Data user interaction monitoring considerations for information security measures

Anand Athavale
Independent Researcher, Decades of Industry Experience in Data Management, USA

* Corresponding Author: **Anand Athavale**

Article Info

ISSN (online): 2582-7138
Volume: 04
Issue: 02
March-April 2023
Received: 03-03-2023
Accepted: 06-04-2023
Page No: 624-627

Abstract
Data leaks are not new to the information security industry. Awareness of malicious insiders trying to steal data has been always there. Much before ransomware, famous cases of insiders stealing the information, including from the government agencies, and then, certain outlets channelizing the leaked information have happened. Along with those, data leak prevention solutions have also been around. But ransomware driven data exfiltration and double extortions have forced organizations to rethink their approach. Besides this, organizations have been forced to validate the nature and granularity of the data loss, or theft, by regulation changes and to cross check the scope of the data loss to verify the accuracy of threat from an attack. However, the current approach to user monitoring is limited in scope while also missing the element of accountability. User data interaction monitoring considerations reach far beyond behavior analysis and response.

DOI: https://doi.org/10.54660/.IJMRGE.2023.4.2.624-627

Keywords: User data interaction monitoring, User Behavior analysis, User Identity mapping, Machine and Human activity

Introduction

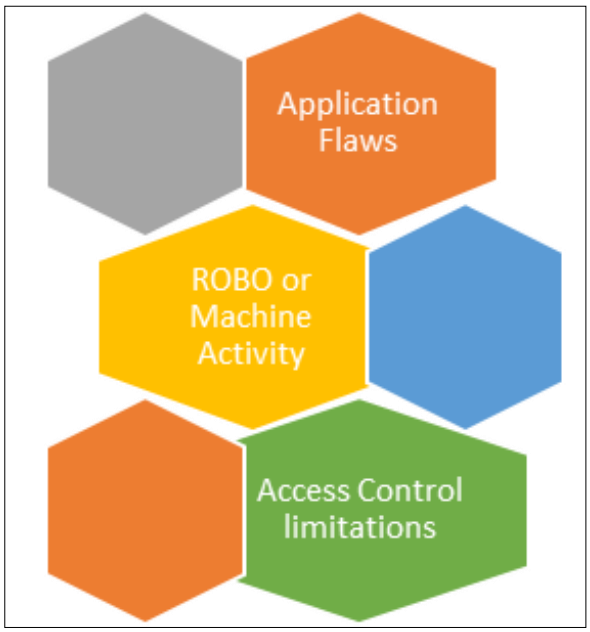
Data leak prevention basic job is to prevent classified data from leaving the boundaries of an organization. Some solutions divide these into data at rest, endpoint, and network data leak prevention. End points in data leak prevention context typically are laptops with devices like USB drives which can be used to steal data. Network typically refers to cross-over from company networks into public network. There is element of Data Leak Prevention, or DLP, in Cloud Access Security Broker (CASB), which is like a firewall which allows the organization to extend their security control beyond their network boundaries. A DLP function of CASB can detect and block the unauthorized transfer of sensitive data, including identifying patterns like keywords, social security numbers, or credit card details [1]. However, whichever forms the DLP function takes, the exfiltration ransomware have seen to consistently bypass it. The cloud era combined with the remote work has broken the protection wide open [2].

Understanding the bypassing methods for Data Leak Prevention

There is a story which can probably explain the issue at heart of the DLP solutions. There was a bi-cycle making factory. One of the employees used to often work late and on Fridays, he would carry a heap of grass from the factory premise on the bi-cycle. The guard at the gate always used to check his heap of grass as he suspected that the employee was stealing something from the factory. But he was not sure what it was. After a few years, the employee retired. After many years, the guard retired too. One day, the guard saw that employee and begged to satisfy the unanswered questions. Was he stealing something, and if yes, what was it? The employee answered that he would work extra and make another bi-cycle. On Fridays, he would leave his own bi-cycle in the factory and carry the newly built bi-cycle with the heap of grass on the back. He said that the guard was too focused on the grass and hence never paid attention to the bi-cycle.

Today, the ransomware is using many techniques to evade detection of data theft successfully. These methods range from network level methods to scripting and file sharing mechanism [3]. Whether it is using encryption, or, breaking the sensitive information into pieces to then evade detection, it is being successfully done. Hence, there is a clear need to monitor the activity users are carrying out on sensitive data where it resides. But it is easier said than done.

Challenges of preventing the user activity on sensitive data



Challenges to blocking user activity

- **Application flaws**
  Apart from the methods mentioned that are being used for data theft, there is one more method which exploits the flaws in application programming. It is popularly known as "SQL injection" which is effective in stealing from structured data sources. In short, the attackers change the scope of the data to be retrieved by a query mechanism, or the code itself tricking it into returning large set of data records instead of limited ones and granting wider privileges [4]. Here, the activity is carried with legitimate means and hence the user cannot be simply blocked from issuing queries. Of course, there are best practices to be followed during coding to prevent this, but the sad fact is that just one mistake is enough to open doors for hackers.
- **ROBO or machine user necessity**
  Usually, any human user retrieves small amount of data at a time, as humans have limited span and capacity to process information. There may be already machine activity going on in an environment, such as, some operations being carried out in thousands, on thousands of data items. There may be a legitimate need like reporting, or end of day batch processing, which necessitates using a user which is meant to carry out data activity in bulk. Thus, such activity cannot be blocked totally and it is hard to separate the malicious machine activity from legitimate machine activity.
- **Access Controls and malicious users**
  It is possible to block access to sensitive data by using access controls. But there are always malicious users with necessary permissions already granted who can make the access controls on sensitive data ineffective. Also, there is always chance of legitimate credentials being compromised.

Challenges of monitoring the user activity on sensitive data

- **Regulations preventing user activity monitoring**

There are restrictions in countries like Japan, blocking blanket monitoring electronic activity of employees [5]. Such restrictions along with employee-employer trust considerations make it difficult for IT to deploy such surveillance barring those involved in trading and sales. That however severely limits usability and effectiveness of user activity monitoring.

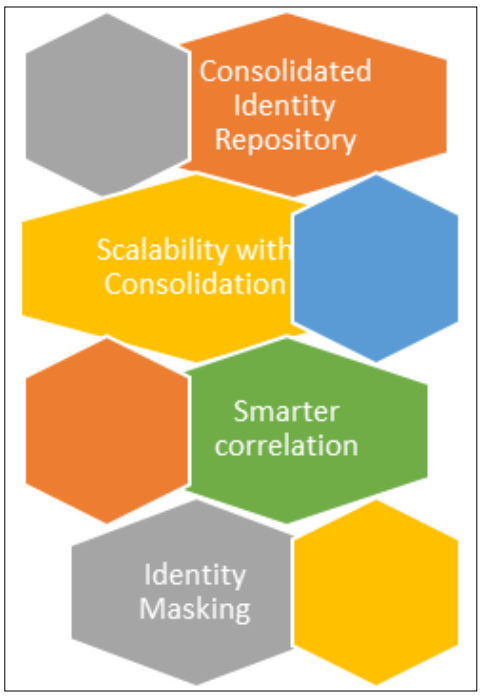
Regulations preventing blanket user monitoring

- **Volume of activity and performance impact of monitoring**
  The first of the second challenge is the volume of activity. Overall, in a large organization, there may be millions of individual data activity on hundreds of thousands of data items per hour. Keeping pace for capturing, processing, and analyzing such volume of activity is scalability challenge. Even if that challenge is addressed, storing historical activity, and correlating that itself becomes additional challenge. This is important because, threat actors lurk in environments for months before acting. This makes longer age of the historical activity relevant and necessary. The second part is the impact of monitoring itself. Typically, data sources, which include unstructured sources such as NAS devices and applications like SharePoint, do not necessarily have audit mechanisms turned on by default. This is mostly because if turned on, auditing takes a toll on end user data access response times. Impact levels vary from activity type to activity type, with something like write activity having less impact than an activity like delete. Hence, even if monitoring solutions exist, IT and storage administrators may not be inclined to turn on monitoring settings with data source applications.
- **Lack of linkage between identity systems across sources**
  Identity systems which are dedicated to only that purpose, providing identity management, are not a lot. However, data systems do not always necessarily use these, or, do not use these exclusively. That creates challenge for mapping one identity to other even when the identity, or, the account, or, the user, or the email belongs to the same human person. Examples of these are Active Directory, LDAP, Various data sources like

Google Drive, Box which are mostly based on email addresses, cloud accounts which are primarily known as security principles and the list goes on. Unless, there is some unification at some level to map all of these to a single entity, or, person, it becomes challenging to confidently establish accountability of activity on sensitive data.

- **Disconnect of identity and activity causing loss of accountability**
  More recent developments in cloud applications have given birth to function, or, code as a service, which does not have a "permanent body" in terms of virtual machine, or cloud instance, or even a container. These are functions or blocks of code which are executed by other services. As an example, AWS Lambda allows you to add custom logic to AWS resources such as Amazon S3 buckets and Amazon DynamoDB tables, so you can easily apply compute to data as it enters or moves through the cloud. Now, when there is automated action taken on any data in S3 bucket, who is the actor? The user which logged in to some application, which issued an API to read S3 bucket, which in turn called the code snippet as it was configured to do? But then what about the human user, or, service principle which created the AWS lambda function, or modified it? The more you analyze these latest advancements, the more the "who" of it seems to be less and less obvious.

Changes to achieve accountability and effectiveness in user monitoring



Recommendations for improving user monitoring effectiveness

- **Consolidated repository of identities**
  A unified list needs to exist, either in the data management or security application, or in the identity management system itself which clearly merges all the form factors of identities assigned to a human user. This means mapping and arriving to a single name, accounting for the rare collision scenarios, by consolidating active directory user account, email address-based identities and other systems like user ids in Unix systems. A more complicated scenario also needs to be handled where a single security principle may be accessible for use to multiple human users. Even then, there needs to be a separate consolidated accountable identity in the unified list, which represents such scenario. As an example, if John Doe and Jane Flame are using a common user account gameadmin@workers.com, the consolidated user repository needs to have separate entries for JohnDoeViagameadmin@workers.com and JaneFlameViagameadmin@workers.com and any activity of gameadmin@workers.com should be shown for both these logical user entities.
- **Scalability combined with smarter consolidation**
  For monitoring, application should be scalable to handle influx of millions of user activity records on data. There are technologies today which can handle the routing and indexing at such scales. Beyond the scalability, there needs to be some optimization which does not lose the value and distinction of the user interaction with the data. If the auditing mechanism is extremely granular to capture each I/O to a data item, or, the underlying file system, it does not mean that each of that interaction needs to be recorded. It could very well be consolidated on a time scale that does not lose usability. For example, it may not be damaging to consolidate these at each interaction type, or, time boundaries of say minute, instead of seconds. However, consolidation happening at the level of a day may lose its usability for forensics and similar needs for information security.
- **Smarter correlation**
  For identifying unusual behavior, typically the standard method is to compare it established pattern over a long duration of time like months. However, this increases the cost of storing historical user activity interaction. However, if there is insufficient history, then accuracy of separating unusual behavior from the typical behavior is lowered. Hence, there needs to be more effort and innovation put into identifying unusual behavior with less historical data. Use of artificial intelligence and machine learning could be one way, but is not the only one. Additionally, some applications may not necessarily generate file-system level operation event as per the effective user interaction. An application may create a temporary copy of a file, make the changes on the temp file and then overwrite the temp file on the actual file being modified by the user. The monitoring correlation needs to translate these file system interactions to the effective access event instead of just showing bunch of temporary file interactions and probably delete and create event of original file, which would be meaningless, for security operations person trying to understand the actual interaction to the original file.
- **Identity masking**
  Regulations and restrictions around blanket monitoring of employee or user activity are always going to act as hurdle to effectively catch and reduce data exfiltration. Hence, for the most parts, the consolidated list of identities in such application, or solution should be masked by default. Only when there is a real need to act, a consensus mechanism requiring more than one person

should translate the masked identity back to the real one. For day-to-day monitoring, identifiers of users appearing in interfaces, and reports and graphics of monitoring solutions should be non-identifiable meaning impossible to map to a human user.

Conclusion

Data Leak Prevention used to be primary measure for blocking exfiltration of sensitive data. However, primary focus of these solutions was on the boundaries and not at source. Today, exfiltration ransomware has gotten sophisticated to bypass the exfiltration blocking mechanisms of data leak prevention solutions. As discussed in this paper, there is a need to shift the focus of monitoring from boundaries to source, or at the least complement boundary monitoring with "at source" monitoring. As discussed in this paper there are a few challenges in effective monitoring at scale. However, with the necessary considerations by the monitoring applications, it is possible to overcome those challenges. Along with those adjustments, centralization of disparate user identities has also become necessary to improve accuracy and effectiveness of user monitoring to enable faster and meaningful collaboration among IT teams and security teams. Finally, masking of user identities for day-to-day monitoring and approval-based unmasking only on positive confirmation of any breach attempt should solve for regulation and other apprehensions around user monitoring.

References

1. O'Connor C. What is a CASB and how does it integrate with DLP? [Internet]. November 2022. Available from: https://www.docontrol.io/blog/what-is-a-casb-and-how-does-it-integrate-with-dlp. [cited 2023 Mar].
2. Spadaccini S. Why DLP solutions struggle to protect today's cloud workspaces. [Internet]. June 2022. Available from: https://www.safeguardcyber.com/blog/security/dlp-issues-why-it-isnt-enough-in-todays-cloud. [cited 2023 Feb].
3. Jaju A. How some tech-savvy employees are bypassing data leakage prevention measures. [Internet]. June 2022. Available from: https://angle.ankura.com/post/102hppz/how-some-tech-savvy-employees-are-bypassing-data-leakage-prevention-measures. [cited 2023 Mar].
4. Brook C. What is SQL injection? Definition, how it works, prevention tips & more. [Internet]. December 2022. Available from: https://www.digitalguardian.com/blog/what-sql-injection-definition-how-it-works-prevention-tips-more. [cited 2023 Feb].
5. Yamashima T. Atsumi and Sakai, L&E Global. Employment law overview Japan 2021-2022. [Internet]. 2021-2022. Available from: https://www.aplawjapan.com/application/files/2816/5542/9654/LEGlobal-Employment-Law-Overview_Japan_2021-2022.pdf. [cited 2023 Feb].