Social Media Forensics and Disinformation Tracking

ABSTRACT

Social media evolved as one of the major hubs of communication, but its misuse has led to many challenges, such as spreading disinformation, fake news, deep fakes and more. By using Digital forensics on social media, we can develop a strong approach to gathering, analyzing and identifying the collected data. Social media forensics evolved into an important and crucial research area, moreover media manipulation is a pressing problem in this field, where many struggle to identify between what is fake and what is real. The main aim of this paper is to dive deeper into the field of social media forensics, tracking fake propaganda, disinformation campaigns, tools used, and challenges faced. By analyzing the real-world cases we can understand the role of social media forensics in safeguarding the digital environment

Contents

1. Introduction

Social media is mostly used to share information, express opinions, and communicate with others over digital landscape. However, Social media applications like Instagram, X, Facebook and Reddit are aimed at manipulating public opinion by spreading misinformation, digital impersonation and deep fakes. These fake posts have emerged and gained millions of views that have affected public opinion. The openness of these platforms gave rise to such threats.

Social media forensics involve gathering, analyzing and identifying data from online to support investigators. It involves tracking user activity, finding fake accounts, detecting deep fakes and fake images which are generated by using AI tools and analyzing any found anomalies. Disinformation tracking mainly tracks the spread of false information or misleading information that has been circulating on social media platforms.

Due to the Continuous growth of social media and its influence on people, many disinformation monitoring and counter groups have been established to tackle the spread of miss information, such as Cogsec Collab and Beacon Project. Investigators must ensure evidence is collected systematically, preserving integrity, while maintaining legal and ethical frameworks. As misinformation techniques evolve, the importance of social media forensics continues to grow.

2. Types of Digital Evidence on Social Media

When investigating crimes involving social media, it's not just checking saved posts, screenshots and comments. Investigators need to collect a wide variety of digital clues to understand what is happening and who is involved.

- **Posts and Comments:** Public Posts, Private messages, comments can give us a lot of information about someone's actions or intentions. So investigating these can give us users intentions and much more.
- **Photos, Videos and Streams:** When we share any posts on social media it also hides extra information which cannot be visible, basically this hidden data is what we call as metadata by using this we can connect all the dots. Ip address, location, device model, lens used and date and time of event all these come under metadata.

- **Account Information:** Details like when the account is created, where and when it is created, what IP addresses were used to login to that account, and any phone numbers or emails linked to that account can help investigators to track back to real people
- **Social Trends:** By finding out who connects with who, followers, following, friends and tagged accounts can help investigators to find fake bots, or organized people who manipulate others by texting them online

3. Disinformation and Deepfake Analysis

3.1. Disinformation Campaigns

Generally, disinformation refers to the deliberate spread of false information intended to mislead audiences. Organized campaigns use fake accounts, bots, and emotional content to manipulate public opinion. Some of the tactics include, amplifying some conspiracy theories, targeting the specific groups with tailored misinformation and suppressing the true information which is through distraction techniques

Example**:** In the 2016 U.S. presidential election, operatives from the Russian-based Internet Research Agency spread divisive political content across social media networks to influence voter perceptions.

3.2. Deepfake Technologies

The Deepfakes use machine learning algorithms (especially GANs - Generative Adversarial Networks) to create some hyper fake videos. These can damage in many ways like reputations, influence political outcomes, or facilitate fraud.

Detection Techniques:

Frame Analysis: In the frame analysis we use Spotting pixel anomalies, lighting inconsistencies, or unnatural facial movements.

Audio Forensics: The audio forensics are mostly used for Identifying synthetic speech patterns.

Blockchain-Based Authenticity Verification: Emerging technologies which allow cryptographic signing of authentic media which is at the point of capture.

4. Tracking Online Harassment and Cybercrimes

Online harassment has become increasingly common today and includes forms such as cyberbullying, doxxing, stalking, and revenge porn. Investigators track perpetrators by:

Analysing Posting Patterns: Investigators look at how often someone posts, at what times (which can hint at their time zone) and their language style or tone. These details can provide important clues about who's behind the harassment.

Tracing IP Addresses: In serious cases, authorities may approach internet service providers (ISPs) with legal requests (like subpoenas) to match IP addresses with real users.

Spotting Fake or Duplicate Accounts: Duplicate accounts Known as "sockpuppet" accounts, these are fake profiles often created to mislead or harass. Investigators check for reused profile pictures, similar email IDs, or slight changes in usernames to uncover them.

Example: In 2021, TikTok saw a surge of "hate raids" where large groups of bots targeted and harassed creators from minority communities. These attacks weren't random and they were planned. Digital forensic experts stepped in and by closely studying the patterns in these bot activities, were able to trace them back to organized groups working behind the scenes.

5. Case Studies

- Case Study 1: 2016 U.S. Election Interference

In Russia, investigators discovered that operatives were using social media platforms like Facebook groups, fake pages and Instagram ads to influence public opinion. Through digital forensic analysis of ads, posts and metadata, experts were able to trace thousands of these accounts back to the Internet Research Agency based in St. Petersburg.

- Case Study 2: Deepfake Celebrity Scandal

In the year 2019, a deepfake scandal surfaced involving synthetic videos that showed celebrities in compromising situations. Forensic experts stepped in and used frame-by-frame analysis to examine the footage closely. They spotted inconsistencies like mismatched lighting between the faces and the background which helped expose the videos as fake and debunk the material.

- Case Study 3: Christchurch Shooting Livestream

There was a livestreamed terror attack in Christchurch in the year 2019, New Zealand, was rapidly shared across various social media platforms. Forensic teams collaborated with tech companies to trace how the video spread online. This investigation played a key role in pushing platforms to introduce stronger content moderation policies to prevent such incidents from going viral in the future.

6. Forensic Tools and Techniques

SpiderFoot: SpiderFoot is an open-source OSINT automation tool. It helps gather public information about usernames, email addresses, domains, and IP addresses. It's very useful for tracing fake social media accounts or mapping disinformation websites.

Maltego: Maltego is a powerful tool for visualizing relationships between entities. Investigators use it to map connections between accounts, domains, and email addresses making it easier to see networks behind disinformation campaigns.

X1 Social Discovery: X1 Social Discovery is a specialized forensic tool for collecting evidence from social media platforms. It preserves posts, messages, comments and metadata in a legally admissible way for court use.

Sherlock: Sherlock was used to identify the presence of suspicious usernames across multiple social media platforms. By mapping these accounts, we established digital linkages that suggested coordinated disinformation activities. This initial reconnaissance complemented deeper investigations carried out using SpiderFoot.

SpiderFoot was utilized as a core tool to demonstrate the application of automated OSINT gathering for social media forensics and disinformation tracking. The test was conducted on different domains and in particular it was victorlivestockfarm.co.az to illustrate how spiderfoot can give the digital footprints linked to potential disinformation.

The SpiderFoot scan was configured to query a wide range of data sources including DNS records, SSL Certificate information, internal and external URLs, email addresses, IP Addresses. Key results included the discovery of 15 unique email addresses potentially reused across different platforms, this can be a finding for disinformation networks. The scan also revealed the existence of cloned domain structures which gives suspicion of domain spoofing. Over 60 internal URLs were identified along with IPv6 address and multiple SSL certificates. SpiderFoot graph's visualisation feature was used to map relationships between the domain, email addresses, IPs and associated infrastructure. This proved highly valuable in illustrating how social media forensic investigators can visualise disinformation networks.

7. Legal and Ethical Considerations

Jurisdictional Complexities:

Social media platforms operate globally. Investigations involving foreign actors, like in election interference cases, face extradition issues, conflicting national laws, and diplomatic roadblocks.

Privacy Concerns:

Investigators must tread carefully to avoid infringing privacy rights. Collecting private messages without proper warrants violates laws like: GDPR (General Data Protection Regulation - Europe) CCPA (California Consumer Privacy Act)

Avoidance of Unauthorized Access (Computer Misuse Laws):

Investigators should not attempt to access private profiles without legal authority, unauthorized access may violate laws.

Preservation of Evidence:

Investigators need to properly document, record and store the evidence without any tampering to maintain its integrity.

Transparency:

If we are collecting data for any research purpose, we need to make sure that we inform people about what is happening and it's better to get consent if possible.

Ethical Use of Technology:

Use tools and methods carefully and responsibly do not use tools like facial recognition unless it is necessary and justified legally.

Minimizing Harm:

Investigators need to make sure that they don’t put individuals at risk and they need to restrict putting people at unnecessary psychological harm.

8. Future Trends in Social Media Forensics

As technology continues to evolve in social media, digital investigations also continue to evolve in adapting to new means of doing. Online communications continue to become more complex driven by technological changes as well as the growing volume of user-created data. Below are four major trends shaping what’s ahead in social media forensics:

- Blockchain for Chain of Custody: Preservation of evidence integrity has been one of the persisting issues in digital forensics. Examiners need to be able to guarantee that digital evidence be it screenshots, messages or logs does not get altered or modified throughout collection, analysis or storage. Blockchain technology is being touted as the answer. By utilizing a decentralized and immutable ledger, it is now feasible to track any access and try to alter an evidence item.

  For instance, when a chat history is extracted by a forensic examiner, the blockchain may record the timestamp of extraction, involved user credentials or even a hash of the original evidence. This open audit trail maintains faith in the authenticity of the evidence.

- AI-Driven OSINT (Open-Source Intelligence): The volume of content on the internet is beyond human capability to track manually. Artificial intelligence is being utilized now to automate the task of gathering and examining large volumes of publicly available information. Using machine learning and natural language processing (NLP), forensic teams are able to identify behavioural patterns, recognize fake accounts or coordinated disinformation campaigns and monitor emerging threats in real time. For instance, AI programs are able to identify sudden bursts of synchronized messaging during political rallies or even spot subtle manipulation in video or audio recordings posted on social media.

- Augmented and Virtual Reality Forensics: As AR and VR worlds become more popular, social engagement is no longer confined to text and pictures. Worlds such as Horizon Worlds or VRChat provide immersive environments where people engage via avatars, voice, gestures and live 3D worlds. This opens up new avenues for digital crime, virtual stalking, harassment and even spying. Forensics experts need to equip themselves to be able to investigate interactions within such worlds. This can include recording metadata such as user movement records, headset direction, session times and in-game chats. All these virtual worlds present themselves as the possible next frontier for both crime and digital forensic investigation.
- Standardization and Legislation: As digital evidence takes centre stage in legal cases, there is increasing call for definite guidelines on what makes up acceptable evidence. This is particularly important in the era of deepfakes and AI-generated content. Legal systems around the globe are starting to enact legislation that outlines how digital content needs to be gathered, stored and authenticated. For example, whether a viral video is authentic or not will eventually demand more than metadata, it will possibly demand cryptographic evidence that the content has not been altered since it was filmed. These legal frameworks strive to balance the admissibility of electronic evidence with the privacy and due process rights.

9. Conclusion

Social media forensics serves as a digital investigative framework essential for uncovering criminal activity, tracking disinformation and combating online harassment in our increasingly interconnected world. Forensic experts analyze content from platforms like Facebook, Instagram, and X , examining posts, messages, metadata, and multimedia content to extract valuable evidence. The field presents unique challenges, including verifying the authenticity of digital content, navigating international legal barriers, and managing massive, diverse datasets.

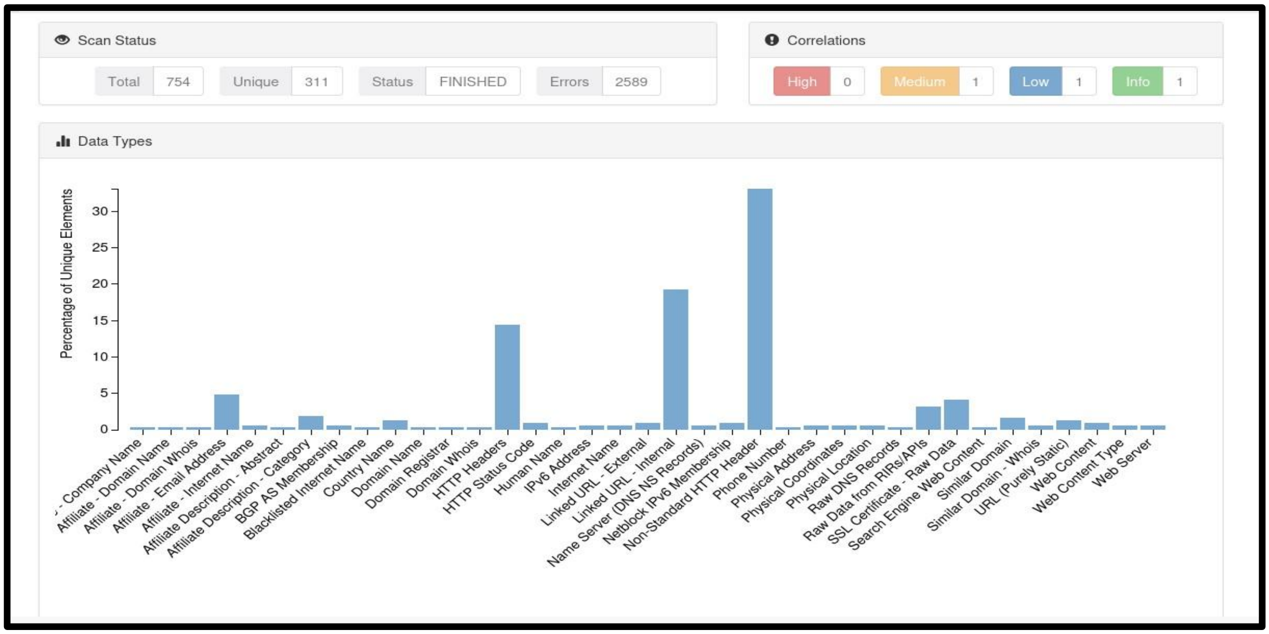
Disinformation tracking, a critical subset of social media forensics, focuses on identifying and analyzing the deliberate spread of misleading content. False news articles, doctorated media, and propaganda campaigns can significantly shape public opinion. Experts use both manual techniques and automated tools to trace the origins, spread patterns and actors behind such efforts, often uncover networks of bots or coordinated influencers.

To effectively address these evolving threats, forensic investigators rely on innovative tools, strict methodological standards, and adherence to ethical and legal principles. Looking ahead, emerging technologies like artificial intelligence and blockchain are poised to transform the collection and authentication of digital evidence. A multidisciplinary approach bridging technology, law and ethics will be vital to ensure that social media forensics continues to evolve responsibly and effectively in protecting digital societies from manipulation and harm.

References

- "The Role of Social Media in the 2016 U.S. Election," Senate Intelligence Committee, 2019. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf
- Westerlund, Mika. "The Emergence of Deepfake Technology: A Review," Technology Innovation Management Review, 2019. https://timreview.ca/article/1282
- Casey, Eoghan. Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet, Third Edition. Academic Press, 2011.
- https://ctc.westpoint.edu/christchurch-attacks-livestream-terror-viral-video-age/

APPENDIX



Scan Status

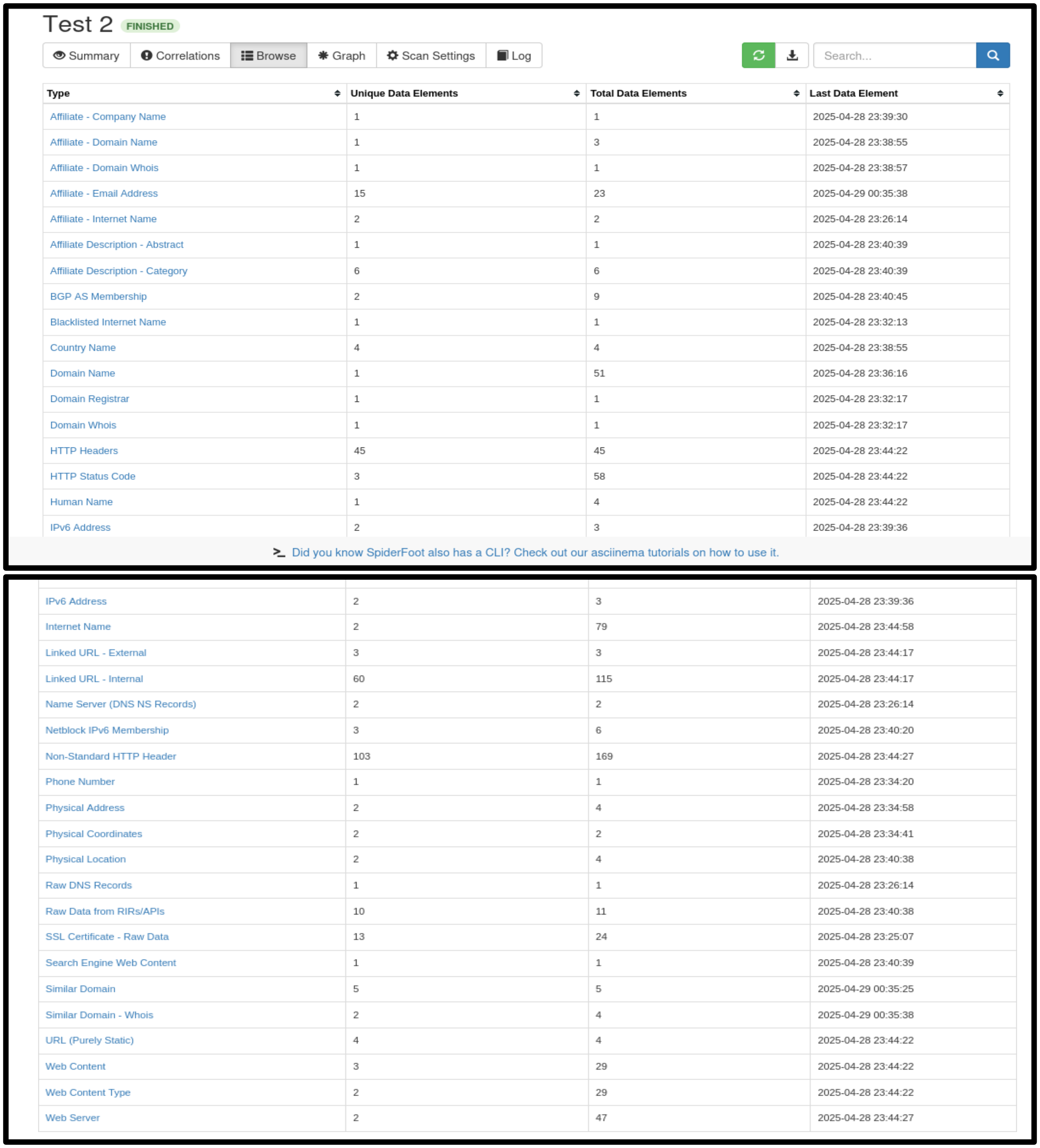
Total	754	Unique	311	Status	FINISHED	Errors	2589

Correlations

High	0	Medium	1	Low	1	Info	1

Data Types

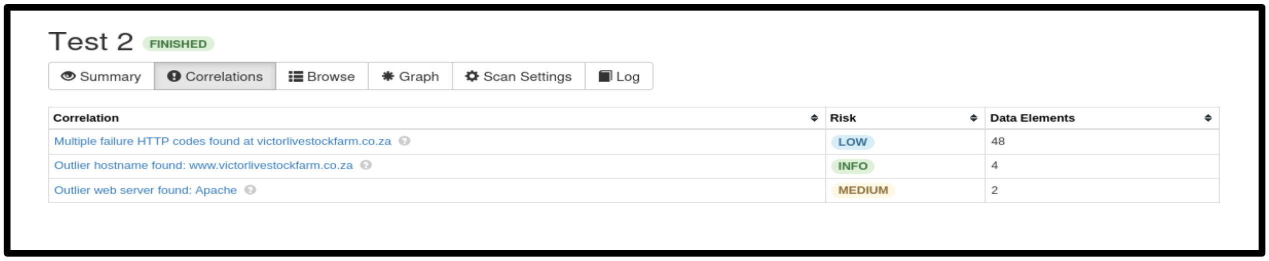
Fig. 1 Summary Results for the domain “victorlivestockfarm.co.za”



Type	Unique Data Elements	Total Data Elements	Last Data Element
Affiliate - Company Name	1	1	2025-04-28 23:39:30
Affiliate - Domain Name	1	3	2025-04-28 23:38:55
Affiliate - Domain Whois	1	1	2025-04-28 23:38:57
Affiliate - Email Address	15	23	2025-04-29 00:35:38
Affiliate - Internet Name	2	2	2025-04-28 23:26:14
Affiliate Description - Abstract	1	1	2025-04-28 23:40:39
Affiliate Description - Category	6	6	2025-04-28 23:40:39
BGP AS Membership	2	9	2025-04-28 23:40:45
Blacklisted Internet Name	1	1	2025-04-28 23:32:13
Country Name	4	4	2025-04-28 23:38:55
Domain Name	1	51	2025-04-28 23:36:16
Domain Registrar	1	1	2025-04-28 23:32:17
Domain Whois	1	1	2025-04-28 23:32:17
HTTP Headers	45	45	2025-04-28 23:44:22
HTTP Status Code	3	58	2025-04-28 23:44:22
Human Name	1	4	2025-04-28 23:44:22
IPv6 Address	2	3	2025-04-28 23:39:36
IPv6 Address	2	3	2025-04-28 23:39:36
Internet Name	2	79	2025-04-28 23:44:58
Linked URL - External	3	3	2025-04-28 23:44:17
Linked URL - Internal	60	115	2025-04-28 23:44:17
Name Server (DNS NS Records)	2	2	2025-04-28 23:26:14
Netblock IPv6 Membership	3	6	2025-04-28 23:40:20
Non-Standard HTTP Header	103	169	2025-04-28 23:44:27
Phone Number	1	1	2025-04-28 23:34:20
Physical Address	2	4	2025-04-28 23:34:58
Physical Coordinates	2	2	2025-04-28 23:34:41
Physical Location	2	4	2025-04-28 23:40:38
Raw DNS Records	1	1	2025-04-28 23:26:14
Raw Data from RIRs/APIs	10	11	2025-04-28 23:40:38
SSL Certificate - Raw Data	13	24	2025-04-28 23:25:07
Search Engine Web Content	1	1	2025-04-28 23:40:39
Similar Domain	5	5	2025-04-29 00:35:25
Similar Domain - Whois	2	4	2025-04-29 00:35:38
URL (Purely Static)	4	4	2025-04-28 23:44:22
Web Content	3	29	2025-04-28 23:44:22
Web Content Type	2	29	2025-04-28 23:44:22
Web Server	2	47	2025-04-28 23:44:27

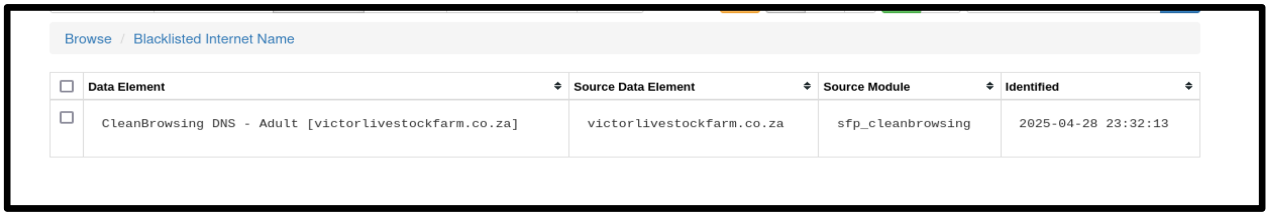
Did you know SpiderFoot also has a CLI? Check out our asciinema tutorials on how to use it.

Fig. 2 Results for the domain “victorlivestockfarm.co.za”



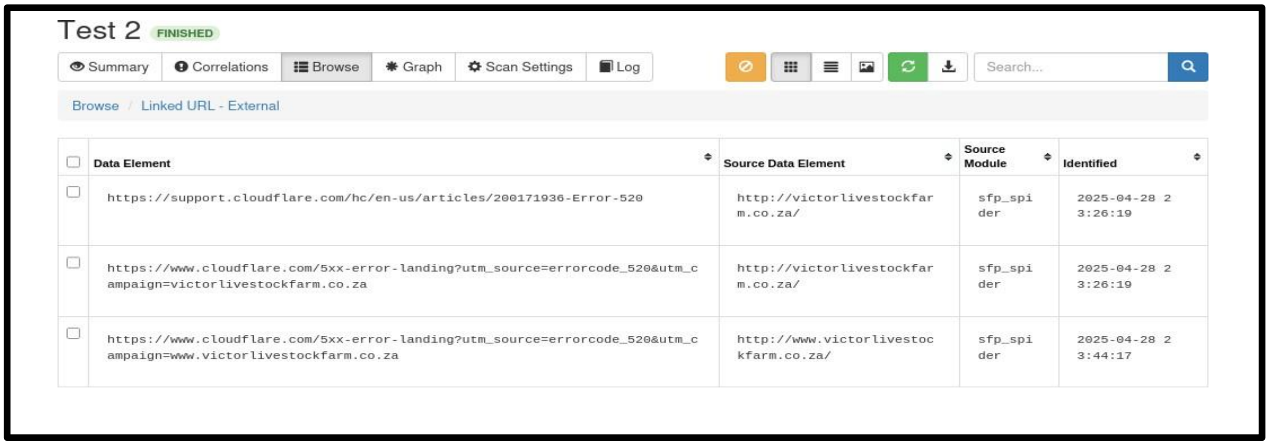
Correlation	Risk	Data Elements
Multiple failure HTTP codes found at victorlivestockfarm.co.za	LOW	48
Outlier hostname found: www.victorlivestockfarm.co.za	INFO	4
Outlier web server found: Apache	MEDIUM	2

Fig. 3 Correlations Results for the domain”victorlivestockfarm.co.za”



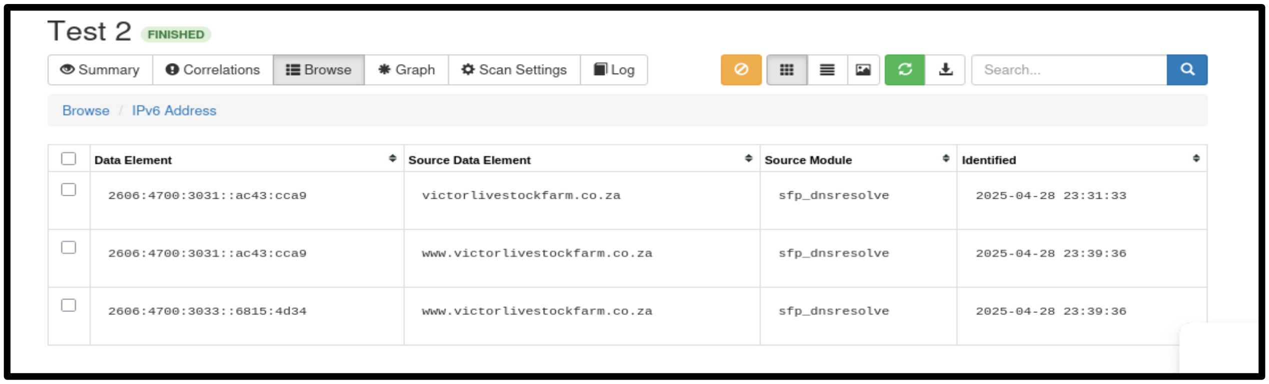
□	Data Element	Source Data Element	Source Module	Identified
□	CleanBrowsing DNS - Adult [victorlivestockfarm.co.za]	victorlivestockfarm.co.za	sfp_cleanbrowsing	2025-04-28 23:32:13

Fig. 4 Blacklisted Internet name of that domain



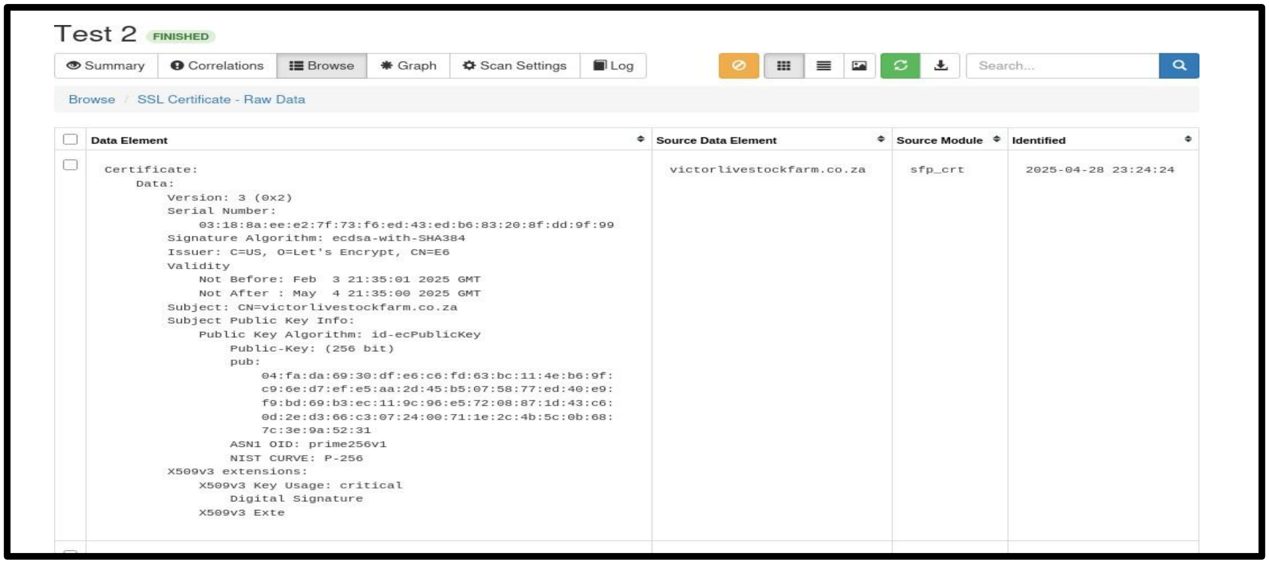
□	Data Element	Source Data Element	Source Module	Identified
□	https://support.cloudflare.com/hc/en-us/articles/200171936-Error-520	http://victorlivestockfarm.co.za/	sfp_spider	2025-04-28 23:26:19
□	https://www.cloudflare.com/5xx-error-landing?utm_source=errorcode_520&utm_campaign=victorlivestockfarm.co.za	http://victorlivestockfarm.co.za/	sfp_spider	2025-04-28 23:26:19
□	https://www.cloudflare.com/5xx-error-landing?utm_source=errorcode_520&utm_campaign=www.victorlivestockfarm.co.za	http://www.victorlivestockfarm.co.za/	sfp_spider	2025-04-28 23:44:17

Fig. 5 Internal Linked URL present in that domain



□	Data Element	Source Data Element	Source Module	Identified
□	2606:4700:3031::ac43:cca9	victorlivestockfarm.co.za	sfp_dnsresolve	2025-04-28 23:31:33
□	2606:4700:3031::ac43:cca9	www.victorlivestockfarm.co.za	sfp_dnsresolve	2025-04-28 23:39:36
□	2606:4700:3033::6815:4d34	www.victorlivestockfarm.co.za	sfp_dnsresolve	2025-04-28 23:39:36

Fig. 6 IPv6 Addresses for that Domain



Data Element	Source Data Element	Source Module	Identified
(certificate dump below)	victorlivestockfarm.co.za	sfp_crt	2025-04-28 23:24:24

```
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            03:18:8a:ee:e2:7f:73:f6:ed:43:ed:b6:83:20:8f:dd:9f:99
        Signature Algorithm: ecdsa-with-SHA384
        Issuer: C=US, O=Let's Encrypt, CN=E6
        Validity
            Not Before: Feb  3 21:35:01 2025 GMT
            Not After : May  4 21:35:00 2025 GMT
        Subject: CN=victorlivestockfarm.co.za
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (256 bit)
                pub:
                    04:fa:da:69:30:df:e6:c6:fd:63:bc:11:4e:b6:9f:
                    c9:6e:d7:ef:e5:aa:2d:45:b5:07:58:77:ed:40:e9:
                    f9:bd:69:b3:ec:11:9c:96:e5:72:08:87:1d:43:c6:
                    0d:2e:d3:66:c3:07:24:00:71:1e:2c:4b:5c:0b:68:
                    7c:3e:9a:52:31
                ASN1 OID: prime256v1
                NIST CURVE: P-256
        X509v3 extensions:
            X509v3 Key Usage: critical
                Digital Signature
            X509v3 Exte
```

Fig. 7 SSL Certificate Data of the Domain

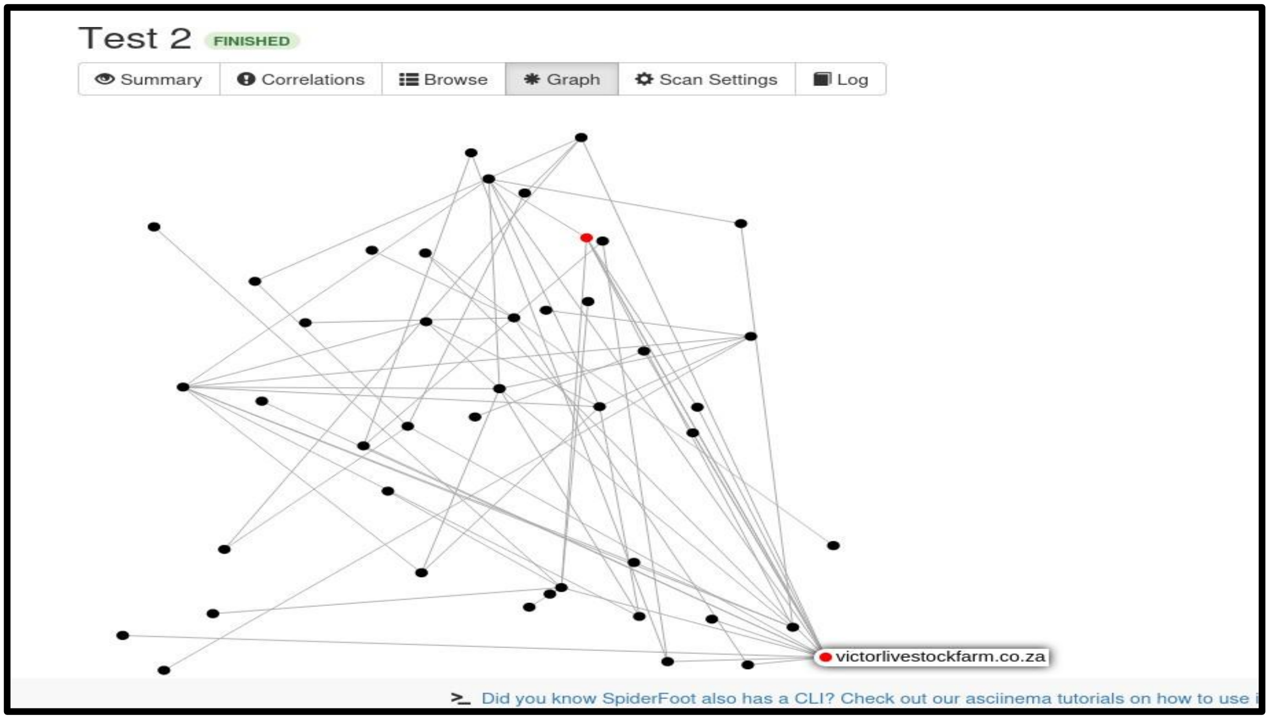

Test 2 FINISHED
Summary
Correlations
Browse
Graph
Scan Settings
Log
victorlivestockfarm.co.za
>_ Did you know SpiderFoot also has a CLI? Check out our asciinema tutorials on how to use i

Fig. 8 Graph result for the Domain.