How to Design a Least Privilege Architecture in AWS

Dave Shackleford
Introduction: What Is Least Privilege?

On the surface, the concept of least privilege is seemingly obvious: In any given scenario or use case, only allow a user, service, application or system to operate with the bare minimum privilege necessary to successfully accomplish the business goals desired. However, over decades of computing, consistently implementing least privilege as a best practice has been a challenge for a number of reasons, including:

- **The ability to determine the appropriate “least privilege” for a given use case is a surprisingly complex issue.** It’s often challenging for administrators, engineers and developers to plan for and think through the exact set of privileges needed to implement a least privilege access model because of widely differing access needs from different types of users and services.

- **It is easier to allocate more privileges than to limit access.** Security professionals have observed this classic problem in many different scenarios over the years, ranging from data center administration to development and application interactions to end users on their workstations. It’s much more convenient for workers to do whatever they need to when they are assigned extensive privileges.

- **The range of permissions and privilege models varies widely between environments and applications/services.** Because there is little to no commonality across the use cases and technologies we employ from one organization and environment to the next, developing a consistent model of least privilege can be time-consuming.
That said, even successful least privilege implementations tend to shift and drift over
time without continuous monitoring and oversight.

**Least Privilege Concepts in the Cloud**

Security professionals are rethinking the approach to least privilege security concepts for
the public cloud. Some key factors to address include:

- **Vanishing perimeter**—It is time to look at the entire environment as potentially untrusted or compromised, rather than to continue thinking in terms of “outside-in” attack vectors. The cloud is a cohesive ecosystem that relies on numerous service and application interactions, and the classic idea of the perimeter is changing.

- **Application workloads**—Security professionals need a better understanding of application behavior at the workload level. They should be looking at the types of network communication approved applications really should be transmitting.

- **Trust relationships**—The focus should be on trust relationships, system-to-system relationships and service-to-service relationships within all parts of the cloud environment. Most communications in enterprise networks today are either wholly unnecessary or irrelevant to the systems or applications really needed for business, and organizations don’t want this problem replicated in cloud deployments.

**Pillars of Least Privilege**

Security teams need strong access controls to effectively secure who can do what and from where. The “who” could be a user or app identity or systems/subnets within the environment. Many cloud access management strategies are starting to revolve around the idea of least privilege at all layers, which some may call “microsegmentation” or a “zero trust” design. Whatever you choose to call it, the three elements of this strategy, illustrated in Figure 1, are:

- Identity and access management (IAM)
- Network access and segmentation design
- Cloud security posture management

For many organizations, designing least privilege access controls often encompasses a blend of cloud-native and third-party controls as well. This area is evolving quickly, so security teams should pay careful attention to the market and open source communities too.
Although the first two pillars are more critical, all three are needed to enable a comprehensive least privilege strategy. In the coming pages, we explore the three pillars of the least privilege model and look at how they can work together to implement an effective least privilege strategy.

**Least Privilege Pillar 1: Identity and Access Management**

Arguably, one of the most important aspects of cloud security is IAM. If you think about it, IAM is a linchpin to controlling most elements of security for who and what can access resources in the cloud. Defining roles, enabling strict access models and limiting the resources available to users and systems is a critical step in enabling a sound cloud security strategy overall. A key element of IAM that security teams need to adapt to is the use of IAM for enveloping assets, allowing them to create least privilege architectures with affinity policies in place.

**User Relationships**

IAM users are associated with credentials for making API calls to interact with cloud services and exist only within the cloud environment itself. By linking directory services like Active Directory to the cloud, security teams can leverage existing in-house users and map them to IAM groups and roles, but a standalone user created within the cloud is only useful in the cloud. New IAM users have no permissions (an implicit “Deny All” policy). This is a good thing, because permissions must be explicitly granted. This policy can also help with the common problem of over-allocating privileges to users and groups in the environment.

IAM users can represent any asset/resource—an IAM user is a simple identity with associated permissions. This means that IAM users can be enabled for application access to Amazon Web Services (AWS) resources too, not just as actual interactive user accounts. Once you create service-oriented users, place them in defined groups, if warranted. Security teams can assign permissions and privileges directly to users (not advised) or groups (better to manage and maintain).

**Service Relationships**

For service interactions within the environment, however, cloud security teams should focus on defining specific roles. There are four types of roles:

- **AWS services**—This type is for provisioning roles that will be assigned to AWS services like Amazon EC2 and AWS CloudFormation. In other words, what resources can access other resources in AWS, and what actions can they take? This type of role forms the basis for instance profiles, which we cover in a moment.

- **Cross-account access**—Teams can provision access to their AWS infrastructure to other AWS accounts the organization owns or to third-party AWS accounts.

---

1 This paper mentions solution names to provide real-life examples of how cloud security tools can be used. The use of these examples is not an endorsement of any solution.
• **Federation**—For federating access with SAML 2.0 to in-house directories, a federation role is available.

• **Identity providers**—These role types work with identity providers (IdPs) for single sign-on (SSO) and federated access to resources. There are three types of IdP roles. The first focuses on web IdPs like Google, Facebook and Amazon Cognito. The second grants web-based SSO to Security Assertion Markup Language (SAML) providers, likely some of the most common for management console access. For direct SSO access to APIs via SAML, a third type of IdP role is available.

There are several distinct types of identity-focused least privilege orientation for cloud deployments and infrastructure. First, there should be a focus on any privileged users that need access to the cloud environment for administration, engineering or security-focused tasks. Ideally, even in large organizations, this should be a relatively small number of users that are carefully set up and monitored. The best practice for these users is to federate their internal user accounts directly to an assigned role within the cloud environment that has the fewest privileges assigned.

The second major type of least privilege access model that all organizations need to consider is associated with deployment pipelines and associated systems and services. Whether on premises or fully hosted within the cloud environment, deployment pipelines need certain privileges to update workload images and containers, access code repositories, assign metadata tags to resources and monitor performance and security metrics and activities.

The third major type of least privilege focus is mapping user, service and application relationships wholly contained within the cloud environment. These might be Amazon EC2 workloads with instance profiles assigned that allow access to other AWS services like Amazon S3 buckets, AWS Lambda functions that need to interact with Amazon CloudWatch logs and database services, or service IAM accounts/groups used to allow access between applications and services in the environment.

Finally, privileges should be carefully reviewed for accounts accessing other accounts’ services when a multi-account strategy is in place.

**Relationship Mapping**

For all of these different least privilege scenarios, organizations need to successfully map user and service relationships to create the most restrictive privilege models needed. Fortunately, a number of tools are available to accomplish this. During AWS IAM account creation, admins can use the AWS Access Advisor feature. Access Advisor shows AWS services allowed by the assigned IAM policy, policies assigned that grant specific permissions and last access times (if relevant). This information is especially helpful for users that are members of multiple groups with a variety of different policies in place. Many organizations have numerous groups, users and accounts that need to be handled differently, and it can get confusing. With this feature, admins can
get a sense of what permissions are being applied, ideally before they are. The AWS Trusted Advisor service also informs account owners of some well-known privilege allocation issues that may be present.

AWS IAM Access Analyzer, a feature within AWS Identity and Access Management (IAM), performs a more thorough analysis of privilege models in use. This tool helps organizations identify potential security risks in the AWS environment by analyzing the resource-based policies applied to resources within their zone of trust (the current account). When AWS IAM Access Analyzer identifies any policy that allows access to those resources by a principal that isn’t within the zone of trust, the service generates a finding/alert. Security teams can use the information in each finding, such as the resource, access level and principal that has access, to determine whether the access is necessary or unintended. If the access is unintended, and therefore a risk, security teams can modify the policy to remove the access and work toward a least privilege identity model.

With an isolation and segmentation technique, each account is a completely isolated set of resources that can be configured to access resources in other accounts. For multi-account strategies employed to limit the post-compromise risk (often referred to as the “blast radius”) and provide highly granular least privilege access models, AWS IAM is a critical element of managing the access between accounts. AWS Organizations is a service that organizations can use to define policies and guardrails to apply across multiple AWS accounts from a master control level.

With AWS Organizations, you can create service control policies (SCPs) that really govern the use of other IAM policies. AWS Organizations can control the entire account, group and role life cycle with regard to policy application, and can do so for accounts that need to interact or have some relationship. Some basic examples of how AWS Organizations could be practical would be governing business unit (BU) account use (because they may have totally different needs, but still need some central control or billing), as well as governing and controlling DevOps and other team accounts (for the same reasons). AWS Organizations is the linchpin of a multi-account blast radius limitation strategy in AWS—limiting the blast radius to the smallest possible surface area prevents attackers from leveraging one compromised asset to access another.

Creating a centralized policy model within AWS Organizations can allow security administrators to create different and least privilege policies for the appropriate accounts and assign them and/or revoke them easily. The service also provides a “master” rollup account that is often also the “payer” account that gets the consolidated billing for AWS accounts.

Setup and configuration of multi-account architectures have long been considered challenging and complicated tasks, especially for large organizations. Fortunately, numerous services and design models have been created within AWS to help with this. A sample multi-account framework to start from, called a “Landing Zone,” was proposed by cloud engineering experts several years ago, but creating and managing even this led
AWS to create a new service, called AWS Control Tower, that can automatically deploy a multi-account starting architecture. Enterprises can then use AWS Control Tower to create and implement defensive guardrails such as AWS Config monitoring rules, infrastructure-as-code definitions in AWS CloudFormation, and strict identity policies that restrict permissions and privileges across accounts, enable data encryption and much more.

**Least Privilege Pillar 2: Network Segmentation for Access Control**

The second major component of a traditional least privilege design model is network segmentation that is closely aligned with a specific type of system or workload. This is often termed “microsegmentation.” A least privilege concept of network segmentation strives to prevent would-be attackers from using unapproved network connections to compromise systems, move laterally from a compromised application or system, or perform any illicit network activity regardless of environment. By potentially eliminating lateral movement, a least privilege microsegmentation model also reduces the blast radius when an attacker has illicitly gained access to an asset within a data center or cloud environment.

The classic model for implementing least privilege at the network level starts with a network access control policy of Deny All and then adds only those types of network access needed.

**Microsegmentation with Cloud-Native Controls**

The first category of focus for any cloud network isolation and segmentation should be the core network zone associated with cloud accounts. In AWS, this is known as the virtual private cloud (VPC), and this can contain any number of distinct network subnets. Cloud-native access controls can be created and applied within the VPC and should be used for isolating and controlling traffic flow into the VPC subnets altogether, as well as to and from instance workloads running applications and services.

AWS has two built-in types of network access and isolation controls: security groups and network access control lists (NACLs). Use security groups and NACLs to control traffic into and out of network deployments. Security groups apply to instances and are stateful, whereas NACLs apply to VPC subnets and are stateless. Table 1 provides a breakdown of security groups versus NACLs.

<table>
<thead>
<tr>
<th>Table 1. Differences Between Security Groups and NACLs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Security Groups</strong></td>
</tr>
<tr>
<td>Apply to instances</td>
</tr>
<tr>
<td>Only support Allow rules (layered on a default Deny)</td>
</tr>
<tr>
<td>Are stateful</td>
</tr>
<tr>
<td>Are considered in their entirety before traffic is allowed</td>
</tr>
<tr>
<td>Must be associated with an instance to apply</td>
</tr>
</tbody>
</table>
In general, it’s best to sparingly apply NACLs to either allow or deny known trusted or malicious IP addresses and subnets. The majority of the network access controls (NACs) should be defined and applied through the use of security groups. Because security groups begin in a “default Deny” state, it’s much easier to create a least privilege model with them. Security personnel can enable Amazon VPC Flow Logs to track communications between assets in a VPC to ensure that no unusual or unexpected access is allowed, but a more complete coverage option to audit large numbers of security groups is to look into third-party network policy analysis tools that can ingest security group definitions and analyze them at scale.

**Advanced Network Security Segmentation and Access Controls**

To segment and control traffic at the application layer, or define policies focused more on application details and protocols, a third-party solution likely makes more sense, and many cloud options are available for enterprise-class networking. Most major cloud providers offer enterprise-class solutions that are capable of providing more granular policies and monitoring. Today’s next-generation firewall (NGFW) platforms are often used to provide network intrusion detection and prevention, traffic inspection and behavioral monitoring, and centralized configuration and administration alongside existing on-premises NGFW platforms if desired. Leading providers include Palo Alto Networks, Fortinet, Sophos and others.

**Segmentation/Isolation Best Practices**

There are many well-known security fundamentals that organizations can follow when planning for and implementing least privilege network isolation and segmentation in the cloud.

First, be sure to consider what types of architectures make the most sense. For example, you can create all distinct assets in one very large VPC and control access with security groups and NACLs, or create a much more granular isolation strategy with multiple accounts and VPCs. Most major cloud providers support the concept of peering between virtual networking boundaries. VPC peering enables organizations to couple distinct VPCs together, allowing assets in one network to talk to assets in another. This capability can be incredibly useful in a design model because you can create true hub-and-spoke network designs that require traffic to pass through a transit zone of some type (through a dedicated security zone with intrusion detection and other controls, for example).

VPC peering is not transitive (i.e., there is no need to specifically allow it for each VPC peered together). In this case another type of platform, called a “transit gateway,” can simplify multi-VPC architectures significantly. This resource, which can be managed through the AWS Resource Access Manager service (for managing assets across accounts), can help teams create a more traditional hub-and-spoke model of network connectivity that will then have security groups and NACLs applied as needed. Much like route control, transit gateways can have IPS or firewall appliances attached as well, making these ideal for a central security control point. For managing multiple transit
gateways, the AWS Transit Gateway Network Manager (AWS Network Manager) service enables organizations to manage all connected hybrid cloud network zones connected to and through transit gateways in a single dashboard. In many cases, teams set up a “transit VPC” with an NGFW platform as described earlier to process traffic to all other zones peered within the network architecture.

To summarize how IAM and core networking controls can facilitate a least privilege cloud deployment, be sure to:

- **Plan IAM roles and permissions to protect access to and use of VPC resources and services.** Many VPC objects and services can easily be controlled through IAM, including EC2 workloads, containers and much more.

- **Leverage security groups and NACLs to the full extent.** These controls provide built-in cloud-native NACLs to workloads and between subnets. If you need more control (and you likely will), consider a third-party virtual firewall/IPS appliance as a gateway.

## Least Privilege Pillar 3:
### Cloud Security Posture Management

Cloud security posture management (CSPM) tools can assess the actual control plane of the cloud environments in use for compliance assessment, operational monitoring, DevOps integrations, risk identification and risk visualization. A CSPM platform should continuously monitor cloud security risk and potentially implement configuration changes in the cloud environment that facilitate least privilege access and much more. These tools also offer threat detection, logging and reports. In addition, they usually provide automation to address issues ranging from cloud service configurations to security settings as they relate to governance, compliance and security for cloud resources. Because many cloud platform settings relate to networking and IAM configuration, having a continuous monitoring engine that highlights over-allocation of privileges and permissive traffic policies can be invaluable.

Having interoperability between monitoring and automation is a critical advantage of a CSPM. For enterprises grappling with hybrid architectures and container environments, where misconfiguration is a common threat to cloud security, a CSPM tool is an excellent step toward implementing continuous monitoring and alerting for the cloud provider fabric configuration. Common misconfigurations tend to be present with identity controls, workload security, logging enablement, network configurations and more. Organizations that are moving to or currently in hybrid deployment scenarios should strongly consider CSPM tools.
A Least Privilege Use Case

For an organization planning on deploying to a platform-as-a-service (PaaS) or infrastructure-as-a-service (IaaS) cloud environment with a focus on least privilege, there are multiple recommended steps:

1. Identify roles and responsibilities for team members requiring access to the cloud infrastructure.
2. Determine the type of network access needed.
3. Evaluate IAM roles and privilege assignments.
4. Monitor the cloud control plane.

In the first step, those responsible for the least privilege strategy carefully identify roles and responsibilities for any cloud engineering, DevOps and security team members that may need access to the cloud infrastructure. This should always be the first priority because the account owner (the root account) is one that should be almost wholly disabled, other than for billing and a potential “break glass” scenario if disaster strikes. These privileged users should be established through federation and role integration if possible, or standalone users within defined groups if not. It’s best to assign the cloud provider’s predefined policies that match these administrative roles whenever possible because these are likely to be the most accurate and well-structured. Even with that said, some of these can be used as beginning policies from which to reduce privileges as needed.

The next step is to determine what type of network access is required from the internet, from a hybrid cloud dedicated network connection and within the cloud environment itself. This step requires a review of application and service architecture to define data flows with TCP/UDP ports and application behavior profiles that planners can use to carefully restrict the types of traffic needed for operations. Organizations should plan to start with cloud-native networking controls like security groups and NACLs, which allow for a strong microsegmentation approach that can be managed through infrastructure-as-code (IaC) templates, such as AWS CloudFormation or HashiCorp Terraform, and monitored through API logs and metadata queries. For more robust network security, many enterprises will want to adopt a VPC peering arrangement for additional isolation, possibly with a third-party NGFW platform introduced to provide additional application-layer protection.

Throughout this entire process, identity, development and security teams should evaluate IAM roles and privilege assignments for workloads, services and all interaction between assets in the environment. Fortunately, tools such as AWS IAM Access Analyzer can be used to perform a deep dive into assigned roles and privileges for all components within a defined trust zone such as an account. All teams involved should be invested in leveraging reports and alerts from tools like this to continuously look for...
opportunities to reduce privilege allocation wherever possible. This is an ongoing effort that will likely continue over time, because applications and assets continuously change and update within dynamic cloud environments.

Finally, it’s a good idea to consider a CSPM platform to continuously monitor the cloud control plane itself, looking for exposure and potential configuration pitfalls that could inadvertently allow for unintended or privileged access into services or the environment as a whole, as well as internal mappings of network and identity orientation that may be improved upon. For large, complex deployments, these types of third-party solutions can provide an extra set of eyes and ears on the cloud deployments overall.

**Conclusion**

A least privilege cloud architecture should include authentication and authorization controls, network access and inspection controls, and monitoring/enforcement controls for both the network and workloads. No single technology currently will provide a full least privilege design and implementation—organizations need to implement a combination of tools and services to provide the full degree of coverage needed. For most organizations, a hybrid approach of both cloud-native and third-party controls will make the most sense.

To implement a least privilege cloud environment, start with user and administrative access, followed by multi-account identity management, if applicable. From there, focus on network architecture and access control design, using cloud-native controls as the first line of defense and applying third-party controls for more robust defenses. Throughout all deployments, continuously evaluate privilege allocation and role assignments to find potential over-allocation of privileges where they may exist. Once the cloud environment is up and running, a CSPM platform may make sense to continuously monitor the configuration.

More tools and services are available than ever before to aid in building and maintaining a cloud infrastructure that adheres to the principle of least privilege. A commitment to continuous oversight is critical because cloud environments tend to change rapidly. Implement tools as needed to provide adequate logging and alerting to ensure security teams are aware of how the environment is operating at all times.
About the Author

Dave Shackleford, a SANS analyst, senior instructor, course author, GIAC technical director and member of the board of directors for the SANS Technology Institute, is the founder and principal consultant with Voodoo Security. He has consulted with hundreds of organizations in the areas of security, regulatory compliance, and network architecture and engineering. A VMware vExpert, Dave has extensive experience designing and configuring secure virtualized infrastructures. He previously worked as chief security officer for Configuresoft and CTO for the Center for Internet Security. Dave currently helps lead the Atlanta chapter of the Cloud Security Alliance.

Sponsor

SANS would like to thank this paper’s sponsor:

aws marketplace
## Upcoming SANS Training

Click here to view a list of all SANS Courses

<table>
<thead>
<tr>
<th>SANS OnDemand</th>
<th>OnlineUS</th>
<th>Anytime</th>
<th>Self Paced</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS SelfStudy</td>
<td>Books &amp; MP3s Only</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
</tbody>
</table>