marketplace for illicit goods, stolen data and cybercrime.

Before diving into the specifics of Dark Web investigation, the understanding of the technology behind the working of it is crucial.

The Onion Router

Tor utilizes Onion Routing as we have already understood which is a layered encryption model.

1. Circuit construction: traffic passes through 3 volunteer nodes i.e, the entry node (Guard), a middle relay and the Exit node.
2. Hidden Services: for sites having .onion suffix, the circuit terminates at a rendezvous point where both client and the hosted service meet. Therefore, ensuring neither discovers the other’s IP or physical location.

Invisible Internet Project (I2P)

I2P implements Garlic Routing where it bundles multiple messages into a single encrypted packet called “cloves”.

1. Tunnels: Different from Tor’s bidirectional circuits, I2P uses unidirectional tunnels. The path for a request is always different from the path of the response.
2. UDP support: Because I2P supports UDP, it makes I2P the primary infrastructure for P2P file sharing and media streaming.

Feature	**The Onion Router (Tor)**	**Invisible Internet Project (I2P)**
Routing Protocol	Onion Routing (Circuit-based)	Garlic Routing (Packet-based)
Tunnel Symmetry	Bidirectional	Unidirectional
Network Model	Semi-centralized (Directory Authorities)	Fully Decentralized (DHT/NetDB)

Primary Use Case	Web browsing and Hidden Services	P2P, messaging, and internal "eepsites"
Protocol Focus	TCP only	TCP and UDP

After getting the history and working behind us, it is now time to reflect on what could be sold or bought from the Dark Web Marketplace. It's true that the Dark Web provides a way to use the internet and the related services without any surveillance but it is also true that this also leads to more complicated crimes one could ever imagine. Most importantly, The Dark Web isn't operated in only one level. It is operated in several layers which gets darker when you get to the next level. There are many opinions about the definition of the Dark Web which almost always leads to a dilemma about the core understanding of this side of the internet.

Why is it called the Dark Web? Why not "secret web" or "intelligence web" or "people's web" or "one-world one-web" or called anything else but The Dark Web. The truth is that everyone knows that it is not just the transaction of illegal items or drugs but far more than that.

On one level, there is a place called "The Red Room" where people actually pay for watching a person torturing someone physically to the point where we can't even imagine. There are lots of websites that are hosted and they openly state that there are several females to be sold from many origins. There are many researchers and experts that claim this is just a hypothetical concept but there is no smoke without a fire.

Let's dive into the tools and techniques that are used for professional Dark Web investigation.

Phase1: Operational Security (OpSec) & investigation environments

Establishing a non-attributable investigative domain is the first technical hurdle. Failing to not prepare a proper preventive environment can lead to "identity leakage" where attributes like IP, time-zone, browser fingerprint can bleed into the research persona.

Hardened Investigative Operating system

1. **Tails(The Amnesic Incognito Live System):** It is a Debian-based OS which is designed to perfectly run from RAM, primarily designed to force all the connections through Tor and wipes the entire system state upon shutdown, which helps in leaving no digital traces on the local hardware. This OS is widely used to access the Dark Web for several reasons.
2. **Whonix:** This implements a dual-VM architecture. The Whonix-Gateway handles all the Tor traffic, while the "Whonix-Workstation" remains isolated from the traffic and connections because the workstation never knows its own public IP-address, preventing de-anonymization via browser exploits.
3. **Qubes OS:** This operates by leveraging Xen hardware-level virtualization to isolate investigative "cubes". Researchers widely use Disposable VMs to open suspicious files, ensuring that the malware doesn't persist across sessions. The best part is that this OS is open-source for single-user desktop computing. There is a predefined set of more than one isolated applications for such projects to manage the network-stack and firewall or other user-defined purposes (for more see here).

Counter-Fingerprinting techniques

As we all know that the websites collect information such as screen resolution, installed fonts and hardware concurrency to generate unique identifier, mitigation steps like using Tor browser at its default window size is the best because maximizing the window will reveal the screen dimensions; writing all the communication in the local editor beforehand to reduce the keyboard biometric analysis and avoid predictable typing rhythms; using a wifi adaptor which reduces the risk for ip leak.

Phase2: Technical Reconnaissance & Infrastructure Mapping

The agenda of reconnaissance is to map technical backend supporting a hidden service to find configuration flaws.

Infrastructure fingerprinting tools

This is a very crucial phase from the perspective of both investigation and security. While it helps to identify many point outs, it also exposes the investigator to fall for fake redirection links leading to compromise of the integrity of the investigator. To prevent this from happening, the investigator use the following tactics:

1. **Shodan/Netlas:** These tools are used for unique HTTP headers or SSL certificates that might be reused across Dark Web and clearnet both. Shodan is especially used for Internet-connected devices which are either open to the internet or somehow compromised. Netlas is a comprehensive internet-wide scanning tool and OSINT platform designed for cyber security professionals to discover, query and analyze connected devices, IP addresses and domains using non-intrusive scanning. It provides real-time data which is searchable

on publicly accessible internet services. Shodan is little prone to the listing of purposefully compromised devices which act like a honeypot where someones connectivity details are gathered for profiling so it is generally advised to be cautious when clicking the IP link provided on shodan as it can lead to identity theft or disclosure.

2. **Favicon and Asset hashing:** This refers to the comparison of the cryptographic hash of a .onion site's favicon or script files against the known clearnet IP addresses to find matches. As we have already established that the Tor hosted websites use .onion domain suffix to remain immune to endpoint identification which makes it difficult to investigate the traffic on the website, also hosting of public forums, communication hubs, and most importantly illicit marketplace. While Tor v.01 was released in 2002, its hidden services were introduced in Tor v.02 around 2004. With version 3, or as we know it as the next gen onion services with 56-character addresses was introduced in 2018 and the current Tor v.04.8.17. Tor circuits are built using telescopic path - building design. This process involves the initiator negotiating independent session keys with each hop. This ensures perfect forward secrecy aka PFS. Therefore even if the intermediate key is compromised, the previously transmitted data or communication remains secure. The question to ask here is "why mapping is important?" The answer to this question is, the Dark Web operates on anonymity, and this anonymity creates a critical blind spot for security or threat teams. Mapping is therefore important as it lets us discover and analyze the technical background that supports such anonymity or hidden services as discussed above, without actually interacting with them which also creates a safe layer between the investigation and the target involved as mapping primarily involves focus on

servers, domains and technical fingerprinting that helps investigators access these hidden services.

As from the perspective of Threat Intelligence teams, mapping gives them a critical visibility on the Dark Web infrastructure:

1. **Threat actors ecosystems:**

   As most of the places involved in the Dark Web space, are the hubs for dark crimes, mapping these infrastructures, threat intelligence teams can where different adversaries involved in crimes are engaging, conducting gatherings, trading data and communicating. This way mapping helps intelligence teams to create an ecosystem-view of these criminals to identify connections that might seem unrelated.

2. **Infrastructure overlapping:**

   Not all the adversaries are geniuses, several are just beginners who're trying to make a quick money, despite the anonymity offered on the Tor, they often reuse the elements for the setup like hosting server, a domain pattern or sometimes misconfigured certificates. TI teams identify these overlaps linking these hidden services to each other and sometimes even to clearnet resources. Bulletproof hosting is used as infrastructure for various Dark Web sites, it is a technical hosting service that remains unaffected by legal complaints about activities, which serves as a basic building block for hidden sites and cyberattacks.

3. **Early warning:**

   Monitoring or mapping also provides an early warning system to organizations that helps prevent or defend cyber attacks effectively. By tracking, TI teams can track the trading of an exploit, compromised asset details, and even the communication involving planned attacks. For example, the famous Medusa ransomware group lists all its victims on its Tor-based leak websites. This provides an immediate visibility into data exposure. The data itself might not be useful for organizations, but it can bring many issues to the light which never seem to be a compromising asset. Nmap is a network mapping tool that implements a full TCP port scan (0-65535) and identifies unreferenced administrative interfaces like phpmyadmin, /wp-admin. Or other exposed database ports.

Network investigative techniques:

These are specialized , often intrusive, law-enforcement tools that are used to hack into computers and identify the suspects operating on the Dark Web or hiding their IP addresses.

There are some aspects related to Network Investigative techniques:

The term was coined by the FBI, used to bypass or pierce, anonymizing technologies and locate individuals involved in criminal activities. Critics and digital rights groups call them government-authorized malware as it is dropped on to the target machine and is referred to as a drive-by-download malware, this also raises significant legal challenges therefore the courts have scrutinized the scope of such techniques especially when one warrant is used to hack thousands of computers.

These tools are not used in day-to-day investigation, but are typically deployed in high-level investigations including terrorism, national security threats and sexual exploitation. It is slightly different from the forensics, as it is an active offensive technique used to extract information from the target machine.

But to deploy NITs, there has to be a vulnerability to exploit in order to gain access to the system. For this, TI teams or other government bodies either identify the vulnerability or purchases zero-day exploits in browsers or OS. As every other exploit, NITs are delivered through phishing or compromised websites. The program executes, captures the computer network traffic including the real IP address, mac address and other identifiable data.

Phase3: Identify Attribution and behavioral profiling

Attribution helps in connecting anonymous handles into real-world individuals through link analysis and linguistic markers. For this, the forensic psychologists develop criminal personas from forum activity patterns:

1. **High-value ransomware operators-**

   Forensic psychologists look at the timing of the post for target timezone alignment. Like US/Europe business hours. Avoidance of slang and the use of business language. They look at shared detailed attack logs to generate reputation building. They also look at the negotiating tactic to see if it is done publicly on any leaked site.

2. **Low-level operators-**

   Forensic psychologists look at their local timing, late night activation, slang-heavy and emoji used in the communication, and the frequent amount of changes to identify any paranoia which leads to panic selling of stolen data post-breach. These are mostly newbies who lack coordination and patience with their criminal activity which leads them into the list of easy suspects.

India represents around 12% of global Dark Web traffic blending domestic criminal needs with international market. Forums like Jamatara carding, sell PAN/Aadhaar dumps at merely ₹42-425 each. UPI phishing kits targeting digital payments at minimum ₹8,500 monthly. Various government exams leaked paper via Tor paste sites and the Dark Web hawala services which converts BTC to INR cash.

The National Crime Records Bureau (NCRB) 2025 data shows 3,847 dark web related cases which has risen by 280% since 2021 with Kerala leading blockchain tracing recoveries(around ₹127 crore).

PGP key correlation

PGP keys(Pretty Good Privacy) are a pair of cryptographic keys - a public key for encrypting data/verifying signatures and a private key to decrypt data/creating signatures. They ensure secure communication, encrypt the email, and file signing with commonly used algorithms like RSA or ECC. PGP keys are the primary trust mechanism on the Dark Web as we have discussed earlier, during the server seizures, investigators extract administrator databases which contain thousands of public keys.

1. Pivot points: It is so crucial that a single PGP key can swivel to link an actor across 50+ different marketplaces. This leads to more elaborated data points which leads to a more sophisticated digital profiling.

2. Tools used: Mostly, tools like Maltego and i2 Analyst's Notebook are used to create a visual representation of the connections between tor handles, PGP keys and social media or digital profiles on the surface web.

Stylometry and Linguistic fingerprinting

Today, with help of the modern developed LLMs and AI tools, matching writing patterns, syntax and slang between Dark Web forum's post and real-world personas provides a way to attribute activities to specific adversaries or criminals. The criminals on the Dark Web mostly try not to repeat the patterns in their communications so that they can avoid profiling but this also makes them spend months to develop communication style and building trust, therefore, the investigators must match that commitment with patience.

Phase4: Cryptocurrency forensics and financial tracing

Cryptocurrency is the economical subway of the Dark Web. To trace the transactions, the investigation is divided into two primary typologies:

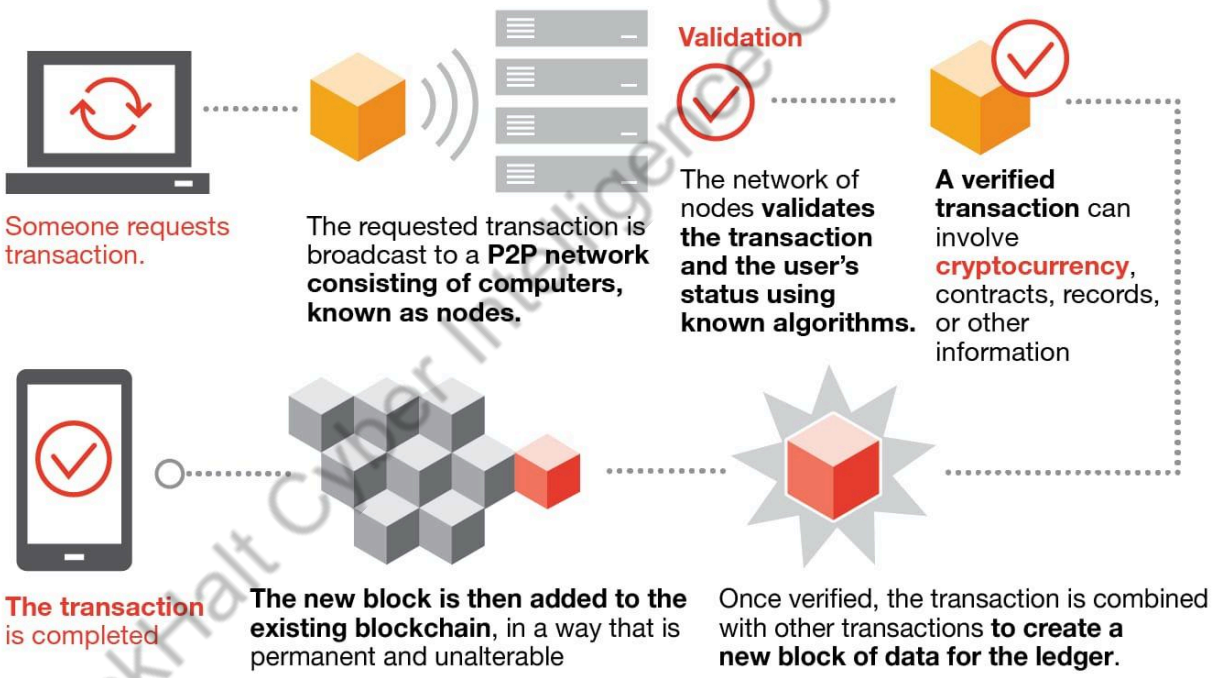
1. **Type A (Off-Chain):**

   The origin of type-a is the suspect's device. Investigators use tools like Autopsy or Magnet AXIOM to find wallet.dat files that hold all the details

about the crypto transactions, it stores the recovery seed phrases, and unencrypted local logs.

2. **Type B(On-Chain):**

This is initiated on the blockchain. As we know, the blockchain decentralized, distributable and immutable digital ledger that records the transactions across a network of computers, ensures integrity and security of data without any intermediaries.



We can see here through this image that there are no intermediaries involved in the transaction of blockchain technology. Investigators use tools like Chainalysis Reactor or TRM Labs to visualize the flow of funds and direction of the transactions. Blockchain works on five basic principles:

1. Distributed database
2. Immutable records
3. Transparent access

4. Algorithm-based computational logic
5. Two-way participant transmission.

   Bitcoin and privacy coins fuel commercial transactions on the Dark Web. But, blockchain's immutability becomes investigators' greatest asset.

Advanced tracing techniques (workflow)

1. **Address clustering:** Using shared-input enables investigators to identify all wallets controlled by a single entity.
2. **Peeling chain analysis:** Exchange of KYC matchings where flat conversion reveals real identity by automatically tracing small "peeled" amounts from a large wallet to identify vendor payouts or "off-ramps" exchange where KYC data can be obtained legally.
3. **Taint analysis:** It is a security analysis technique that tracks the flow of untrusted user input(sources) through an application or sensitive operations. This analysis is used for tracking dirty coins through obfuscation layers like mixers or tumblers. Also the timing evaluation helps in correlation with previously known Dark Web sale timestamps.

Phase 5: Forensic evidence acquisition

The forensics team requires a blend of traditional disk imaging and specialized memory analysis. **Traditional disk imaging** is a process that creates an exact, sector-by-sector copy of the storage device (hard drive, partition, USB) into a single, compressed file. This includes the OS, applications, settings, and files that allows complete system restoration or cloning into a new hardware. Whereas **specialized memory analysis** is a branch of DFIR that involves examining the

volatile data of a computer or device to identify malicious activity, unauthorized access which are counted as artifacts from malware.

Live capturing and memory forensics

In a scenario where the device is captured live or logged in, the investigators can bypass full-disk encryption. The standard forensic softwares such as EnCase, FTK or Magnet AXIOM, are used for imaging the hard drives and comb through data. But in the case of Dark Web, there are some special focuses:

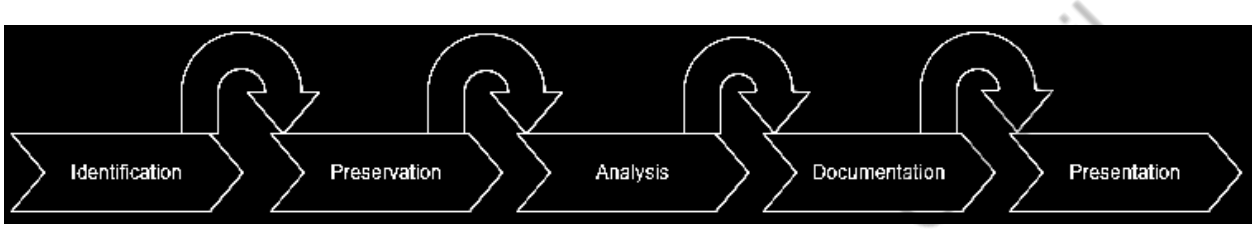
> "Criminals almost always use some kind of encryption to protect their data in case of any emergency. If the forensics team gets their hands on a live device then they can get every piece of information in a plaintext. But, due to the encryption, it usually takes a bit of time to crack the passwords and find the keys. Sometimes the court of law can compel the suspect to disclose the encryption key. There have been some instances where the hidden cameras, keyloggers have been installed to avoid the battle of encryption and decryption. It sounds a bit tricky, but it usually works."

Volatility framework is the primary tool for analyzing RAM dumps to extract private keys which are unencrypted, active Tor circuits, and chat histories in plaintext. But, there is also an order in which the data acquisition must be done so that there is minimal to no data loss. Investigators capture data in RAM first as it is lost the instant the device is powered down. This ensures that all the volatile data - running processes, network connections, and encryption keys are preserved.

Standard order of Volatility (Most to Least volatile)

1. **CPU Register and cache:** Extremely fast, stores temporary data and is lost instantly when powered down. This stores the most transient data and this data can be lost in milliseconds if not handled properly. Primarily it has CPU state, active running processes information.
2. **Routing tables, ARP cache, Process tables, kernel statistics:** This has the dynamic data structures that can change frequently during system operations. Its volatility is very high and the data here changes with the network and the activities of the system.
3. **Memory(RAM):** RAM stores memory of active processes, temp files and data which are currently in use. As we know all the data in the RAM is instantly lost when the system is powered down, which makes it most volatile.
4. **TEMP files:** It is a storage location for temporary files which are used by OS and applications running on the device. Its volatility is low as the data remains as is until it is intentionally deleted or overwritten.
5. **Remote logging and monitoring data:** These are the logs and monitoring data stored on remote servers. Its volatility depends on the retention policies or the storage practices. As some servers store data based on particular time duration after that the data is automatically deleted.
6. **Physical configuration and network topology:** This refers to the hardware setup and network structure of the system. Its volatility is low but it changes frequently.
7. **Archival media:** This refers to the long term preservation media used for the backups and archiving purposes. Its volatility is considered extremely low as the data persists over a long period of time(See here).

Now we can understand the importance of the Order of Volatility because it ensures the most volatile data is preserved before it gets lost due to some very minor mistake, which also implies the importance of the critical evidence. It guides investigators and makes it easier for them to follow the protocols and maintain a proper chain of custody throughout the operation. The complete process can look like:



Web capture and preservation

It is a process of collecting, recording and maintaining digital evidence from the websites and online platforms in a manner that ensures the integrity, authenticity and admissibility in the court of law. As web content can be highly volatile and dynamic in nature can also be easily altered. Therefore specialized forensics tools are required to create snapshots of the data which are reliable and impactful for meaningful forensic results. The capture and preservation are critically important for both defensible and tamper-proof records of online evidence. The widely used tools are:

1. **Hunchly:**

   It is a specialized OSINT tool designed for investigative agents to capture, preserve and organize digital evidence during online research. It works by automatically recording every visited webpage, making it invaluable for

investigative, journalistic and ethical environment purposes. With hunchly, professionals can perform tasks like:

1. Capture and archive every web page visited.
2. Preserve evidence with time-stamps and metadata.
3. Automatically arrange and tag the research findings.
4. Produce a report that is court-ready.
5. Conduct anonymous and secure investigations.

Feature	Description
Automated Web Capture	Automatically records every webpage visited, ensuring no evidence is lost.
Searchable Evidence	Enables investigators to search captured pages by **keywords, URLs, or timestamps**.
Court-Ready Reports	Generates **legally admissible forensic reports**.
Cryptographic Hashing	Ensures the integrity of captured evidence remains **unchanged**.
Custom Tagging & Annotation	Allows users to **categorize and annotate** collected data.
Multiple Export Options	Supports export in **PDF, HTML, and JSON formats** for further analysis.
Integration with OSINT Tools	Works with **Maltego, SpiderFoot, and**

	Shodan for expanded research capabilities.
Cross-Browser Support	Compatible with **Chrome, Firefox, and Chromium-based browsers**.
Privacy-Focused	Supports **Tor, VPNs, and ProxyChains** for secure browsing.

Summary

The above table shows how effectively hunchly can help digital investigation without any unnecessary hassle, providing best choices for investigative purpose.

So far we have seen how the Dark Web investigation is carried out and the process of acquiring the evidence via various tools and techniques which also requires a proper understanding of what type of data is useful for forensics and what data should be left out so as to not increase the chances of mishap, and the methods to preserve the evidences to analyze, document and present in the court of law for further actions. It may seem unsophisticated but in reality the more we go in depth of the Dark Web, the more complicated issues are ready to be addressed.

Sometimes a little carelessness can lead to such situations where it becomes life threatening for the investigators involved in the operation. Therefore, proper preventive techniques are required for safe and sound environment where the investigators aren't worried about something personal getting affected by the work at hand.

References

https://clocked-out.com/what-is-order-of-volatility/

https://the420.in/dark-web-forensics-tor-blockchain-india/#cryptocurrency-transaction-forensics

https://blog.college.ch/blockchain-technology/all-you-need-to-know-about-cryptocurrency-and-blockchain-technology/

Get in touch

+91 8595440110

info@hackhalt.org

www.hackhalt.org