OSINT-Based Threat Intelligence: Investigating Leaked Data on the Dark Web

Dr. Mukesh Patidar
Professor, Cybersecurity Department (B-Tech)
Parul University
Vadodara, Gujarat
Email: mukesh.patidar34885@paruluniversity.ac.in

Kasani Vignesh Kumar
Cybersecurity Department (B-Tech)
Parul University
Vadodara, Gujarat
Email: 210303126122@paruluniversity.ac.in

***Abstract*—The Dark Web has become a hotspot for cybercrime, serving as a market for stolen credentials, financial data, and sensitive corporate information. It poses an emerging threat for organizations to identify and counter threats that are created from leaked data, as cybercriminals utilize advanced anonymization tools and encryption in an effort to remain anonymous to law enforcement. Open-Source Intelligence (OSINT) has been a motivating factor for cybersecurity researchers to track, analyze, and assess such threats on the basis of publicly available information as well as automated reconnaissance techniques.**

This study paper examines OSINT-driven threat intelligence processes to analyze leaked data on the Dark Web. The paper explores the processes through which cybersecurity analysts use tools such as Maltego, SpiderFoot, and Scrapy in order to monitor Dark Web markets, forums, and hidden sites. The paper explores data collection methods, legal and ethical issues, and methods of evading cybercriminal detection. The study also presents real-life case studies of data breaches, breaking down the patterns in cybercrime attacks and the type of information that is most often leaked.

The findings of this research provide information on the extent and impact of leaked data, establishing trends in cybercrime activity and offering countermeasures. The study highlights the importance of proactive monitoring, real-time processing, and automation in OSINT-driven threat intelligence. Future research will include the integration of AI-driven threat intelligence systems to enhance detection rates and automate Dark Web investigations, thereby improving cybersecurity defenses for organizations and government agencies.

***Index Terms*—OSINT (Open-Source Intelligence),Threat Intelligence,Leaked Data Analysis**

I. Introduction

Cybercrime has expanded exponentially, and more cases of data breaches, ransomware, and identity theft are on the increase. Organizations across all sectors, including finance, healthcare, and government, are continuously attacked by cybercriminals after sensitive information. The Dark Web, an invisible part of the internet that is reached via anonymizing networks such as Tor, is a black market where stolen credentials, corporate data, and financial information are sold.

Traditional cybersecurity protocols cannot track and respond to such threats since the cybercriminals employ anonymity and encryption. Open-Source Intelligence (OSINT), however, has proved to be an effective approach to investigate, track, and analyze information leaks from Dark Web sources. OSINT exploits publicly accessible data, web scraping, and network analysis to spot cyber threats, track bad actors, and enable threat intelligence activities.

A. *Problem Statement*

Dark Web leaked information poses severe cybersecurity threats. Despite the presence of OSINT tools, there are no established frameworks for effectively analyzing leaked information. Additionally, legal and ethical considerations limit the way organizations can collect and analyze Dark Web intelligence.

B. *Research Objectives*

This study aims to:

- Examine OSINT techniques for investigating leaked data on the Dark Web.
- Identify patterns in cybercriminal marketplaces and ransomware operations.
- Evaluate the effectiveness of OSINT tools such as Maltego, SpiderFoot, and Scrapy.
- Identify patterns in cybercriminal marketplaces and ransomware operations.
- Discuss ethical and legal considerations in Dark Web forensics.

II. Background & Related Work

A. *Understanding the Dark Web*

The internet is commonly categorized into three layers:

- **Surface Web:** The publicly accessible part of the internet indexed by search engines (e.g., Google, Bing).
- **Deep Web:** The hidden part of the web that requires authentication or special permissions to access (e.g., databases, academic journals, cloud storage).
- **Dark Web:** A segment of the Deep Web intentionally hidden and accessible only through specialized anonymization networks like Tor (The Onion Router), I2P (Invisible Internet Project), and Freenet.

B. *Leaked Data and Breaches*

- Cybercriminals sell stolen credentials, corporate databases, and financial information. Leaked data sources include **BreachForums**, Telegram groups, and ransomware leak sites.
- Cybercriminals sell stolen credentials, corporate databases, and financial information. Leaked data sources include **BreachForums**, Telegram groups, and ransomware leak sites.
- Ransomware groups such as **LockBit**, **Conti**, and **BlackCat** leak sensitive data if victims refuse to pay ransom.

III. Methodology

A. *OSINT Tools and Techniques*

1) ***Maltego***:

- A graph-based tool for mapping cybercriminal networks and analyzing leaked data relationships.

2) ***SpideFoot***:

- Automates reconnaissance by gathering metadata, domain intelligence, and leaked credentials.

3) ***sacpy***:

- A Python-based web scraping framework used for extracting data from Dark Web marketplaces.

IV. Case Study: Investigating a Real-World Data Leak

A. ***Identifying a Leaked Database***

- Using OSINT tools, we identify a leaked database from a major ransomware leak site.

B. ***Extracting and Analyzing Data***

- The extracted data includes emails, hashed passwords, and financial records. Hash cracking techniques help determine password vulnerabilities.

C. ***Security Implications***

- Leaked data increases risks of identity theft, financial fraud, and corporate espionage.

V. Challenges & Limitations

A. ***Legal and Ethical Concerns***

- Tracking Dark Web leaks involves ethical dilemmas in data collection and legal boundaries for accessing sensitive information.

B. ***Accuracy and Validity of OSINT Datas***

- Dark Web data is often staged leaks and created content.

C. ***Evasion Methods for Cybercrime***

- Encryption, anonymizers, and forum restriction are employed by threat actors to evade OSINT detection.

D. ***Limitations of Current OSINT Tools***

- While OSINT tools provide insights, they lack real-time intelligence and AI-driven automation.

VI. Findings & Insights

A. ***Key Trends in Leaked Data***

- Increased use of ransomware leaks, credential stuffing, and dark marketplaces.

B. ***Trends in Threat Actor Activities***

- Tor forums, Telegram channels, and Monero (XMR) transactions are favored by threat actors.

C. ***Best Practices for Threat Intelligence Analysts***

- Continuous tracking of Dark Web leaks.
- Adherence to laws during OSINT research.
- Integration with AI for automation.

VII. Conclusion

1) ***Summary of Key Findings***:

- OSINT is essential for Dark Web forensics but faces challenges in accuracy, ethics, and cybercriminal evasion.

2) ***Recommendations***:

- Organizations should invest in AI-driven OSINT and collaborate with law enforcement agencies.

References

[1] Using Dark Web for OSINT Investigations
[2] George Kalpakis,OSINT and the Dark Web
[3] Top 7 OSINT Tools for Dark Web Investigations
[4] OSINT Tools and Techniques for Unmasking Dark Web Operations
[5] Springer International Publishing , OSINT and the Dark Web
[6] Meenakshi Munjal , ETHICAL HACKING: AN IMPACT ON SOCIETY(2013)