ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

SSN and PII Information Inside Logs

Aakash Aluwala

USA

ABSTRACT

The paper deals with the threat of SSN and PII disclosure within log systems, especially those monitoring tools employed to improve customer experience. This paper studies the problems of SSN and PII disclosure, including but not limited to the chance of identity theft, financial fraud, and negative reputation to institutions. It offers a holistic plan covering data analysis, coding changes, continuous monitoring, and keeping a good relationship with stakeholders. Organizations will reinforce their data security posture through robust solutions and proactive measures, follow all needed regulations, and develop trust with their customers and stakeholders in a rapidly growing, data-driven environment.

***Corresponding author**

Aakash Aluwala, USA.

Received: June 11, 2024; **Accepted:** June 17, 2024; **Published:** June 25, 2024

Keywords: SSN and PII, Monitoring Tools, Data Security, Data masking, Privacy Protection, Compliance

Introduction

In the digital era, when unlimited volumes of data are generated and kept, the issue of confidentiality and the safety of private information is of the utmost importance [1]. One of the major security challenges organizations face is that Social Security Number (SNS) and personally identifiable information (PII) are regularly exposed in log files, which are usually unnoticed in traditional security measures. This article highlights the urgency of the SSN and PII information exposure by the logs, taking into account the applications utilized by the organizations to enhance customer experience, for instance, Tealeaf, Glassbox, and Splunk.

With the majority of logs revealed to contain information containing SSN and PII, companies handling large-size workforces comprising a mix of full-time employees and contractors, including global offshore contractors, run a great risk of leaving themselves vulnerable [2]. Access to over 10 thousand persons leads to a complicated and realistic process of manually identifying fields within logs that demonstrate PII-sensitive data. The old techniques of log monitoring could be more efficient in responding to this challenge. Thus, robust monitoring tools and data analytics solutions are crucial.

Literature Review

Research on the SSN and the person identification number (PII) extracted from logs indicates the broad base of this problem and the devastating losses that are a side effect of this problem with individuals and organizations [3]. Data breaches may be the first item on the concern list. Still, the erroneous sharing of secure information like SSNs and PII can lead to more serious consequences such as identity theft, financial fraud and reputational damage.

One of the organization's main issues is that many log data come from different sources. Codifying personal details using a manual method is unacceptable, and it is easy to make several mistakes [4]. Traditional monitoring systems usually cannot keep up with the increasing complexity and scale of modern IT infrastructure, leaving organizations open to undetected data leaks and possible cyber-attacks [2].

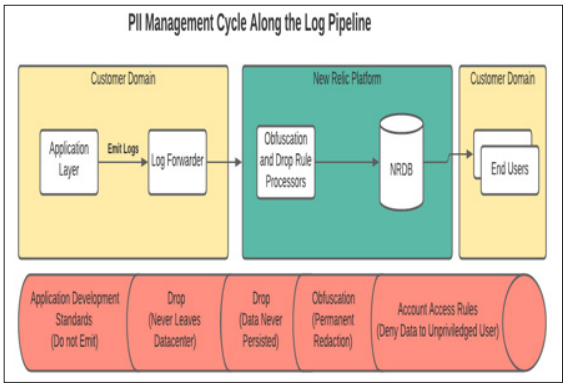

Figure 1: PII Management Cycle

In contrast to these obstacles, analysts recommend implementing state-of-the-art alert systems with AI-based analytics. These tools utilize machine learning algorithms for self-learning purposes that can shortly enable the detection of log files with signs of PII and SSN data exposure [5]. The introduction of smart analytics for dealing with big data in real time allows companies to detect and neutralize incipient security risks, thus helping them prevent their critical data from being breached [1].

Moreover, key data security practices such as data masking, encryption, and access controls are the fundamental building blocks of a complete data security plan that keeps the SSN and PII data safe in the logs [6]. Data masking techniques include

obfuscating the data by encrypting and rendering it unreadable but allowing those with authorization to view the data to read it as they normally would. With encryption and access controls, however, organizations can control who has access to the data based on pre-defined permissions, thus protecting data in motion and at rest [7].

Some other works have also looked at the application of technologies like differential privacy that can be used to un-identify private information from logs. Differential privacy is a method of offering privacy to statistical databases by introducing noise to the responses of queries [8]. It has been applied in various domains, from publishing census data to machine learning to balancing privacy and utility. Some recent work has also proposed differentially private algorithms for log and event streaming data that can be useful in reducing the risk of PII leakage through auditing and analysis [9].

One way to maintain privacy and confidentiality is data anonymization. K-anonymity and l-diversity are formal anonymization techniques that attempt to make attribute linkage easier and provide reidentification protection if external information is merged with the anonymized data [10]. Researchers have investigated how these concepts could extend to sequential data typically found in logs and suggested methods like generalization and suppression to achieve appropriate anonymity while preserving log utility [11]. Continued research on privacy-enhancing technologies may yield solutions that help mitigate PII risks in system logs.

Monitoring Tools Impacted

The impact of PII and SSN leakage inculcates a number of tools that augment customer experience and manage system infrastructure [4]. Tools used in different industries, such as Tealeaf, Glassbox, and Splunk, may also expose the risk of unauthorized access to confidential data.

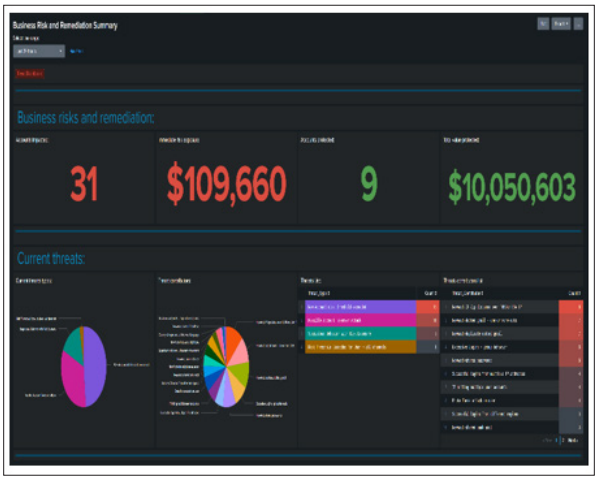

Figure 2: Splunk App for Fraud Analytics

While Tealeaf and Glassbox are well-known for their strong feature in customer experience management and session replay, they inadvertently also capture and store user interactions and sensitive data like SSNs and PII. Although these tools make it easy to collect user behaviour data and website performance stats, the fact that any user session is logged indiscriminately means users' privacy is at risk, and there is a higher chance that information is exposed.

In the same way, Splunk, the leading system for collecting, indexing, and analyzing machine-generated data, such as logs, also suffers from challenges posed by the security of confidential data within its large repositories [12]. Due to Splunk's prevalence in security monitoring, compliance, and IT operations, irresponsible data protection endangers regulatory bodies, which are consequently established by reputational damage.

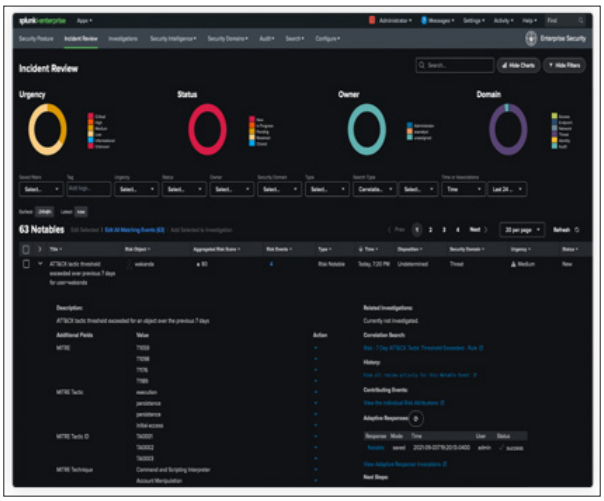

Figure 3: Splunk Products

A wide array of monitoring tools and challenges that arise from managing access for thousands of people, including employees, contractors, and offshore teams, have resulted in a higher possibility of SSN and PII exposure in the logs [3]. Furthermore, the amount of data handled by such tools makes manual identification of exposed information tremendously hard, demanding automated systems and systems measures to mitigate the risks adequately [13].

These customer experience and monitoring tools collect huge amounts of data from various sources daily. For example, a medium-sized company using these tools could generate terabytes of daily log data [6]. With the scale and complexity of modern IT environments, it becomes incredibly difficult for security teams to manually inspect this volume of data and identify any PII that may have been inadvertently captured. This difficulty is compounded when access must be provided to the various employee groups mentioned previously. The risks of improper access or exposure are greatly increased without adequate controls and protections.

Tools like Splunk that aggregate logs from many different applications and systems potentially create a single point of failure [3]. If sensitive data can find its way into these repositories, it represents an attraction for attackers or malicious insiders. Therefore, The tools must implement strong access governance and data security best practices to prevent such compromises. Failure to do so places an undue burden on the organizations that rely on these platforms for monitoring and compliance tasks. Careful oversight of data handling policies is warranted across all stages, from collection to storage and analysis.

Tasks

The operations related to SSN and PII exposure within logs consist of several strategic activities committed to reducing risks, improving data security features, and staying compliant with regulations associated with monitoring instruments [14].

The first step is to determine possible exposure to the monitored tools through agents such as Tealeaf, Glassbox, and Splunk. It involves a thorough investigation of logging mechanisms, data storage methods, and access restrictions to determine the areas where the SSN or PII data may be exposed to the risk of unauthorized access or leakage.

When exposure points are located, solid measures that can prevent them from happening must be enforced. The implementation set of the measures is based on blocking PII-related data at the initial stage, which means preventing sensitive data from entering the monitoring tools' repositories in the first place.

The need to cover PII information must be balanced. It is necessary for the sake of individuals' privacy and also fulfils data-protecting regulations. Even though log data masking mechanisms don't afford sensitive data its proper privacy, they still allow them to contribute to legal operations.

Creating filters that restrict the PII data inside the log is essential for avoiding abuses of access and reducing the risk of data compromise. By using filters and access restrictions, organizations can limit the transfer of sensitive data and thus avoid unintended data exposure.

Solution and Implementation

Solving the SSN and PII exposure in log control requires a systematic method that involves data analytics, communication, cooperation, and frequent supervision. Initially, data analysis and testing must be conducted for the organization's purpose; the lower environments will be loaded with test data to find the potential SSN and PII data logs. Establishing effective relations between application owners and development teams is critical to prevent PII data from entering applications and ensure that the data complies with data protection laws and regulations. Subsequently, codes must be made available in all the surrounding environments, including development, QA, UAT and final production stages, and stringent testing sets have to be developed to substantiate the code's efficacy in stopping PII disclosure within the logs.

Enabling AI dimensions and events across customer experience tools will allow organizations to create intelligent reports that aim to strategically target and identify patterns of PII information, allowing for a proactive response to potential security threats. Stable logging monitoring and logging validation are key to avoiding PII information that can be seen across monitoring tools. Alerts should be put in place that will notify relevant players the second PII exposure is detected.

Figure 4: PII: Strategies for Securing Data [15]

Subsequently, codes must be made available in all the surrounding environments, including development, QA, UAT and final production stages, and along with that s, stringent testing sets have to be developed to substantiate the code's efficacy in stopping PII disclosure within the logs. Automated scanning and validation tools will evaluate code diffs and packages before deployment, checking for compliance with privacy and security standards. Static application security testing (SAST), interactive application security testing (IAST) and runtime application self-protection (RASP) mechanisms will continuously monitor for insecure logging.

Enabling AI dimensions and events across customer experience tools will allow organizations to reorganize reports that aim to strategically target and identify patterns of PII information, allowing for a proactive response to potential security threats. Dimensional attributes like data types, value lengths and formats will be defined through workshops with data owners. A metadata catalogue and data dictionary detailing sensitive field definitions will be maintained. Machine learning clustering and anomaly detection algorithms will then profile typical and atypical log contents to pinpoint records likely containing PII.

Stable logging monitoring and logging validation are key to avoiding PII information that can be seen across monitoring tools. Centralized logging infrastructuresCentralizedn storage, processing and visualization components will be buvisualizationLog-matching schemas using regular expressions and static dictionaries will be created to scan logs in real-time. Anomalous logs flagged by the AI reports will be fed back to the monitoring consoles for human verification and escalation. Syslog-NG, Fluentd, and Logstash forwarding and parsing agents will be configured across applications and platforms.

Alerts should be put in place to immediately notify relevant stakeholders like privacy officers, audit teams, application managers, and security ops whenever PII exposure is detected. Detection rules chaining together multiple dimensions and anomaly conditions will reduce false positives. Alerts will provide drill-down links to the underlying log records and associated transactions for further examination. Notifications will establish clear guidelines on initial containment and mitigation response until root causes are identified and patched.

Results

Smart implementation of strong tools for protecting SSN and PII in log files opens great opportunities for organizations, increasing data risks, and corresponding with regulatory requirements. Enterprise-wide security is the key here, where timely access to sensitive information is prevented from getting into monitoring instruments' repositories, which, in turn, minimizes the risk of data breaches and minimizes consequences like financial loss and reputation damage. Continuous authentication and validation activities as a part of data security also prevent PII information from being visible in any monitoring tools, thus prompting remediation actions upon detection. Overall, these initiatives foster a culture of security awareness and stakeholder engagement, leading to support for proactive measures, all aiming to strengthen the organization's data security posture.

Conclusion

One organization that should strictly guard confidential information and maintain regulatory compliance is recording and monitoring SSN and PII exposure within logs. Through a

robust approach, which includes data analysis, code changes, and constant monitoring, organizations may successfully guard against risks and keep their data secure, ready to face new and current challenges. Proactive measures, multi-stakeholder cooperation, and iterative improvements must be taken as the major components of a comprehensive strategy for resolving this issue in a data-driven world. As organizations implement these steps, they show that they are considering protecting the confidentiality of their customers and people interested in their affairs.

References

1. David Kolevski, Katina Michael, Roba Abbas, Mark Freeman (2021) Cloud data breach disclosures: the consumer and their personally identifiable information (PII)? 2021. IEEE Conference on norbert wiener in the 21st century (21CW) https://ieeexplore.ieee.org/abstract/document/9532579.
2. Aryee Daniel (2020) Cybersecurity Threats to the Hotel Industry and Mitigation Strategies. Diss Utica College https://www.proquest.com/openview/64221b72cf1d4d9a39a0a17cbe763b25/1?pq-origsite=gscholar&cbl=18750&diss=y.
3. Rana Rima, Razieh Nokhbeh Zaeem, Suzanne Barber K (2019) An assessment of blockchain identity solutions: Minimizing risk and liability of authentication. IEEE/WIC/Minimizingational Conference on Web Intelligence https://dl.acm.org/doi/abs/10.1145/3350546.3352497.
4. Shankar Nithya, Zareef Mohammed (2020) Surviving data breaches: A multiple case study analysis. Journal of Comparative International Management 23: 35-54.
5. Miller Jr Ronald Ernest (2023) Measuring the Effects of Trust Between E-Tailers and Customers when Personally Identifiable Information is Compromised in a Data Breach: A Quantitative Study. Diss Trident University International https://www.proquest.com/openview/097f5a71b6d12aa59e7f5a1736025f2d/1?pq-origsite=gscholar&cbl=18750&diss=y.
6. Batista Edgar, Agusti Solanas (2021) A uniformization-based approach to preserve individuals' privuniformization-basedining analyses. Peer-to-Peer Networking and Applications 14: 1500-1519.
7. Stallings William (2019) Information privacy engineering and privacy by design: Understanding privacy threats, technology, and regulations based on standards and best practices. Addison-Wesley Professional https://www.oreilly.com/library/view/information-privacy-engineering/9780135278383/.
8. Bogatov Dmytro, Georgios Kellaris, George Kollios, Kobbi Nissim, Adam O'Neill (2021) Epsolute: Efficiently querying databases while providing differential privacy. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security 2262-2276.
9. Ávila Ricardo, Raphaël Khoury, Richard Khoury, Fábio Petrillo (2021) Use of security logs for data leak detection: a systematic literature review. Security and communication networks 2021: 1-29.
10. Li Shaobo, Matthew J Schneider, Yan Yu, Sachin Gupta (2023) Reidentification risk in panel data: Protecting for k-anonymity. Information Systems Research 34: 1066-1088.
11. Rafiei Majid, Wil MP van der Aalst (2021) Group-based privacy preservation techniques for process mining. Data & Knowledge Engineering 134: 101908.
12. Hristov Marian, Maria Nenova, Georgi Iliev, Dimiter Avresky (2021) Integration of Splunk Enterprise SIEM for DDoS Attack Detection in IoT. 2021 IEEE 20th International Symposium on Network Computing and Applications (NCA) https://ieeexplore.ieee.org/abstract/document/9685977/.
13. Stoyanova Maria, Yannis Nikoloudakis, Spyridon Panagiotakis, Evangelos Pallis, Evangelos K Markakis (2020) A survey on the Internet of Things (IoT) forensics: challenges, approaches, and open issues. IEEE Communications Surveys & Tutorials 22: 1191-1221.
14. Saleem Hamza, Muhammad Naveed (2020) Sok: Anatomy of data breaches. Proceedings on Privacy Enhancing Technologies https://petsymposium.org/popets/2020/popets-2020-0067.php.
15. (2024) Empowering PII Management: Strategies for Security & Data. Data Dynamics, Inc. - Intelligent Data Management. Transforming Organizations into Better Data Custodians. https://Organizationsicsinc.com/blog-balancing-control-and-democratization-six-pii-management-strategies-for-data-empowerment/.

Copyright: ©2024 Aakash Aluwala. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.