

An integrated framework for data breach on the dark web in brand monitoring data hunting

Siti Arpah Ahmad¹, Muhammad Al'Imran Mohd Khairuddin²,
Nor Shahniza Kamal Bashah¹, Nurul Aishah Ab Raman³

¹School of Computing Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia

²44th Floor, Kuala Lumpur Main Branch, Maybank Tower, 100, Kuala Lumpur, Malaysia

³Faculty of Communication and Media Studies, Universiti Teknologi MARA, Shah Alam, Malaysia

Article Info

Article history:

Received May 31, 2024

Revised Jan 10, 2025

Accepted Mar 3, 2025

Keywords:

Dark web

Data breach

Data safety

Personal identifiable information

Sensitive data

ABSTRACT

In today's digital landscape, data breaches pose a substantial threat, with the dark web serving as a prevalent platform for malevolent actors to perpetrate such incidents. Currently, security analysts use various tools to solve the problem, which is very time-consuming. This paper introduces a novel framework that integrates data breach monitoring within the dark web, focusing on brand monitoring and data hunting. The framework starts from the scraping process and continues with the utilization of the Splunk dashboard. The dashboard provides an exhaustive overview of data breaches related to brands for both manual inquiries and rule-based detection mechanisms. The framework comprises five phases: data sourcing, data collection, integration, monitoring, and visualization. The visualization phase encompasses alert generation, notification mechanisms, and reporting functionalities. Moreover, the monitoring phase provides real-time surveillance, advanced search capabilities, brand monitoring, and threat intelligence integration. The integration phase involves security information and event management (SIEM) systems and security orchestration, automation, and response (SOAR) systems. This paper's result contributes to enhancing the National Institute of Standards and Technology (NIST) cybersecurity framework, offering a comprehensive solution to the data breaches challenge within the dark web and the frontiers of knowledge and security practices.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Siti Arpah Ahmad

School of Computing Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA

40450 Shah Alam, Selangor, Malaysia

Email: arpah340@uitm.edu.my

1. INTRODUCTION

In the context of the rapidly evolving digital environments, the potential occurrence of data breaches [1]–[3] has become increasingly common. These breaches, which entail unauthorized entry into confidential data [4], present a significant peril to the integrity and security of information. In recent years, the escalating dependence on digital systems and the advancing complexity of cyberattacks have led to notable uneasiness surrounding data breaches. Since 2020, around 10.88 billion personal information (full name, email addresses, sexual orientation, chat and email transcripts, IP address and payment logs) has been stolen by cybercriminal [5]. These breaches can lead to significant consequences for both persons and organizations, including financial

disadvantage, harm to reputation, legal consequences, and compromised personal privacy [6]. The findings by Verizon data breach investigation report [5] observe that a significant majority, precisely 76%, of data breaches pertain to financial matters. Political motivations can also drive certain breaches, as shown when government-owned public data is exposed to express dissent or advocate for social justice. The exclusive driving force behind other breaches is the pursuit of financial benefit. Individuals addicted to seeking excitement may be the culprits behind security breaches, as they are motivated by a desire to test the limits of their damaging hacking skills.

The fears have been further heightened by the rise of the dark web [7], [8], a hidden Internet sector that facilitates various cybercriminal operations, such as the illegal exchange of compromised data. The dark web refers to content unindexed by ordinary search engines, necessitating specialized software or prior authorization for access [9]. The content accessible on the dark web hosts a specific segment of the Internet, commonly called the darknet [10], [11]. Access to this specific portion of the Internet is limited to certain web browsers or specified network setups.

Given this potential danger, the concept of brand monitoring in cybersecurity is emerging. However, previous works [12]–[14] have limitations on how the data breach affects companies' brands without suggesting or introducing methods to solve the problem. This work defines brand monitoring as an early threat detection to identify cyber threats. Examples of the threats are phishing attacks, fake websites and social media scams. These threats are breaches that manipulate company brands, customers and employees' data for illegal purposes. Brand monitoring includes collecting and analyzing content from various digital channels, detecting brand risks as they appear and allowing cybersecurity teams to remediate instances of brand abuse before they negatively impact the customer experience. In today's environment, where data breaches are increasingly common, proactively identifying and mitigating potential threats is essential. Current early threat detection methods rely on cybersecurity analysts' manual analysis of raw network traffic, utilizing frameworks such as the MITRE ATT&CK framework [15].

Furthermore, existing studies addressing data breach issues often overlook the need for real-time monitoring and analysis and the integration of advanced technologies, as noted in [16]–[21]. This oversight leads to inefficiencies in addressing data breach challenges. To tackle this problem, the integrated framework proposed in this research aims to incorporate threat intelligence. By utilizing data visualization on the dashboard, the framework facilitates real-time surveillance and provides a notification mechanism for any incidents.

This research delves into the cyberspace environment of Malaysia, which distinguishes the distinctive challenges and opportunities. The increasing frequency and complexity of cyber threats [22] necessitate a comprehensive understanding of the nature of data breaches [23], [24], which holds significant importance. These breaches highlight the importance of proactive monitoring as a form of defense [25] as human mistakes, social engineering, and oversights in cyber hygiene often cause them. This study aims to establish a connection between cybersecurity and the human component, emphasizing their interrelationship.

The main objective of this study is to develop a customized brand monitoring methodology specifically designed for the digital landscape in Malaysia. By utilizing data analytics and demonstrating expertise in cybersecurity, this project aims to reveal potential hazards within the hidden depths of the dark web. By engaging in this practice, the objective is to equip organizations with the necessary resources to detect potential security breaches [26] and understand the tactics employed by cyber attackers. Situated at the convergence of digital advancement and security complexities, this research is beyond theoretical discussion. This technology has the potential to provide practical insights that can strengthen the ability of both organizations and individuals to withstand and overcome imminent threats. The emphasis on Malaysia's cyberspace highlights the importance of implementing proactive cybersecurity measures and providing stakeholders with the essential knowledge to protect their digital domain. This study aims to explore the dark web, specifically its associations with data breaches, examine brand monitoring methods and develop a complete framework for enhancing Malaysia's digital security measures. The primary objective of this study is to enhance comprehension of data breach risks and implement measures to protect digital environments.

This study enhances the detection phase in the National Institute of Standards and Technology (NIST) Cybersecurity Framework. NIST framework is the leading cybersecurity standard [27] and is referred to by many. One of the elements in the detect phase is data breach, and this work introduces a new solution related to data breaches. It combines the data breach strategy with the brand monitoring concept in the cybersecurity domain. The phase in the new framework consists of data collection, integration, monitoring capabilities, open-source utilization, and data visualization, promises a holistic approach towards combating the evolving security threats.

In summary, this research explores the evolving landscape of data breaches and brand monitoring in the ever-changing digital realm. Section 1, the introduction, has set the stage by highlighting the escalating Threat of data breaches, their far-reaching consequences, and the rising prominence of the dark web as a breeding ground for cybercriminal activities. It has also underscored the significance of proactive brand monitoring, and the unique challenges and opportunities present in Malaysia's cyberspace. Section 2 introduces the novel brand monitoring framework proposed in this study, shedding light on detecting data breach events

[28] across sectors and locations and their correlation. The introduction of the novel brand monitoring framework signifies a significant contribution to cybersecurity, offering an innovative approach to safeguarding brand integrity in the digital age. Next in section 3 is the methodology, which delves into the systematic framework employed for this research, encompassing data collection, integration, monitoring capabilities, open-source utilization, and data visualization, which includes alerts, dashboards, reports, and notifications. Section 4 contains the results and discussion, which provides a comprehensive analysis of the findings and comparisons with the existing work in the market. Finally, Section 5 is the conclusion, which will synthesize these insights, emphasizing the practical implications of this research in enhancing brand security and resilience in the face of emerging cyber threats.

2. PROPOSED FRAMEWORK OF DARK WEB BRAND MONITORING (DWBM) SYSTEM

The framework for detecting and monitoring data breaches comprises essential components that cooperate to deliver a comprehensive cybersecurity solution. Figure 1 shows the proposed integrated framework for data breaches on the dark web [29], [30] in brand monitoring data hunting. The collection of probable data breach signs [31] facilitates a variety of data sources, including dark web forums [32], underground channels, and the messaging platform Telegram. Python scripts utilize the purpose of extracting data effectively, while a Telegram bot is employed to simplify the ingestion of this data into the Splunk platform. In the context of Splunk, the data undergoes organized transformation, facilitating efficient processing and analysis.

Real-time monitoring, incorporating rule-based detection [33], [14], constitutes a vital characteristic that rapidly notifies relevant stakeholders of potentially suspicious activities. Open-source tools utilize cost-effectiveness and optimize the utilization of resources. Data visualization plays a crucial role in various aspects, including providing real-time warnings, creating visual dashboards, generating comprehensive reports, and distributing notifications via Discord webhooks and Telegram channels. The presented comprehensive framework provides organizations with the tools to protect digital assets and effectively address emerging cyber risks promptly.

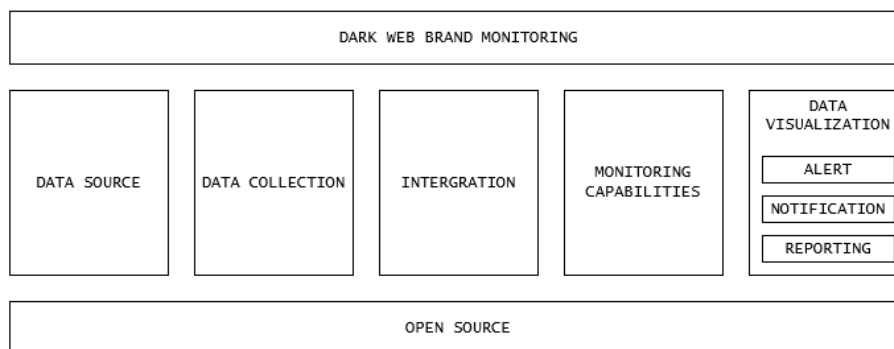


Figure 1. The proposed data breach framework

3. RESEARCH METHODS

The diagram illustrated in Figure 2 offers a visual representation of the operational workflow related to the system's functionality in the context of real-time detection of data breaches and subsequent alerting. The first phase involves using a Python script to methodically collect and retrieve relevant information and data from Internet platforms on the dark web. After the first data-collecting procedure, the obtained data is ingested into the Splunk platform. The Splunk platform enables a thorough analysis process, which involves using custom rules that are carefully adjusted to successfully handle the distinct local brand names that are popular in Malaysia. The preceding regulations' design is to streamline the process of conducting daily searches by a predetermined schedule. If the data fails to meet the predetermined criteria for a data breach, the system will discard it promptly, leading to the completion of the complete procedure sequence. Once a data breach event identifies threats positively, the pre-established protocols within the Splunk system are immediately initiated. The initiation of Splunk prompts a series of actions in which notifications are disseminated to users through the utilization of a Discord webhook integration and a designated Telegram channel. This technique enhances the timely and efficient reporting of incidents, enhancing the overall efficacy of incident management.

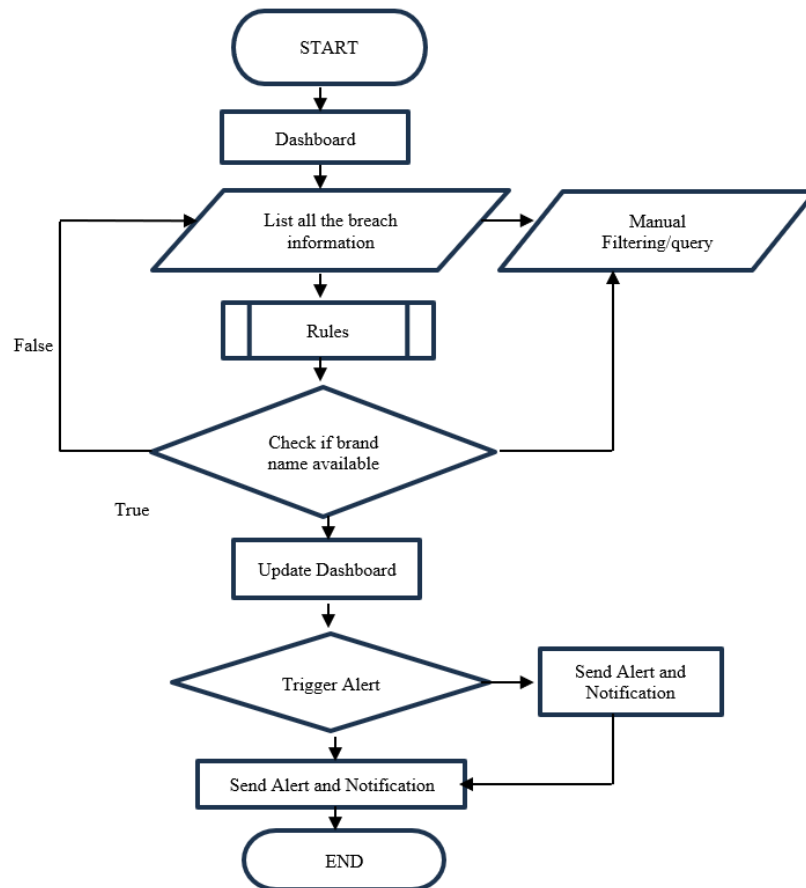


Figure 2. Process flow of the brand monitoring framework (real time detection)

4. RESULTS AND DISCUSSION

Results are illustrated using several figures. Figure 3 is the dark web brand monitoring (DWBM) system interface. Figure 4 is the real-time alert process script for the alert notification in discord (as seen in Figure 5) and telegram (as seen in Figure 6). Table 1 shows the testing results of data breach trends in Malaysia and Table 2 displays the comparison with other available tools online.

4.1. DWBM system

The DWBM system represents a significant advancement in the realm of cybersecurity. Figure 3 shows the interface of the DWBM platform. The input for DWBM is time range, country and brand. These inputs act as a filter to search for potential data breach events. These events are displayed on the dashboard based on regions and sectors. There are three (3) regions: all regions or all over the world, Malaysia and others. As for sectors, there are four (4) sectors: government, education, banking and others. With its extensive coverage of data sources, including the Dark Web, stolen data underground forums, and channels, DWBM provides a broader view of potential brand threats. Its comprehensive set of features sets it apart, including real-time monitoring, advanced search functionalities, brand monitoring, threat intelligence, and offensive security features. Its seamless integration with security information and event management (SIEM) and security orchestration, automation, and response (SOAR) platforms allows for efficient collaboration and automation in the incident response process. Moreover, DWBM's open-source nature eliminates licensing fees, making it an asset for NGOs, businesses, and community initiatives. In an increasingly digital and interconnected world, DWBM emerges as a powerful tool to bolster cybersecurity, protect brands, and contribute positively to the broader community. Figure 4 shows the DWBM platform's real-time alerting process. When the rules triggered contain information about Malaya or Malaysia, notifications will be sent to Discord and Telegram. Figures 5 and 6 are examples of alerts related to Malaya Signature detected and notified using Discord and Telegram. The display in Discord shows the level of severity and the link for the system administrator to do further investigation. The message in Telegram contains the alert name, severity, and link for the system administrator to take further action.

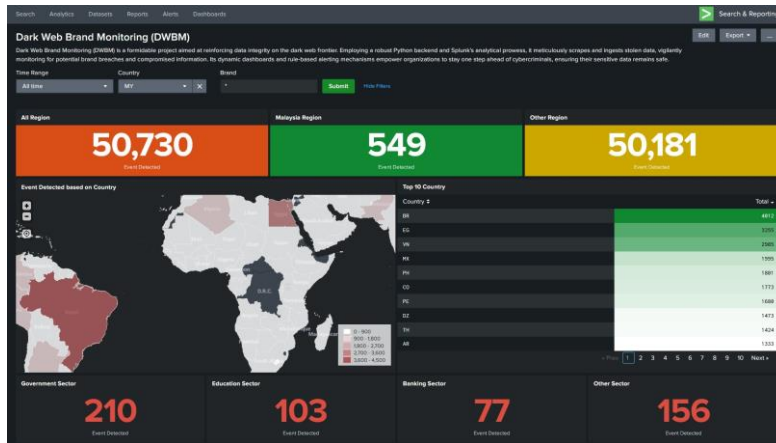


Figure 3. The interface of the DWBM platform

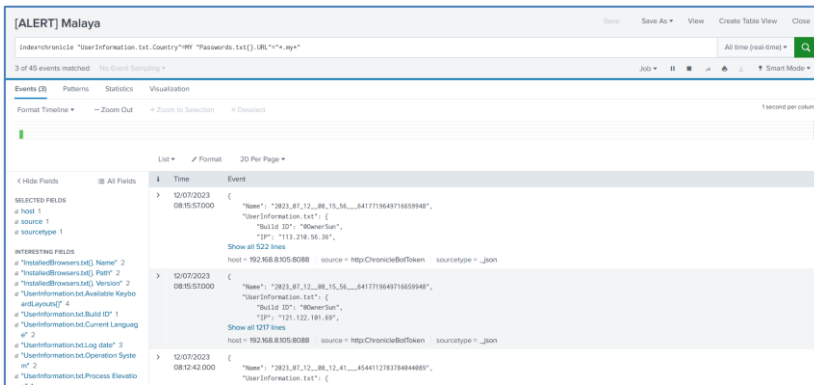


Figure 4. Example of the real-time alerting process

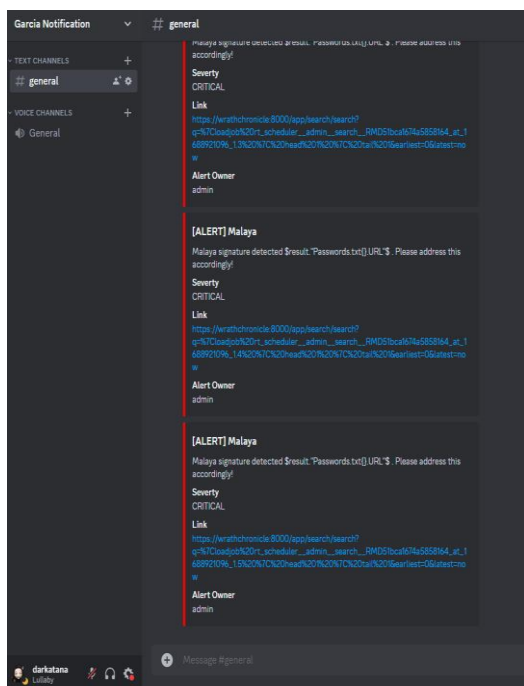


Figure 5. Example of alert notification in Discord



Figure 6. Example of alert notification in Telegram

Table 1. Most locations related to data breach based on the data findings

Location	Count	Percent
Shah Alam, Selangor	20	19.417%
Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur	9	8.738%
Alor Setar, Kedah	6	5.825%
Kota Kinabalu, Sabah	5	4.854%
Bayan Lepas, Pulau Pinang	4	3.883%
Melaka, Melaka	4	3.883%
Kota Bharu Kelantan	3	2.913%
Sitiawan, Perak	3	2.913%
Sungai Petani, Kedah	3	2.913%
Tasek Gelugor, Pulau Pinang	3	2.913%

4.2. Comparison with other tools available online

In order to assess the strengths and weaknesses of the proposed DWBM, a comprehensive analysis was conducted against three leading competitors: Dark Owl, Digital Shadows, and Flashpoint as depicted in Table 2. The comparison covers key features including data sources, monitoring capabilities, integration options, data visualization, data collection methods and subscription models. This detailed analysis was performed not only to understand how the proposed platform compares to existing solutions but also to identify areas where the proposed DWBM tool offers superior functionality, flexibility, and affordability. The goal is to provide businesses and security professionals with an advanced, open-source alternative that meets their brand monitoring and cybersecurity needs in a comprehensive and cost-effective manner.

Table 2. The comparisons of a few DWBM platforms

	Dark Owl	Digital Shadows	Flashpoint	DWBM (Proposed)
Data source covered	Wide range of data sources	Dark web and digital risk	Underground forums	Dark web, stolen data, underground forum, and channel
Monitoring capabilities	Real-time monitoring, advanced search	Brand monitoring, risk scoring	Threat intelligence, incident response	Real-time monitoring, advanced search, brand monitoring, threat intelligence, offensive security
Integration options	Integration with SIEM, TIP, SOAR	Integration with SIEM, TIP	Integration with SIEM, TIP	Integration with SIEM, SOAR
Data visualization	Graph, chart, dashboard	Interactive dashboard, Customizable report	Visualizations, relationship mapping	Customizable dashboard, visualization, alerts, notifications, and report
Data collection method	Web scrapping, API integrations	Web crawling and data scrapping	Web crawling, OSINT collection	Data scrapping, API integrations, web crawling
Subscription	Subscription-based	Subscription-based	Yes, Custom pricing	No, Open Source

5. CONCLUSION

In conclusion, this study contributes to understanding how cybersecurity should integrate brand monitoring strategies to protect brand reputation, customer data, and the proposed Data Breach Monitoring Framework. The generalizations derived from the research findings underscore the pivotal role of proactive cybersecurity measures in mitigating data breach risks, fostering brand trust, and maintaining a strong brand presence in the dynamic digital landscape.

ACKNOWLEDGMENTS

The authors would like to thank the School of Computing Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia, for supporting the publication of this paper.

FUNDING INFORMATION

This work is funded by Universiti Teknologi MARA, 40450 Shah Alam, Malaysia.

AUTHOR CONTRIBUTIONS STATEMENT

Siti Arpah Ahmad, Muhammad Al'Imran Mohd Khairuddin, and Nor Shahniza Kamal Bashah conceived the overall concept of the research and the methodology. Muhammad Al'Imran Mohd Khairuddin developed the DWBM system, carried out the investigation, and conducted the experiments. Siti Arpah Ahmad, Nor Shahniza Kamal Bashah, and Nurul Aishah Ab Raman verified the system and the results. Formal analysis

is done by Siti Arpah Ahmad and Nor Shahniza Kamal Bashah. Data curation is done by Siti Arpah Ahmad, Muhammad Al’Imran Mohd Khairuddin, and Nor Shahniza Kamal Bashah. The original draft was done by Muhammad Al’Imran Mohd Khairuddin. Writing, reviewing, editing, and visualization are done by Siti Arpah Ahmad, Nor Shahniza Kamal Bashah, and Nurul Aishah Ab Raman. Supervision is done by Siti Arpah Ahmad. Project administration is done by Muhammad Al’Imran Mohd Khairuddin. All authors discussed the results and contributed to the final manuscript.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Siti Arpah Ahmad	✓	✓		✓	✓			✓		✓	✓	✓		
Muhammad Al’Imran Mohd Khairuddin	✓	✓	✓			✓		✓	✓	✓	✓		✓	
Nor Shahniza Kamal Bashah	✓	✓		✓	✓			✓			✓			
Nurul Aishah Ab Raman				✓						✓	✓			

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nteraction

R : **R**esources

D : **D**ata Curation

O : **O**riginal Draft

E : **E**valuation & **E**dit

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors declare no conflicts of interest

ETHICAL APPROVAL

This research has applied for an ethics exemption from Universiti Teknologi MARA and has been approved under reference no. 100-KPPIM(PI.9/10/3) (EX/798).

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.




REFERENCES

- [1] H. Saleem and M. Naveed, “SoK: Anatomy of data breaches,” in *Proceedings on Privacy Enhancing Technologies*, vol. 2020, no. 4, pp. 153–174, Oct. 2020, doi: 10.2478/popets-2020-0067.
- [2] J. Li, W. Xiao, and C. Zhang, “Data security crisis in universities: identification of key factors affecting data breach incidents,” *Humanities and Social Sciences Communications*, vol. 10, no. 1, May 2023, doi: 10.1057/s41599-023-01757-0.
- [3] S. Chatterjee, X. Gao, S. Sarkar, and C. Uzmanoglu, “Reacting to the scope of a data breach: The differential role of fear and anger,” *Journal of Business Research*, vol. 101, pp. 183–193, Aug. 2019, doi: 10.1016/j.jbusres.2019.04.024.
- [4] H. Tao *et al.*, “Economic perspective analysis of protecting big data security and privacy,” *Future Generation Computer Systems*, vol. 98, pp. 660–671, Sep. 2019, doi: 10.1016/j.future.2019.03.042.
- [5] D. Gibson and C. Harfield, “Amplifying victim vulnerability: Unanticipated harm and consequence in data breach notification policy,” *International Review of Victimology*, vol. 29, no. 3, pp. 341–365, Sep. 2023, doi: 10.1177/02697580221107683.
- [6] T. J. Holt and A. M. Bossler, *The palgrave handbook of international cybercrime and cyberdeviance*. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-319-78440-3.
- [7] S. Kaur and S. Randhawa, “Dark web: A web of crimes,” *Wireless Personal Communications*, vol. 112, no. 4, pp. 2131–2158, Jun. 2020, doi: 10.1007/s11277-020-07143-2.
- [8] D. Yadav, B. Bhushan, and S. Saxena, “The dark web: A dive into the darkest side of the Internet,” *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3598902.
- [9] F. T. Ngo, C. Marcum, and S. Belshaw, “The dark web: What is it, how to access it, and why we need to study it,” *Journal of Contemporary Criminal Justice*, vol. 39, no. 2, pp. 160–166, May 2023, doi: 10.1177/10439862231159774.
- [10] M. Hatta, “Deep web, dark web, dark net,” *Annals of Business Administrative Science*, vol. 19, no. 6, pp. 277–292, Dec. 2020, doi: 10.7880/abas.0200908a.
- [11] A. S. Beshiri and A. Susuri, “Dark web and its impact on online anonymity and privacy: a critical analysis and review,” *Journal of Computer and Communications*, vol. 07, no. 03, pp. 30–43, 2019, doi: 10.4236/jcc.2019.73004.
- [12] C. A. Makridis, “Do data breaches damage reputation? Evidence from 45 companies between 2002 and 2018,” *Journal of Cybersecurity*, vol. 7, no. 1, Sep. 2021, doi: 10.1093/cybsec/tyab021.
- [13] F. N. Ho, N. Ho-Dac, and J. S. Huang, “The effects of privacy and data breaches on consumers’ online self-disclosure, protection behavior, and message valence,” *Sage Open*, vol. 13, no. 3, Jul. 2023, doi: 10.1177/21582440231181395.
- [14] I. Confente, G. G. Siciliano, B. Gaudenzi, and M. Eickhoff, “Effects of data breaches from user-generated content: A corporate reputation analysis,” *European Management Journal*, vol. 37, no. 4, pp. 492–504, Aug. 2019, doi: 10.1016/j.emj.2019.01.007.



- [15] N. Singh and S. Tripathy, "It's too late if exfiltrate: Early stage Android ransomware detection," *Computers & Security*, vol. 141, Jun. 2024, doi: 10.1016/j.cose.2024.103819.
- [16] X. Zhang, M. M. Yadollahi, S. Dadkhah, H. Isah, D. P. Le, and A. A. Ghorbani, "Data breach: analysis, countermeasures and challenges," *International Journal of Information and Computer Security*, vol. 19, no. 3/4, pp. 402–442, 2022, doi: 10.1504/IJICS.2022.127169.
- [17] H. Hammouchi, O. Cherqi, G. Mezzour, M. Ghogho, and M. El Koutbi, "Digging deeper into data breaches: An exploratory data analysis of hacking breaches over time," *Procedia Computer Science*, vol. 151, pp. 1004–1009, 2019, doi: 10.1016/j.procs.2019.04.141.
- [18] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects," *Annals of Data Science*, vol. 10, no. 6, pp. 1473–1498, Dec. 2023, doi: 10.1007/s40745-022-00444-2.
- [19] D. Molitor, A. Saharia, V. Raghupathi, and W. Raghupathi, "Exploring the characteristics of data breaches: A descriptive analytic study," *Journal of Information Security*, vol. 15, no. 02, pp. 168–195, 2024, doi: 10.4236/jis.2024.152011.
- [20] P. Mayer *et al.*, "Awareness, intention, (in) action: Individuals' reactions to data breaches," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 5, pp. 1–53, Oct. 2023, doi: 10.1145/3589958.
- [21] H. Teymourlouei and V. E. Harris, "Preventing data breaches: Utilising log analysis and machine learning for insider attack detection," in *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2022, pp. 1023–1028.
- [22] Y. Fang, Y. Guo, C. Huang, and L. Liu, "Analyzing and identifying data breaches in underground forums," *IEEE Access*, vol. 7, pp. 48770–48777, 2019, doi: 10.1109/ACCESS.2019.2910229.
- [23] D. Chen, M. M. Chowdhury, and S. Latif, "Data breaches in corporate setting," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Oct. 2021, pp. 01–06. doi: 10.1109/ICECCME52200.2021.9590974.
- [24] S. Trabelsi, "Monitoring leaked confidential data," in *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Jun. 2019, pp. 1–5. doi: 10.1109/NTMS.2019.8763811.
- [25] Z. Mohammed, "Data breach recovery areas: an exploration of organization's recovery strategies for surviving data breaches," *Organizational Cybersecurity Journal: Practice, Process and People*, vol. 2, no. 1, pp. 41–59, Apr. 2022, doi: 10.1108/O CJ-05-2021-0014.
- [26] M. R. Uddin, S. Akter, and W. J. T. Lee, "Developing a data breach protection capability framework in retailing," *International Journal of Production Economics*, vol. 271, May 2024, doi: 10.1016/j.ijpe.2024.109202.
- [27] H. Taherdoost, "Understanding cybersecurity frameworks and information security standards - a review and comprehensive overview," *Electronics*, vol. 11, no. 14, Jul. 2022, doi: 10.3390/electronics11142181.
- [28] H. Alnabulsi and R. Islam, "Identification of illegal forum activities inside the dark net," in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, Dec. 2018, pp. 22–29. doi: 10.1109/iCMLDE.2018.00015.
- [29] R. Liggett, J. R. Lee, A. L. Roddy, and M. A. Wallin, "The dark web as a platform for crime: An exploration of illicit drug, firearm, CSAM, and cybercrime markets," in *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, Cham: Springer International Publishing, 2020, pp. 91–116. doi: 10.1007/978-3-319-78440-3_17.
- [30] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso, and S. N. Mbaye, "Web scraping: State-of-the-art and areas of application," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 6040–6042. doi: 10.1109/BigData47090.2019.9005594.
- [31] R. Naik and M. N. Gaonkar, "Data leakage detection in cloud using watermarking technique," in *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2019, pp. 1–6. doi: 10.1109/ICCCI.2019.8821894.
- [32] R. Avila, R. Khoury, R. Khoury, and F. Petrillo, "Use of security logs for data leak detection: A systematic literature review," *Security and Communication Networks*, vol. 2021, pp. 1–29, Mar. 2021, doi: 10.1155/2021/6615899.
- [33] F. Lin *et al.*, "Linking personally identifiable information from the dark web to the surface web: A deep entity resolution approach," in *2020 International Conference on Data Mining Workshops (ICDMW)*, Nov. 2020, pp. 488–495. doi: 10.1109/ICDMW51313.2020.00072.

BIOGRAPHIES OF AUTHORS






Siti Arpah Ahmad    has received a Ph.D. in electrical engineering (2017) from Universiti Teknologi MARA. She is a senior lecturer at the college of computing, informatics, and mathematics at Universiti Teknologi MARA, Malaysia. She is currently teaching computer projects for degree students in the computer networking degree and the cybersecurity and digital forensics master's degree in the college of computing, informatics, and mathematics at Universiti Teknologi MARA, Malaysia. Her current research interests are in cybersecurity security, digital forensics, image processing, and machine learning. She can be contacted at email: arpah340@uitm.edu.my.






Muhammad Al Imran   is a cybersecurity and digital forensic specialist known for his expertise in the field. He graduated with a degree in information technology (network technology) from Universiti Islam Selangor, Malaysia, in 2020 and an M.Sc. degree in science in Cybersecurity and Digital Forensic at Universiti Teknologi MARA, Malaysia in 2023. Imran's professional journey has taken him to Netbytesec, where he serves as a Senior Security Analyst, specialising in digital forensics and incident response. His work involves investigating and mitigating digital security breaches. Imran is also a dedicated researcher with a focus on various cybersecurity domains such as the dark web, credential leakage, forensic techniques, SIEM, computer forensic analysis, memory forensics, detection analysis, and malware analysis. He actively contributes to the cybersecurity community and can be reached at imran_khairuddin@yahoo.com.



Nor Shahniza Kamal Bashah    holds a Ph.D degree in Telematics from Norwegian University of Science and Technology, Norway. She is an Associate Professor at the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Malaysia. Her research interest is involved in the field of mobile and wireless communication and semantic web. She can be contacted at email: shahniza@uitm.edu.my.



Nurul Aishah Ab Raman    is currently serving as an academic at the Faculty of Communication and Media Studies in UiTM. Certified Professional Technologist (Ts.) from Malaysian Board of Technologist (MBOT) in Arts and Creative Multimedia. Her areas of interest are political communication, digital media, strategic communication, and international relations. She obtained her MA in strategic and defense studies from Universiti Malaya, in 2014. Prior to joining academia, she worked at Media Prima Berhad at the News and Current Affairs Department (NCA) as broadcast journalist specializing in political and governmental issues. She can be contacted at email: aishah7502@uitm.edu.my.