International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Data Exposure Considerations for Data Users in Identifying Security Measures

Anand Athavale

Independent Researcher, Decades of Industry experience in Data Management
andyathavale@gmail.com

Abstract
Data exposure plays a vital role in identifying security measures for Information Security. If data is considered to live in a house, data exposure methods are the windows and doors which need to be blocked with only select persons having access to those. This article elaborates on data exposure considerations. Due to significant evolution in data management, considering data exposure only from direct data users' aspect is no longer enough. Hence, data exposure considerations need to be broken down into direct data users and indirect ones. In this article we will only discuss direct data user exposure. This typically also comes very close to data access control. As discussed in the data content consideration paper, direct data user exposure consideration also has dependency on data content consideration. Lastly, the identity management features also come into play for data exposure control considerations.

Keywords: Data Loss Prevention, Information Security, Ransomware resilience, Sensitive Data, Data Access Control, Access inheritance

Introduction

Data users are a known entity created in any user management system which integrates with the data storage system or the application interacting with the data. While there are non-user entities these days which interact with data, this article is limited to data exposure to direct users. In that context, data exposure means a prominently static literal list of users who are allowed do some data operations on the data.

At a very high level, data operations can be grouped into creation, modification, read or access, deletion, and access control operations. There are other data operations, commonly known as renames, but those do not impact the content or the exposure for most of the data storage systems, unless the rename is not just a rename, but is coupled with a movement of data.

When considering data exposure, the next level granularity of type of exposure is a necessary sub-element to consider. The reason for the next level exposure consideration ties very closely in what type of risk it may pose if not handled correctly. A simple example of this is as follows. To note, in the real banking system, there are many guardrails to catch and prevent this. This example is used as a generic example to make the association of next level exposure consideration to the type of risk. Consider a simple bank record in a system with say account number and balance. If the exposure is limited to only reading, it poses the risk of data theft. If the exposure also has modification and or creation access, then the risk becomes a business risk where someone may increase the balance, or create fake account balances and

thus may put the bank at a loss. If the exposure is of type deletion, then all the balance data could be destroyed and this will simply shutdown the bank as they can no longer offer those services.

Data Location and exposure relation

Exposure considerations change with the data location and the storage layers as follows.

Structured Data

Structured data for the most parts has low necessity of direct data user exposure. The exposure mostly is within the application that interacts with the data. However, exposure considerations are required to ensure that the next level exposure-based data operation type is considered. Structured data access typically has one or more application users. These application users must have only the required data operation type exposure. A simple example is, a reporting application user should not be able to create, modify, or delete anything.

Semi-Structured and Unstructured Data

Both these types have the highest necessity of direct data user exposure to enable the productivity for the direct data users. The data items such as files, folders, shares in case of a file system or cloud-based drives, or, libraries, lists etc. typical to applications like SharePoint must have direct data user exposure for them to use the data. Nevertheless, these data location types also need data operation level consideration from exposure aspects. Typically, the administration role data users would require creation, deletion etc. at the higher level like shares for file system, sites or site collection for something like SharePoint and accounts for something like OneDrive or Google Drive. The actual data users have no requirement for such level of access. Conversely, while administration role users need access at the higher level, wherever the technology allows, they should not have any data operations permissions at the lower level. Typically, these administrator type roles end up having full access by default, and hence it becomes necessary to change those. Many technologies refer to this as “inheritance of permissions”. In hierarchical technologies, these can be carried down from the first level folders all the way to the leaf or the last level folders and files. Many of these technologies permit breaking the inheritance. Modern technologies offer alternative methods to automating permission assignments based on the “jobs and attributes” of direct data users. Microsoft Dynamic Access Control is one such example [1]. Even with the automation, the data location type remains an important aspect to consider when identifying security measures for exposure control. The simple reason being while automation exists to control access, manual ways of allocating permissions still can be used, which in some cases may defeat the purpose of automation.

Data Content and Data Exposure relation

At a very high level, both the data privacy/compliance regulations and data security best practices promote the limit of exposure on a need-to-know basis. While the “Know” word may imply a read type of data operation, it is to be considered broadly. Sensitive data content should not only be protected from theft, but from destruction point of view, need to be guarded from modifications and deletions. Hence, the data operations level exposure consideration should consider the data content labels. Data Content with personal data labels would highly focus on read operation while internal operation data content needs to be looked at from modify and deletion exposure.

The more sensitive the information, usually, less is the target number of direct data user exposure. The analogy here is for a delicate fragile item. The more delicate and fragile the item, the less "crowd" you want handling it. In real world, this also gets more complicated due to malicious insider threats. Malicious insiders are often associated with malicious employees intending to directly harm the company through theft or sabotage [2]. There is a subtle aspect to such employees which is tied to exposure control. Employees may not have a malicious intent initially. But, consider this scenario. Due to financial pressures, geo-political situations, or, labor conditions, a company decides to shutdown specific facilities. The plans and impacts of these shutdowns are exposed to some or all the employees including the ones who are part of the "to be impacted" facilities. This could be a trigger for turning otherwise harmless employees to turn malicious. This is another reason that data content must factor in identifying exposure control security measures.

Identify Management and Data Exposure relation

Type of identities

There are quite a few identity management systems. The original ones which are still popular are Active Directory, LDAP etc. Lately more as-a-Service IDP such as Azure Active Directory have also become available. Typically, there is affinity of security style and identity management system used for permission assignments to direct data users. Windows or CIFS typically go with Active Directory type of users or local Windows server level users. Unix/NFS go with the LDAP or local operating system users.

In Identity management systems, there are typically three types of users which are sometimes also referred to as accounts.

Certain users are present by default. Some of these default users have powerful permissions than others. Typical example for Windows or AD is the Administrator account. Typical Unix operating systems like Linux or AIX have a user called root. These accounts or users are sometimes also referred to as superusers [3]. Some of the built-in users do not have permission to log in. Some of these can be enabled to be able to login but by default it is disabled. These are the users where possible, data operations such as modify and read should be prevented as such as possible. Even some applications can have default "superusers." SharePoint as an example has a default account called admin. Sometimes there are ways to restrict the abilities of users like root such as using SELinux (Security-Enhanced Linux) in case of Linux Operating System [4]. There are also necessary steps like changing the default password for such accounts. But those steps fall more under infrastructure security than data security. Irrespective of the other measures being taken, data exposure to such users should always be a factor in the security measure identification process.

Second type of users are application centric. Similar to some of the default users, these users do not possess an ability to explicitly be used for login. Mostly these are users that can be used programmatically or in other words, only through specific applications or executables. Even then, data exposure control should still cover these types of direct data users, to limit the type of operations allowed only on the data that the specific application interacts with.

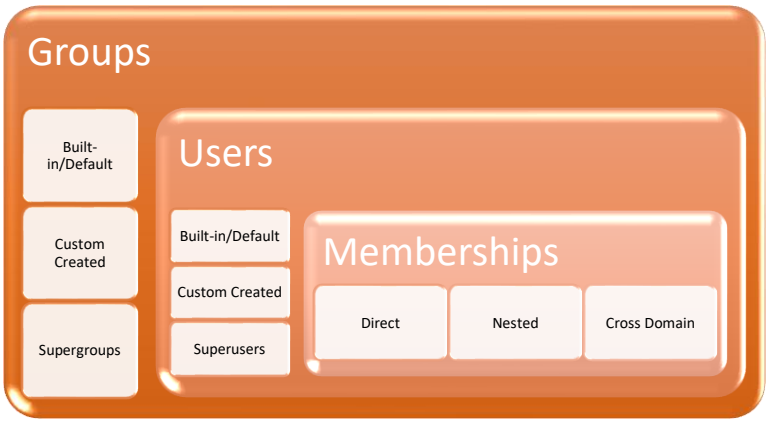
Third type of users are the users which get created for specific persons to give them access to several systems other than just the data systems which we are talking about. These users have a direct association with a specific person. For example, a user johndoe in an Active Directory domain pinkrose would be created for a person named John Doe. These users are typically very high in numbers compared to the other types and deserve the most attention from data exposure aspects. Here the many things associated with these types of users, such as job profile, grade, department need to be considered for data exposure

decisions. While Identity management systems are the easiest way to allocate direct user permissions, sometimes, other systems which are not strictly identity management type, also get used. An obvious example is an email address. An email address while not having a strict one to one mapping is also used for allocating permissions. We have seen this when we share content from OneDrive or Google Drive using email addresses for sending links to the content. The email address used may be part of the Identity Management system, but it is not strictly necessary, or always the case. Email addresses from free or paid email applications can also be used for exposure control. This adds a certain level of complexity for exposure control. The good part is, these complexities do not apply to structured applications like databases, as those do not give a wide flexibility of data sharing.

Groups and roles

Groups and roles concepts were introduced to simplify access control. However, often we see that a measure introduced to simplify a critical process or a security measure can also be the cause of risk. We have all seen the fire alarm trippers installed in common places. While they are there to alert about fire in an easy way, sometimes they get misused for malicious intent.

Before groups and roles were a thing, access control assignment needed to be repeated for every single user. Also, access control assignment needed to be changed for each user, if the user's work profile changed. To ease this burden, groups and roles came into existence. Instead of assigning access to each individual user one by one, groups, which are logical collections of users, reduce a thousand operations in just one. Group concepts exist in most of the identity management systems, and within almost all applications. As Groups are logical collections of users, roles are logical collections of permissible operations. Together these simplify the access control process.



Data User Elements related to Exposure

Just like default users, there are default groups, and they work in different ways when it comes to data exposure control. As an example, a group called Everyone, is a logical representation of all user entities which can "reach" the specific data item. There is no specifically assigned membership for each user to everyone group. If any data item, say /Employeeresources/Employeehandbook.pdf, has a read permission to the group "Everyone", it means any user which can reach /Employeeresources/Employeehandbook.pdf can read it. This makes sense because it will be hard to keep track of which new employee joined an organization and which one left and to keep updating the permissions for such data items. But care must be taken to not use such groups for access control to more restrictive data items like source code, or trade

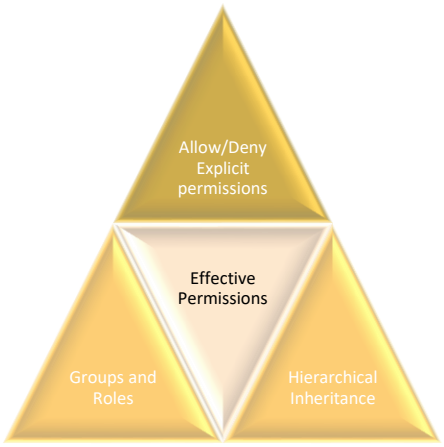
secret documents. Some of these groups themselves can be of type "supergroups" with more administrative powers. In Windows, examples such groups are Administrators or Domain Admins. This is why exposure control to direct users is a critical aspect in identifying security measures for data and it ties with data content on many levels.

It is important to also remember that groups can exist on individual systems and at identity management system level and many times users from one can be added to groups from the other. Sometimes this is referred to as cross domain membership. As an example, Windows Servers have local groups. Local users from the specific Windows Servers domain and users from active directory domain both, can be members of such groups. Focusing on one or the other type of users for controlling memberships can create risk of unintended exposures.

This can get further complicated. There is also the concept of nested groups, where group A11 and A12 can be part of a group called A1, which could in turn be part of a group called Group A. Because of this, any permission assigned to group A is also "inherited" by the member user of group A12. There could be another group LG1 which is a local group on a Windows Server. These are some of the aspects bad actors may take advantage of for carrying out their intended attack operation, whether encryption, destruction, theft, or all of them [5].

There is another consideration when looking at exposure control considerations for security measures. This is often overlooked as this is less visible compared to the traditional access control which is direct. Often this overlooked aspect is referred to as "effective permissions." Effective permissions are net result of considering inheritance, membership and allocated direct permissions often known as "explicit permissions." There are also concepts of explicit permissions for allow and deny making effective permission identification complex.

Effective Permissions Calculation



Effective permissions primarily are impacted by two related concepts. First are the roles, or grouped permissions. Since it is always possible to assign more than one role to any user, while the intent may be to segregate duties of read-type users from the admin-type users by creating separate roles, both roles may still end up assigned to the same user. A classic example of this is a job change. If a previous developer shifts job to become administrator, then the carry forward role may give that user both types of permissions, the read-write on code repository, and the create-destroy on the storage for the code repository. Second method of inadvertently getting both types of permissions, or unintended permissions

is through different groups. As discussed earlier, the same user can be part of multiple groups. So, even if separate groups are created to segregate job duties, the same user can end up being made part of those groups, effectively getting unintended, or, unnecessary permissions. Finally, in certain technologies like Windows, the creators or owners have some default permissions [6]. These are some of the underlying reasons which generate the need for monitoring the access control at regular intervals, manually or using automation. Monitoring of access control is by itself a vast topic which is separate that static exposure control considerations for direct data users.

Conclusion

Data content is the source of information security, which gives birth to data exposure control as one of the measures to minimize the security risks related to data. The direct data user-based exposure considerations revolve around the purpose, content labels, data location types and the intended user profiles as we discussed during this article. Methods, tools and technologies made to ease the burden of IT operations for access control also introduce the risk to the very objective those are created for. IT practitioners need to remember the quote, with great power comes great responsibility, when utilizing access control simplification mechanisms. To reiterate, the identity management itself has moved past the direct data user concepts due to various cloud technologies and constructs which need separate consideration. Use of email address to identify the data user is the first example of such scenarios, but it has gone passed email addresses in cloud technologies with function-as-a-Service like AWS Lambda and service principles concepts.

References

1. Andy Morales, Configuring Dynamic Access Control in a Lab, (March 2019), https://www.amorales.org/2019/03/basic-dynamic-access-control.html , (May, 2020)
2. Holger Schulze, 2019 INSIDER THREAT REPORT [Cybersecurity INSIDERS with Fortinet] (November 2019), https://www.fortinet.com/content/dam/fortinet/assets/threat-reports/insider-threat-report.pdf, (June, 2020)
3. Jess Phillips, Scott Walker, Senior Solutions Engineer EMEA, BeyondTrust, Superuser accounts: What they are and how to secure them [intelligent CISO] (March, 2019), https://www.intelligentciso.com/2019/03/19/superuser-accounts-what-they-are-and-how-to-secure-them/ , (June, 2020)
4. [RedHat Documentation], What is SELinux [Security-Enhanced Linux]? (August 2019), https://www.redhat.com/en/topics/linux/what-is-selinux , (May 2020)
5. Jim B, Security Risks of a One-Way Trust Relationship between Domains (May 2015), https://serverfault.com/questions/690692/security-risks-of-a-one-way-trust-relationship-between-domains, [May 2020]
6. A. Köhler, Folder owner should not be able to change permissions / OWNER Rights explained for NTFS security (January 2017), https://blog.it-koehler.com/en/Archive/1313 , [July 2020]