Tiako P.F. (ed) *Competitive Tools, Techniques, and Methods*. Chronicle of Computing. OkIP. CAIF24#15
© 2024 Oklahoma International Publishing https://doi.org/10.55432/978-1-6692-0007-9_4

Exploring Vulnerabilities: An In-Depth Analysis of Security Weaknesses Leading to Dark Web Exposure and Improved Security Defenses Using Artificial Intelligence (AI) & Machine Learning (ML)

Joseph Squillace, Justice Cappella, Andrew Sepp

Penn State University, jms10943@psu.edu, Schuylkill Haven,17972, United States
Penn State University, jlc6927@psu.edu, Schuylkill Haven, 17972, United States
Penn State University, afs6479@psu.edu, Schuylkill Haven, 17972, United States

ABSTRACT

In modern society, demand and value of stolen personally identifiable information (PII) has risen dramatically. A significant portion of this information can be found within the Dark Web. Growing exposure and exchange of PII is largely a resultant of inadequate organizational security measures. This investigation uses Artificial Intelligence (AI) and Machine Learning (ML) to explore the underlying causes of data breach events to examine how adaptation of more robust security strategies can mitigate PII exposure.

Keywords: Artificial Intelligence, Cybersecurity, Dark Web, Cybercrime, Personal Identifiable Information (PII)

I. INTRODUCTION

The *Dark Web*, described by Liu et al. (2020) as "a collection of illegal and covert platforms that facilitate communication and transactions among cybercriminals" is where a significant portion of illicit *Personal Identifiable Information* (PII) can be found and procured. It is more vital than ever to use Artificial Intelligence (AI) and Machine Learning (ML) as a tool to integrate better security measures to integrate more with the Dark Web, a pivotal platform for the exchange of PII. The continuous inability of organizations to effectively secure and protect PII will only further increase its market demand, particularly in relation to the relative ease with which this sensitive information can be accessed.

Through utilization of case study analysis, we will thoroughly investigate, analyze, and assess the impact of implementing modified security protocols and policies, with a focus on determining their associated effectiveness. The overarching objective of this research study is not only bolster privacy and security measures for mitigating PII exposure, but also make substantial contributions to the extant body of literature across various modalities. By examining inadequate privacy policies, studying the outcomes of previous PII exposure on the dark web, and proposing an alternative set of improved policy guidelines, organizations can further strengthen users' privacy and their personal data.

II. BACKGROUND

Ineffective and inefficient privacy policy is not only a concern for end-users, but also presents substantial data security risks to organizations. Effectively securing end- users' PII allows organizations to directly protect their assets and eliminate the possibility of legal ramifications and consumer lawsuits as a resultant of stolen information during data breach events and cybersecurity attacks. Many businesses possess substandard privacy and security policies; moreover, a significant number of organizations do not implement. This negligence contributes to cyber-attacks and data breaches; facilitating the sale of all Personally Identifiable Information (PII) on the dark web (Thomas et al., 2022). Externally, organizations may be perceived to possess strong security and privacy policies and security personnel, however internally they may not be effectively safeguarding their systems due to outdated, non-compliant, insufficient, and/or incorrect policies that are unable to be properly enforced by the organization. Research done by Moore states "Perfect security is impossible, but even if it were, it would not be desirable" (Moore, 2010). While security is never completely safe, organizations that are not proactive in their security posture are more susceptible to leaking PII, data breach events, and privacy inequality.

This research investigation seeks to explore the underlying causes of data breach events and examine how the adoption of more robust privacy and security strategies can better mitigate PII exposure. Moreover, this study aims to evaluate the consequences of poor privacy and security policies to showcase the potential negative resultant output from those actions; once identified this research will introduce a novel set of enhanced user privacy practices and security controls aimed at mitigating the risk of PII exposure. These enhancement efforts will include, but not limited to, tighter security measures, password policies, security audits, encryption of data, and more.

Through utilization of case study analysis, on publicly disclosed data breach events, we will aim to answer the following questions i) Determine the causes behind the breach event (why it happened) and ii) Establishing methods to prevent future occurrences through implementation of robust privacy and security measures. The research will further explore the Dark Web to examine in detail the value (economic) of the data and the popularity (download interest) of the identified PII stolen and leaked during the security breach event. The overarching objective of this research study is not only bolster privacy and security measures for mitigating PII exposure, but also make substantial contributions to the extant body of literature across various modalities. By examining inadequate privacy policies, studying the outcomes of previous PII exposure on the dark web, and proposing an alternative set of improved policy guidelines, organizations can further

Presented at the 2024 OkIP International Conference
on Artificial Intelligence Frontiers (CAIF)
in Oklahoma City, OK, USA, and Online, on April 3, 2024

strengthen users' privacy and their data. The study will expand upon the traditional Security, Education, Training, and Awareness (SETA) model, isolating the most effective strategies and security measures. Thus, leading to the development of an enhanced education framework and reinforced security guidelines, ultimately aiming to build a more secure cyber infrastructure.

III. Proposed Methodology and Innovation

This research examination will consist of a longitudinal study and deploy a Mixed-Methods Methodological approach. Secondary data will be utilized to better understand how organizational privacy and security policies affect the handling and protection of (PII). Identification of both privacy and security policies that are least effective and/or inadequately implemented is crucial for the data collection process. This investigation will focus on the identification of vulnerabilities within both the privacy and security policies and introduce a novel set of enhanced user privacy practices and security controls aimed at mitigating the risk of PII exposure; specifically in the context of exploitation via the Dark Web. Table 1 provides a comprehensive visual representation of the entire research process, encompassing each stage from the initial data collection to the final analysis of implemented security measures. This research will help offer insights into the effectiveness of current privacy policy strategies and the potential for improvement.

TABLE I. RESEARCH PROCESS

RESEARCH STEP	PROCEDURE(s)
1.Data Collection	Analyzing standards like ISO 27799, ISO 27018, and ISO 27037
2.Policy Analysis	Assessing policy characteristics, comprehensiveness, adaptability to threats, and effectiveness
3.Comparativ e Study	Comparing companies with minimal breach events against those with frequent incidents
4.Dark Web Examination	Using tools like Tor, Tails, and VPNs for safe data examination
5. High- Value PII Identification	Utilizing web scraping and manual analysis for data collection
6. Policy Enhancement Development	Analyzing collected data to identify gaps; proposing realistic enhancements
7.Comprehensive Evaluation	Comparing new recommendations with existing strategies

A. *Privacy/Security Standards and Data Collection*

This study will involve the collection of current organizational privacy and security policies, focusing on prevalent organizational privacy and security standards such as ISO 27799, ISO 27018, and ISO 27037. Data analysis aims to examine the distinct characteristics of these policies, exploring their impact on the protection of PII, and the potential for user data exploitation arising from weak policy implementation. Furthermore, an exhaustive examination will be conducted to identify companies with enhanced security and privacy policies that result in minimal breach events, and how these enhancements compare to companies with a high frequency of security breach incidents. This will include an assessment of the policies' comprehensiveness, adaptability to emerging security threats, and the effectiveness of their implementation within various organizational contexts. This multifaceted approach ensures a comprehensive understanding of the policy design, implementation, and their associated real-world outcomes; essential for the development of more effective strategies in data protection and shaping robust approaches to safeguarding sensitive data and information.

B. *Dark Web and PII Information*

Examination of the use and transfer of PII within the Dark Web will stem from second stage of this research investigation. The examination of PII on the Dark Web will provide insight into the consequences beyond damaged organizational reputation. Internet hotspots and Virtual Private Network (VPN) connections will be utilized to ensure a secure and secluded platform is used, disconnected from personal information disclosure. Analysis of the PII on the Dark Web will be conducted utilizing various security tools such as, but not limited to, the Tor web browsers and Tails. The data collected will assist in better understanding the value and demand of stolen PII on the Dark Web, and the impacts of privacy and security breach events on PII due to poor organizational policies.

C. *PII Data Collection Methods*

The initial collection of data will focus on identifying the types of (PII) that hold the highest value on the Dark Web. This process involves systematically analyzing various marketplaces and forums, where PII is often sold or exchanged. Our approach will involve collecting data on the varieties of listed information, the frequency in their occurrence, and the prices at which they are presented. Our methodology will include collecting data on the types of information listed (e.g., Social Security Numbers (SNN), financial information, full identity packages, etc.), the frequency of their appearance, and the prices at which they are offered. Data filtering will consist of categorization of PII into groups, such as, financial information, personal, medical, and other unique identifiers that could be used for malicious intent. A breakdown of the categorical groups of PII can be seen within Table 2. To ensure comprehensive data collection, a web scraping tool will be employed during the research process.

This tool is designed to efficiently extract PII from a wide array of Dark Web sources. Utilizing this approach enables extensive coverage across various Dark Web marketplaces, ensuring a thorough capture of prevalent information. Moreover, a manual analysis of Dark Web Markets will be conducted to obtain valuable insight into PII trading practices that could have been missed by the automated tool. Our analysis aims to uncover not only the most commonly traded types of PII but also factors such as the level of completeness of the data sets.

TABLE II. RESEARCH PROCESS

Information	Examples
1. Financial Information	• Credit Card Details Bank Account Numbers
2. Personal Information	• Social Security Numbers Date of Birth • Home Address Phone Number
3. Medical Records	• Health History Prescriptions
4. Unique Identifiers	• Personal Identification Numbers • Biometric Data

D. *Data Analysis*

Once the data is collected, we will analyze existing organizational policies and how they are equipped to protect the types of PII identified as most valuable on the Dark Web. An in-depth analysis of these policies will then be conducted, examining their specific conditions to determine their effectiveness in safeguarding sensitive data. This analysis will focus on uncovering any deficiencies or weaknesses in the policies that may leave valuable PII vulnerable to exploitation by attackers. By comparing these findings, we aim to identify critical areas where current policies fall short. Based on this comprehensive evaluation, we will develop a set of refined policy recommendations, tailored to address the identified gaps, and enhance overall data protection. Our goal is to propose practical, applicable enhancements that can be realistically implemented across various organizational structures, thereby significantly strengthening the defenses against PII misuse and exposure. Data identified in this step will be crucial, as it not only focuses on existing data protection strategies but also contributes to effective data security measures in an ever-changing cyber threat landscape.

E. *Privacy Analysis*

After the implementation of the proposed set of enhanced user privacy practices and security controls, the proposed changes will be measured for effectiveness. The primary focus will be on monitoring and analyzing organizations that adopt these new measures. This assessment will involve a comprehensive review of security incident reports, breach frequencies, and the types of PII compromised post-implementation compared to prior records. Furthermore, we will conduct future studies on the Dark Web to observe any noticeable decrease in the availability and/or demand for the previously identified high-risk types of PII. This step is crucial for validating the effectiveness of our proposed policy enhancements. By quantitatively and qualitatively measuring the impact of these new measures, we aim to provide concrete evidence of their efficacy in mitigating PII exposure. The insights derived from this analysis will assist organizations in enhancing their data protection strategies against ever- evolving cyber threats landscape.

IV. RESULTS AND DISCUSSION

Below are details about the results identified during the initial investigation and discussion about the future proposed research work and intended investigations.

A. *Abbreviations and Acronyms*

After the completion of our study, we anticipate revealing critical insights into the current organizational privacy and security policies in protecting (PII) and the exposure of users' data found of the Dark Web through weak privacy policy. We expect our research will highlight discrepancies in policy effectiveness across various organizations, highlighting the need for policies that are, not only comprehensive, but also adaptable to ever-evolving cyber threats landscape. Our analysis is likely to show a contrast in breach frequencies between organizations with heightened security measures and those with weaker protocols. Additionally, our exploration of the Dark Web should offer a deeper understanding of the market dynamics for stolen PII, revealing how weaknesses in organizational policies can fuel the Dark Web market through the exchange of PII and other sensitive personal and identifiable information.

These anticipated results, while subject to the typical limitations of secondary data analysis and the ever-changing landscape of cyber threats, aim to contribute valuable insights into effective data protection strategies. The study's outcomes could serve as a foundational guide for future research, particularly in developing more resilient privacy and security frameworks. Through this research, we seek to provide meaningful recommendations for policy enhancements, thereby aiding organizations in fortifying their defenses against sophisticated cyber threats.

V. CONCLUSIONS

The main objective of this study will be to analyze the effectiveness of current organizational privacy and security policies in protecting Personally Identifiable Information (PII). While the actual results are pending, our research aims to highlight key areas where privacy and security policies can be improved to mitigate the risks associated with PII. The expected findings will likely emphasize the importance of dynamic and adaptable security policies in the rapidly evolving landscape of cyber threats. This study aims to shed light on the complex relationship between organizational policies and the actual security of personal identifiable information (PII).

VI. LIMITATIONS

This research recognizes several limitations in our study. Initially, the reliance on secondary data sources introduced the potential for bias and/or gaps in available information. Additionally, the dynamic nature of cyber threats institutes that our findings are always evolving, necessitating ongoing research in the field. Lastly, the analysis of organizational policy and Dark Web markets, may not fully capture the breadth of factors influencing PII security. Future research could expand on these areas, exploring the efficacy of specific policy enhancements and the long-term impact of cybersecurity

A. *Research Implications*

The implications of our mixed-methods methodology research will serve as a basis for informing future strategies in cybersecurity and data protection. By identifying high-risk PII categories and evaluating policy effectiveness, our work will guide organizations in prioritizing their security initiatives. In addition, this research also serves as a scientific reference resource for academics and researchers interested in better understanding how organizational decisions within the technology domain can have an impactful effect on the Dark Web market. The study's findings are expected to reinforce the necessity for a proactive approach to cybersecurity, emphasizing continuous policy updates and employee training. These insights could prove invaluable for organizations striving to safeguard their data assets in an increasingly digitized world.

VII. Future Work

The consequences of the negative actions have been identified, this research will introduce a novel set of enhanced user privacy practices and security controls aimed at mitigating the risk of PII exposure.

Proposed future work includes:

- Expanding upon the traditional Security, Education, Training, and Awareness (SETA) model
- Isolating the most effective strategies and security measures needed to protect and secure data
- Development of an enhanced education framework for increased security defenses
- Creation of a more secure cyber infrastructure for organizational security use

VIII. About the authors

Joseph Squillace, Ph.D., is an Assistant Teaching Professor of Cybersecurity at Penn State University. Joseph received his Ph.D. in Information Systems (DISS) with a concentration in Information Security from the College of Engineering and Computing at Nova Southeastern University (NSU). Joseph's research interests include Cyberbullying, Cybersecurity, Privacy, Cybersecurity in Artificial Intelligence (AI), Secure Supply Chain, Data Integrity, and Economics of Information Security and Privacy Breaches. Joseph has previously published scholarly research in Cybersecurity, Artificial Intelligence, Privacy, Information Systems, Computer Science, the Internet of Things (IoT), and Climate Change domains. Joseph is also engaged with International research grants, including NSF and the European Commission (EU), and multiple International Research Studies while currently collaborating with academic research teams in both the Czech Republic and Hungary.

Justice Cappella is a Senior Research Assistant attending Penn State University. Justice is graduating in May 2024 with a Bachelor of Science degree in Business Management and Marketing and a minor in Project Supply Chain Management. Justice's research interests include Business, (Secure) Supply Chain Management, Marketing, Cyberbullying, Cybercrime, and Climate Change, and she is currently collaborating on multiple research studies at Penn State Schuylkill. Justice is also engaged with multiple International Research Studies and is currently collaborating with academic research teams in the Czech Republic and Hungary.

Andrew Sepp is a junior researcher attending Penn State University. Andrew is Junior and will be graduating in May 2025 with a Bachelor of Science degree in Cybersecurity Analytics and Operations. Andrew's research interests include Cybersecurity, the DarkWeb, Ransomware, Malware, Cyberbullying, and Cybercrime. Andrew is currently involved in multiple research studies at Penn State Schuylkill.

References

Eason, G., Noble, B., & Sneddon, N.I. (1995). "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London, vol. A247*, pp. 529–551.

Jacobs, I.S. & Bean, C.P. (1963). "Fine particles, thin films and exchange anisotropy," *Magnetism, vol. III*, G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271–350.

Leave no trace on the computer. Tails. (n.d.-a). https://tails.net/about/index.en.html

Liu, Y., Lin, F. Y., Ahmad-Post, Z., Ebrahimi, M., Zhang, N., Hu, J. L., & Chen, H. (2020). "Identifying, collecting, and monitoring personally identifiable information: From the dark web to the surface web," *IEEE International Conference on Intelligence and Security Informatics (ISI)* (pp. 1-6). IEEE.

Maxwell, C.J. (1892). "A Treatise on Electricity and Magnetism," *3rd ed., Vol. 2*. Oxford: Clarendon, pp.68–73.

Moore, T. (2010). "The economics of cybersecurity: Principles and policy options," *International Journal of Critical Infrastructure Protection, 3*(3-4), 103-117.

The Tor Project, Inc. (n.d.). Tor. Retrieved from https://2019.www.torproject.org/about/overview

Thomas, L., Gondal, I., Oseni, T., & Firmin, S. S. (2022). "A framework for data privacy and security accountability in data breach communications," *Computers & Security, 116*, 102-157.

Yorozu, Y., Hirano, M., Oka, K., & Tagawa, Y. (1982). "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan, Vol. 2*, pp. 740–741, [Digests 9th Annual Conf. Magnetics Japan, p. 301].

Young, M. (1989). "The Technical Writer's Handbook," *Mill Valley, CA:* University Science.