Machine Identity in Cybersecurity and IAM

cloud security alliance®

The permanent and official location for Identity and Access Management Working Group is https://cloudsecurityalliance.org/research/working-groups/identity-and-access-management/

Acknowledgments

Lead Authors

Ravi Erukulla
Ramesh Gupta
Shruti Kulkarni
Ansuman Mishra
Alon Nachmany

Contributors

Kapil Bareja
Faye Dixon
Jonathan Flack
Paul Mezzera
Michael Raggo
Venkat Raghavan
Heinrich Smit
David Strommer

Reviewers

Iain Beveridge
Guillaume Cesbron
Senthilkumar Chandrasekaran
Rajat Dubey
Murali Palanisamy
Shraddha Patil
Chandrasekaran Rajagopalan
Michael Roza
Gaurav Singh

CSA Analysts

Ryan Gifford

CSA Global Staff

Claire Lehnert

Editor

Larry Hughes

Table of Contents

Abstract:

Identity management is a crucial aspect of information security, as it ensures that only authorized individuals and entities have access to sensitive data and resources. With the increasing use of technology in today's organizations, identity management has expanded to include (but is not limited to) machine identities (anything other than human), such as device identities, digital identities, and workload identities. This whitepaper aims to define machine identities, explore their history and significance, and provide best practices for managing and governing the risks associated with them. The target audience for this whitepaper includes InfoSec professionals, risk office/owners, IT/cybersecurity liaisons, technology/Site Reliability Engineers (SRE)/DevOps teams, business process owners, application developers, and government/regulatory bodies.

1. Introduction

Identity management ensures that the right individuals, such as people or machines, have access to the right resources, at the right time, for the right length of time, and for the right reasons. This is vital for maintaining the security of an organization's resources. With the advent of new technologies, identity management has evolved to include not only human identities, but also machine identities, such as devices, digital workloads, and robotic process automation (RPA) bots. This document aims to provide an understanding of machine identities and the implications of their use.

2. Definition of Machine Identities:

An identity, in general, is a set of one or more attributes that uniquely describe a subject within a given context - a person, organization, device, hardware, network, software, workload, or service (Source: NIST SP 800-63).

Identities are distinct from the credentials (e.g., passwords, keys, certificates) used to authenticate an identity.

The attributes may include a name, email address, IP address, or other identifying characteristics. Human identities are associated with individuals, whereas machine identities are associated with devices, digital workloads, and other types of entities.

Device identities are associated with physical devices such as laptops, smartphones, servers, and operational technology (OT) devices such as Internet of Things (IoT). These identities are used to authenticate and authorize access to resources and applications on a device.

Digital identities are associated with digital entities such as workloads, services, applications, virtual machines, containers, clouds, RPA bots, and APIs. These identities are used to authenticate and authorize access to resources and applications on an on-prem network or in the cloud, either by verifying credentials or certificates.

Machine identities are digital identities that can use either symmetric or asymmetric cryptographic keys, tokens, or passkeys.

- Asymmetric key encryption,also known as public key encryption, uses a public-private key pair. The public key is never a secret, and is used to lock or encrypt the payload, while the private key is used to unlock or decrypt the ciphertext. The private key must be kept secure and is typically stored in a hardware or software key vault or key store. The vast majority of machines use asymmetric keys like digital certificates to identify themselves and to establish trust. Examples include web server certificates, client certificates, SSH host keys, and SSH host certificates.
- Symmetric key encryption is not as widely used as asymmetric encryption. It is typically used only in simplistic use cases. Symmetric key encryption uses only one key, not a key pair. Examples of symmetric key machine identities include API keys, tokens, and shared secrets.

3. Background History:

The concept of machine identities has its roots in the early days of computer networks. As networks grew in complexity and size, securing access to resources and applications became increasingly important. One of the earliest methods of securing access was to assign unique identities to devices, such as IP or MAC addresses, and restricting access to the devices based on those network identities. As technology evolved, workloads became more prominent. Over time workloads have become ephemeral, leading to challenges in identifying them. Machine identities have expanded to include digital workloads, service accounts, RPA bots, APIs, etc. In addition, the proliferation of IoT and smart devices, at both homes and businesses, have added other complexities to machine identities. With the explosive growth of connected devices and machines, and a large increase in the amount of machines vs. humans, there is a strong need to focus on securing and managing machine identities.

4. Differences from Human Identities:

Machine identities are used by entities that can neither change passwords nor support multi-factor authentication. More often than not, machine identities have long passwords that do not expire. To secure the credentials, many organizations implement policies to rotate or change the password regularly. However, this poses a problem if the machine identity is embedded in an application, or is in use by a tool, and the rotation of its password breaks the dependency the application or the tool may have. One way to address these situations is to use managed identities (Azure) / roles (AWS) in cloud environments. In on-premises environments, this can be addressed with a privileged access management tool (e.g., Thycotic, CyberArk) that can discover machine identities and their dependencies.

5. Protecting Machine Identities:

Protecting machine identities is essential for maintaining the security and integrity of an organization's information and assets. Unlike human identities, machine identities do not have biometrics or other forms of secondary verification embedded into the software.

Machine identities can be allocated to any device or even used to mimic one. Therefore, it's crucial to ensure humans don't directly handle or access the private aspects of these identities. Instead, humans should focus on establishing policies and governance, while automation takes care of verification, issuance, and management of these identities.

Machine identities are typically authenticated using asymmetric key pairs. Humans should not have clear text access to the private keys for any reason.

Machine identities can potentially be compromised by malicious users, and a malicious actor can hide behind the compromised machine identity. A machine identity typically is the assigned name or a fully qualified domain name. The logs and log events of the machine will record the name of the machine identity as the actor carrying out malicious activities. A common example is the compromise of a service account in an Active Directory environment, where the service account has "interactive logon" enabled. Any user having access to its password can log in as (spoof) the service account. Activity logs record the name of the service account, not the malicious actor. Hence, it is always preferable to disable "interactive logon" on a service account unless there is a compelling business requirement for it to be enabled.

We cannot protect assets that we do not know of, so discovering machine identities and creating an accurate inventory of them is an essential first step in protecting them. This includes service accounts, managed identities, and APIs.

Root of Trust (RoT) is the foundation of trust in an organization. It is essential for any organization to secure and protect machine identities leveraging a secure and highly reliable hardware and software. Ideally, the private keys of machine identities are stored in a Hardware Root of Trust, this adds additional cost and complexity in managing and maintaining the Identities. Software key stores are widely used to secure private keys due to its flexibility. Every organization should secure the machine identities or private keys in a software key store and essentially automating the whole process to remove the human element from accessing or managing the software stores.

6. Challenges with Machine Identities

Due to the nature of machine identities and how they are managed, they pose several challenges for organizations. Some of them include:

Discoverability and Backdoor Machine Identities:

Not all organizations follow a consistent approach to discovering and inventorying machine identities. Unlike human identities, machine identities can manifest from anywhere within an organization. Insecure coding practices may introduce backdoor machine identities, such as hard-coded credentials within an application/service/script, whether intentionally created or unintentionally left behind. Note that these are not the same as the default identities such as the admin account of a device. Those are easy to manage, but backdoor identities are difficult to discover and may require specialized tools.

Legacy Machine Identities:

Legacy machine identities can pose a significant challenge for organizations, as they may lack documentation, use vulnerable cryptographic algorithms or outdated security controls, or have uncertain ownership. Examples of legacy identities include but are not limited to identities of printers, CCTVs, projectors, wireless routers, etc. When dealing with legacy identities, it is important to take a risk-based approach and prioritize efforts based on the level of risk posed by each identity. This may involve retiring unused identities, rotating keys, or updating security controls.

Alternatively, when a known legacy identity is in use, an organization may look towards implementing compensating controls to prevent compromise of the credentials.

One example of a compensating control is to ensure that devices with such identities are on a separate network that is air-gapped or segmented from the network that processes sensitive data.

Another example is to ensure that such devices are on internal segmented networks and not on public-facing networks.

Lifecycle Management of Machine Identities:

An important aspect of managing machine identities is ensuring that they are kept up to date throughout their lifecycle. This includes tasks such as provisioning new identities, revoking or deactivating old identities, and ensuring that existing identities are still active and in use. Assigning a distinct identifier to each machine identity, and documenting its dependencies, will help to enable access provisioning and policy enforcement.

A well-defined process for managing the lifecycle of machine identities helps ensure that identities are created and used appropriately, and that they are revoked or deactivated when no longer needed.

Perpetual Ownership:

Another challenge in managing machine identities is dealing with the issue of perpetual ownership. Unlike human identities, which are associated with specific individuals, machine identities may be owned by multiple individuals, devices, or entities over time. For example, an RPA bot may be owned by an individual or organization who is responsible for its development and operation, but it may also be used by multiple organizations or teams. Ensuring that there is a clear process for managing ownership and repurposing of machine identities ensures appropriate governance.

It is equally important that proper controls are in place to ensure owners of these machine identities do not intentionally or unintentionally abuse their privilege and commit fraud through a toxic combination of transactions. Such transactions often go unnoticed and are hard to correlate against after the fact, so prevention through proactive controls is best. Controls such as storing the credentials in a vault and rotating them on a regular basis may prevent such compromises.

Governance of Machine Identities:

Ensuring that machine identities are governed effectively is essential for maintaining the security of an organization's information and assets. This includes ensuring that identities are created and used appropriately, are revoked or deactivated when they are no longer needed, and are owned and managed by the right individuals or entities. Developing and implementing a comprehensive lifecycle helps make sure that identities are managed effectively and the associated risks are minimized.

Centralized Management of Machine Identities:

Machine identities, originating from various organizational departments, are frequently mishandled due lacking recognition of the implications of improper management. Collaboration across diverse teams - such as Application Development, IT, Security, IAM, DevOps, Identity Governance, and Cloud Infrastructure - is crucial to incorporate a centralized management system for these identities. Centralized management not only enhances visibility and control but also permits the application and enforcement of standardized policies and procedures. It ensures consistent security practices, boosts the efficiency and effectiveness of processes, and aids in risk, audit, control, and compliance activities. Additionally, centralized management helps event investigations, vulnerability identification, and patching activities.

7. Best Practices:

Effective management of non-human identities requires a combination of technical and organizational controls. Some best practices for managing machine identities are given below.

Life-cycle Management

- Implementing a formal process for managing the lifecycle of machine identities, including provisioning, de-provisioning, and rotation of keys and certificates.
- Establishing ownership and accountability for machine identities, including assigning a responsible party for each machine identity and ensuring that this information is well-documented.
- Defining a clear relationship between identity and role grants, and ensuring visibility into this relationship.
- Making "crypto-modularity" a best design practice in your application development so the application can be changed or rerolled without having to recode the entire application.
- Treating machine identities similar to human identities, by implementing the principles of least privilege and Just-in-Time (JIT) Access, which ensures that machine identities have only the necessary access for a limited time to effectively limit the scope of privilege and exploitation.
- Using managed identities (Azure) / roles (AWS) in cloud environments to keep the probability of identity compromise low. This is because the credential secrets are managed by the cloud provider.
- Reducing manual compliance tasks across machine identities. Apply automation for elevated access requests, issuance, renewal, and revocation tasks.
- Implementing a centralized system to manage machine identities, to provide complete visibility into all machine identities it owns.
- Treating the maturity of devices and workloads as distinct factors, since improving the infrastructure for devices may require different methods than those needed for onboarding tools.
- Establishing continuous monitoring (monthly, quarterly, etc.) for access reviews of machine or workload identities, to determine if any are dormant.
- Decommissioning inactive machine or workload identities as part of normal hygiene.
- Identifying and right-sizing overly-permissioned machine identities, by comparing permissions granted vs. those being used.
- Providing tailored guidance to developers, I&O, DevOps and security teams by defining how the tools in technology stacks should and should not be used, and under what circumstances they can be deployed.
- Implementing a secure key orchestration mechanism to automatically validate, verify and issue identities based on inherited trust, thereby preventing human access to machine identities.

Vaulting and Authentication

- Centralizing and storing digital certificates, SSH keys, and secrets in secured locations, preferably in Hardware Security Modules (HSM) or key vaults. Moreover, access to these devices should be limited to privileged users with strong passwords or RBAC.
- Moving to an identity-centric approach on top of common tools like gateways, encryption, or key management when possible.

Continuous Controls and Monitoring

- Ensuring continuous monitoring and auditing of machine identities for suspicious activity.
- Using anomaly detection, when possible, to determine when there are abnormal machine identity activities.
- Periodically detecting compromised machine identities, to disable or deactivate them.
- Incorporating machine identity management into overall security and risk management processes, and implementing compensating controls where risks cannot be mitigated.
- Identifying and documenting machine identity-related outages, and creating an awareness program to prevent further ones.
- Ensuring compliance with relevant government and industry regulations.
- Enforcing separation of duties, not just for machine identities, but also across combinations of identities and owners. A person with control over a machine identity should not be able to perform a toxic combination of transactions, e.g., perform part of a malicious transaction as the machine identity and another part as themselves.
- Ensuring that machine identities do not possess admin-level interactive human-type permissions that allows changing of role permissions, creating additional users, etc. Human-type permissions should not be assigned to a machine identity without business justification.
- Alerting and monitoring privilege escalation activities carried out by machine identities, by tracking tasks that are tied to known TTPs (Tactics, Techniques and Procedures) or behaviors associated with permission changes, lateral movement (e.g., cross-account access), and sensitive infrastructure components such as virtual firewalling.

8. Conclusion:

In conclusion, machine identities are an essential aspect of identity management. Understanding the unique characteristics and risks associated with these identities, and developing best practices for managing and governing them, is crucial for maintaining the security of information and assets. By implementing effective identity management strategies, organizations can ensure that the right machine identities, as with individuals, have the right access to the right resources, at the right time, for the right intention, thereby minimizing the risk of unauthorized access. This document provides the foundation for understanding machine identities and their management within an organization.